Høgskolen i Østfold

MASTEROPPGÅVE

Applying etymology to increase the readability of vocational texts in the ESL classroom

Roger Jan Støyva

30.11.2022

Master framandspråk i skulen

Fakultet for lærarutdanningar og språk

Institutt for språk, litteratur og kultur

Abstract

Norwegian vocational students often face challenges when learning English. This is not least the case in the work with vocational texts, where general and technical vocabulary may pose a double difficulty. This study aims at facilitating Norwegian students' work with vocational English by developing a method for analysing and simplifying the general language of vocational texts. An important component of this effort takes into account the similarities and differences in vocabulary between Norwegian and English. While the two languages are closely related and share many cognate words, English has a great number of words with French and Latin etymology. Using words of Germanic etymology rather than Romance etymology can contribute to making the general English of vocational texts more easily readable for the target group.

The methods for investigating the issue were threefold. Firstly, a number of vocational students' knowledge of English vocabulary was tested in a 40-word vocabulary test. The answers were analysed in terms of frequency, eymology and the number of syllables. The answers seem to confirm that vocational students have the greatest mastery of words that are high-frequent, are of Germanic etymology and have few syllables. Secondly, the students answered a survey about which types of vocabulary they found most worthwhile learning. The results show that the majority of the respondents prioritise learning high-frequent words, vocational words and phrasal verbs. Thirdly, the results of the vocabulary test and the survey were used to construct a readability index. The values of the index were set to indicate which words are suitable to facilitate the general language in vocational texts. The readability index was then applied to a selection of vocational texts, general textbook texts and authentic texts, and it seems to give a realistic impression of their degree of difficulty. For future research and use of the index, it is recommended to develop a digital version that can make further adjustment and testing more efficient.

Contents

1. Introduction

1.1 Background

It is a challenge to design learning materials for vocational English that are well adapted to the students' levels of proficiency. Textbooks of vocational English contain two different types of text. Firstly, they contain general subject texts that treat topics within geography, literature etc. like most other English textbooks. Secondly, they have texts that specifically treat vocational topics like welding, construction work, health care etc. Such vocational texts inevitably contain many subject-specific words that will be new to the learners and pose a challenge for them. Given that the vocational vocabulary often has a high level of difficulty, it will be beneficial for the learners if the remaining vocabulary in the text, i.e. the general vocabulary, is made to be as easy to comprehend as possible. If that can be done, the students will be spared an unnecessary double vocabulary workload. In this master thesis the focus will be on the general language of the vocational texts. How can the general language be adapted and simplified, so that the students can be allowed to focus fully on learning the vocational subject matter and vocabulary?

The traditional approach to determining the degree of difficulty of individual words has been word frequency. Considerable work has been done to map the frequency of English words, with the assumption that the most frequent words are the easiest to learn and understand. Another approach to word difficulty can be the etymology of words. Knowledge of language history can be of help in the efforts of adapting and simplifying the general vocabulary. Parts of the English vocabulary have common linguistic roots with the other Germanic languages, e.g. the Scandinavian languages and German. A considerable number of English words, on the other hand, have their roots in French, Latin and Greek. In this thesis it will be hypothesised that, generally speaking, the Germanic-based lexicon of English will be easier to comprehend by Norwegian vocational students than the part of the lexicon which is derived from French, Latin and Greek. Using Germanic-based vocabulary will thus contribute to the desired simplification of the general language.

It can be said that the English language has acquired several "layers" of vocabulary through its history. Arguably, some of the most basic and highly frequent words originate from Anglo-Saxon and partly from the Old Norse language that influenced English in the Viking Age.

Today, the surviving stock of these words is typically used by "ordinary" people in everyday language situations (Francis Katamba 2005, pp. 163-164). Apart from the Germanic words being, generally speaking, relatively simple and easy to comprehend, the common linguistic heritage within the Germanic branch of the Indo-European language family also contributes to making Germanic-derived words easier for Norwegian students. The Germanic languages – like Norwegian, Swedish and German – contain numerous inherited words that are cognates with English. Some examples may be *hus / Haus / house*, *båt / Boot / boat* and *veke / Woche / week in* Norwegian (Nynorsk), German and English, respectively.

After the Norman invasion in 1066, a great number of French words were introduced. French was the language of the ruling classes, and much of the French-derived lexicon came to be associated with social and academic prestige (see Albert C. Baugh & Thomas Cable 2013, pp. 163-169). Later, many terms based on Greek and Latin – for a large part covering scientific topics – were added (Baugh & Cable 2013, pp. 296-297). Many of those words can also be said to belong to an "upper" stratum of the English language. This said, there are also many words that are based on French, Latin or Greek that can be counted among the simplest and most basic words in the English lexicon (see French examples of this in Baugh & Cable 2013, pp. 168-169).

It can be argued that English, perhaps to a greater extent than other languages, has one set of terms typically used by "ordinary" people for everyday purposes and another set of terms more commonly applied for academic and formal use. This is a resource that can be used to vary and adjust the levels of difficulty in textbook texts. Likewise, the similarities in vocabulary between e.g. Norwegian and English can also be used to simplify the language where that is desirable. In this study it is hypothesised that using terms from the more commonly used Germanic-based lexical set, and/or terms that are linguistically related to the students' mother tongue, will make reading comprehension easier for Norwegian students. Baugh & Cable (2013, p. 10) argue that "cognates generally are learned more rapidly and retained longer than words that are unrelated to words in the native language lexicon". Since Norwegians use the word "hus" in their mother tongue, the English word "house" sounds more familiar and should be easier to learn than French "maison".

For vocational students' comprehension of vocational texts, clarity of the language in textbooks and teaching materials is of a great importance. In a description of English in vocational studies, Torill Irene Hestetræet and Sigrid Ørevik (2018, p. 327) state that "relatively high demands in theoretical subjects represent a considerable challenge to many

students who are more inclined towards practical work". Vocational students are required to learn many subject-specific terms. According to Paul Nation (2013, p. 20) as much as 20-30% of technical texts may consist of technical words. If too many advanced academic words are used in vocational texts together with the vocational terms, the combined level of difficulty may become too high. Thus, it will be advantageous if the general vocabulary in vocational texts is chosen predominantly among the "familiar" words of Anglo-Saxon and Old Norse origin rather than among the more "advanced" terms that are based on French, Greek and Latin. Replacing advanced English general words with simpler English words that have cognates in Norwegian can make both reading and learning easier. Vocational students must of course also acquire a command of necessary academic terms, but this can be prioritised in the general-language texts instead of in the vocational texts.

1.2 Aim and research questions

The study aims at identifying which words are the most suitable to facilitate understanding of general English for Norwegian vocational students, so that processing capacity is not wasted on comprehending general words in the vocational texts, but rather can be freed to learn the vocational terminology more efficiently.

- This study explores how words of different frequency, origin and syllable structure influence the readability of texts.
- As a starting point, it is investigated which knowledge vocational students have of vocabulary, and which types of vocabulary words they find worthwhile learning.
- The results are used to construct a readability index that can be used to analyse student texts, textbook texts and authentic texts.
- The final aim is to develop and adjust the readability index so it can be an easily applicable tool for selecting reader-friendly vocabulary that can be used in practical teaching.

1.3 Overview of thesis

In Chapter 2 of this thesis, the theoretical framework will be explained. Particular weight will be laid on word frequency, etymology and syllable structure, and how these factors can be integrated in a readability index. Language history will also be emphasised, as it gives a necessary basis for understanding the present language situation and why it can be relevant to use different approaches to readability for students with different linguistic backgrounds. Chapter 3 describes the methods and materials. Firstly, a major diagnostic vocabulary test

gives an impression of vocational students' knowledge of vocabulary. Secondly, a survey shows which types of words students find important learning, and lastly a readability index will be constructed and used to analyse student texts, textbook texts and authentical texts. In Chapter 4 the results are presented and analysed, while Chapter 5 gives the conclusion of the study.

2. Theoretical framework

2.1 Introduction

The literature for this thesis covers three factors that can contribute to simplifying / adapting the general language of vocational texts. One factor is etymology, based on research that shows that cognate words from closely related languages are easier to process by learners. An overview of the linguistic origins of the English lexicon will be given, showing influences from Anglo-Saxon and Old Norse on the one side, and from French, Latin and Greek on the other. The second factor is word frequency, which is the most common approach to determining the degree of difficulty of vocabulary. A third factor is the number of syllables in each word (syllable count). An attempt will be made at combining these three factors and developing a readability index which can be applied to make English texts more reader-friendly for Norwegian vocational students.

2.2 Some relevant factors for readability

From the 1920s onwards, a number of readability formulas, or readability indexes, have been developed. The aim is to attain the optimal readability of texts, for each group of readers, by carrying out calculations with text properties like the average length of words, average sentence length and word frequency. William H. DuBay (2004) gives an overview of the development of readability indexes. He has a both practical and academic approach, and

outlines (p. 1) how better readability can enhance comprehension, which may in turn improve people's ability to cope in society and even save lives.

The difference between readability and legibility is emphasised by Dubay (p. 3). Readability focuses on writing style with the aim of creating a clear language, a measure of how easy it is to read words and sentences. Another important aspect of readability is how well the text matches the reading skills and knowledge of the target reader group. Good readability contributes to the readers' success in comprehending and benefiting from a text. Legibility, on the other hand, deals with the (typo)graphic qualities of a text, like the choice of fonts and the division into columns. Neither legibility nor the content of a text – how coherent and well organised it is – is included in the notion of readability. It must be emphasised that a text with simple words is not necessarily always easy to comprehend. The sentence "The Earth is weightless" contains simple words but it nevertheless expresses a concept which is difficult to grasp (National Partnership for Reinventing Government, 17 Sep 2022).

Frequency – that is, how often the learner hears, sees and understands the word – is an important factor that can make a new word easier to learn and process for the second language (L2) learner. Estimates vary as to how many times a learner must encounter a new word before remembering it. Stuart Webb et al. (2012, p. 96) have looked at how fast single words are learnt incidentally through reading. The form and meaning of a word will be learnt better the more frequently it is encountered in the reading context. A study by Marlise Horst et al. (1998) concluded that learners must encounter the new word eight times. Rob Waring and Misako Takaki (2003), on the other hand, found that 20 encounters were necessary. Webb et al. (2012, pp. 96-97) explained the differences by pointing out that reading contexts differ – some texts contain more useful information than others, while other texts may appear to be confusing or misleading to the reader.

Patsy Lightbown and Nina Spada (2006, p. 98) point out that even more encounters may be necessary before the learner is able to grasp the word in a fluent conversation, or to understand it in new contexts. Even though estimates may vary, these studies demonstrate that vocabulary learning may be difficult. In demanding contexts, e.g. when working with new vocational vocabulary, it makes sense to simplify non-target vocabulary in the texts to make the learning process more manageable. This conclusion is further strengthened by findings suggesting that readers with the largest L2 vocabularies learn the greatest number of words (Horst et al., 1998, abstract). This phenomenon gives reason to adapt the L2 language that vocational learners have to comprehend in textbooks and learning materials. When the results

of that effect accumulate over several years of school, students who are slow learners of vocabulary are left with comparatively small resources to tackle vocabulary learning, and this could be compensated by simplifying the general language of the vocational texts.

The frequency of a word within the English language itself is not the only determining factor for how easily it is recognized and learnt by foreign learners. Cognates and other words which resemble each other in the learner's first language (L1) and second language (L2) also contribute towards making learning easier. Nation (2013, pp. 44-45) states that the lexical learning burden will be lessened if the new word reflects patterns and knowledge that the learner is familiar with. According to Nation, the new L2 word will be maximally easy to learn if it

- has sounds that are used in the L1
- is spelt similarly to L1 words
- is a loanword with generally the same meaning in the L1 as in the L2
- follows the ordinary grammatical patterns of the L1
- has roughly the same collocations and constraints

Nation (ibid.) adds that teachers can ease the lexical learning burden by pointing out patterns, analogies and connections between L1 and L2. Teachers should estimate the learning burden for each of the factors of a word, and actively seek to explain patterns that will improve the students' comprehension. It can be argued that it will be easier to fulfil the five criteria stated by Nation if the L1 and L2 words are closely related, notably if they are cognates.

2.2.1 Adapting lexical learning to the learners' linguistic background

Lightbown and Spada (2006, pp. 98-99) have constructed three lists of words and discussed how easy they would be to recognise and understand for new L2 learners of English. These lists can serve as an illustration of how English learners with a Germanic language background may have different needs than learners with a Romance language background.

List 1 (basic English words)		List 2 (international words)		List 3 (Latin-derived words)	
ENGLISH	NORWEGIAN	ENGLISH	NORWEGIAN	ENGLISH	NORWEGIAN
friend	ven [frende]	hamburger	hamburgar	government	regjering
more	meir	coke	coke, cola	responsibility	ansvar

town	by	t-shirt	t-skjorte	dictionary	ordbok
book	bok	walkman	walkman	elementary	grunnleggjande
hunt	jakt	taxi	taxi, drosje	remarkable	merkverdig
sing	synge	pizza	pizza	description	skildring
box	boks	hotel	hotell	expression	uttrykk
smile	smile	dollar	dollar	international	internasjonal
eye	auge	internet	internett	preparation	førebuing
night	natt	disco	disko(tek)	activity	aktivitet

Table 1 The table of example words from Spada and Lightbown (2006, p. 98), with Norwegian translations.

In Table 1, Lightbown and Spada's table of examples has been replicated, and Norwegian (Nynorsk) translations have been added. Lightbown and Spada connected the following comments to each of the lists:

- The words in List 1 are simple, monosyllabic, and belong to the 1,000 most frequent words in English. Despite of this, they are not necessarily easy to recognise for learners who have no previous knowledge of English. Lightbown and Spada state that neither the orthographic form nor the pronunciation reveals the meanings of these words, therefore the learners must be exposed to them many times before they are learned.
- The words in List 2 are international loanwords that occur in many languages. Lightbown and Spada point to the fact that many students know such words before they start studying English.
- The words in List 3 are rather of a technical and academic character. They are polysyllabic, and most of them are "fairly infrequent" in English according to Lightbown and Spada (p. 99). Despite this, Lightbown and Spada claim that many students either know them from their L1 or will learn them very easily, "because they have a clear resemblance to their translation equivalent in other languages" (ibid.). Lighbown and Spada add that even though these words are of Latin origin, they occur not only in Romance languages. Lightbown and Spada use the words in List 3 as examples of cognates.

When the Norwegian translations in Table 1 are taken into account, this leads to different conclusions than Lightbown and Spada's. Words that are similar in both languages have been marked with a green filler colour in the table.

- In List 1, which contains high-frequent basic words, seven out of ten have similar forms in English and Norwegian. This reflects the close linguistic relationship

between English and Norwegian, which notably comes to expression in the short, basic words. Lightbown and Spada found no cognates in List 1, but many such words can clearly be regarded as Norwegian-English cognates.

- List 2 shows that Norwegian has adopted as loanwords all the English words that Lightbown and Spada listed from popular culture, with only minor adaptations of their orthography.
- From List 3, however, it can be seen see that Norwegian uses few of the Latin-based words that English has adopted. This difference between Lightbown and Spada's description of List 3 as familiar words, and the Norwegian language users' impression of the same words as foreign-sounding, reflects the fact that Norwegian has a lower incidence of Latin-derived words.

As a conclusion, it seems that Lightbown and Spada have based their analysis on how speakers of Romance languages will experience their encounter with English. Speakers of Italian, French and Spanish are familiar with the words in List 3, while the List 1 words are unfamiliar to them. For speakers of Germanic languages like Norwegian, Swedish and Danish, the situation is quite the opposite. Speakers of Germanic languages will learn the basic English words more easily, because many of them are cognate with words in their own L1, while many Latin-derived words are experienced as foreign and unknown. This shows the importance of taking the local language situation into account, rather than uncritically adopting research findings that have been conducted in other contexts.

2.2.2 Word relatedness – cognates and loanwords

Etymologically speaking, the word *cognate* means "born together". It stems from the Latin prefix co-, which signifies "together", and the Latin verb *nasci*, which means "to be born"

(Merriam-Webster 2019, the entry for "Cognate"). Two words that are cognates are thus "born together". The word "cognate" is not unambiguous in English, however. According to the Vocabulary.com Dictionary, "a word is cognate with another if both derive from the same word in an ancestral language". The word "cognate" can also be used about blood relatedness between persons, e.g. a brother and a sister. However, the dictionary also gives a wider definition where both words and persons can be termed cognates even though they are not "genetically" related but merely share the same characteristics (Vocabulary.com Dictionary 2019). The Merriam-Webster online dictionary confirms this ambiguity by saying that cognate words are "related by derivation, borrowing, or descent" (Merriam-Webster 2019). This wide definition from general-purpose dictionaries reflects how the word "cognate" is used by the general public. Lightbown and Spada (2006, pp. 98-99) use the wider definition of the word cognate when they state that "[w]ords that look similar and have the same meaning in two languages are called COGNATES".

Other linguists use a narrower definition of "cognate". Martha Lengeling (1995, pp. 17-18) defines "true cognates" as words that two languages have inherited from the same ancestor language. Lengeling, who is concerned with the relationship between English and Spanish, uses the words *animal* and *secret* / *secreto* as examples of true cognates. Both words are derived from Latin. *Animal* has got an identical orthographic form in both languages (although the pronunciation differs), while *secreto* has received a suffix in Spanish. In the case of English and Norwegian, the words *house* / *hus* and *boat* / *båt* can serve as examples of true cognates, since both languages have inherited them from Proto-Germanic (Douglas Harper 2019).

However, words that are merely borrowed from another language are not recognized as cognates in this narrower sense of the term. Lengeling mentions the words *sandwich* (borrowed into Spanish from English) and *canyon* (borrowed into English from Spanish) as examples of such borrowings. According to this narrow definition, the same word, *sandwich* (borrowed into Norwegian from English), and the words *they*, *them* and *gift* (borrowed into English from Old Norse in the Viking Age) are not English-Norwegian cognates but borrowings. (Harper, 2019.)

Finally, Lengeling points out the challenge of false cognates ("false friends"), which are words that are formally identical or near identical in two languages. Even though they look similar, they have different meanings. False cognates may either have a common etymology, or not. As an example, Lengeling mentions the risk of confusing the Spanish word *librería*

(bookstore) with its English false cognate *library*. Both words share the same Latin origin, but they have developed different meanings through history.

Lightbown and Spada use Latin-derived words like *government*, *elementary* and *description* as examples of cognates, because such words stem from a common Latin root that has developed into parallel and equivalent words in many different languages. An illustration of this process is the Latin verb *gubernare*, which is the root word of the modern-day "parallel" words *gouvernment* in French, *governo* in Italian, *gobierno* in Spanish and *government* in English. These four words can be said to be cognates since they stem from a common ancestor word, and since they are (by and large) formally similar enough to be recognized by language learners. Another question is whether e.g. the English and the Spanish systems of government are similar enough to warrant the use of the words *government* and *gobierno* as synonyms. In some cases, originally synonymous words have grown semantically dissimilar in different languages and have thus become "false cognates" because the reality they describe is different from country to country.

2.2.3 How cognates can facilitate language learning

Cognates are easier to learn and retain than other vocabulary, according to multiple research studies (see Agnieszka Otwinowska and Jakub Szewczyk 2017, p. 2). This is called the "cognate facilitation effect". Cognates (both identical and nearly identical cognates) are translated faster and more correctly than other words, and cognates are easier to identify. Ton Dijkstra and Walter van Heuven (in Otwinowska and Szewczyk, 2017) claim that the cognate facilitation effect is due to a parallel activation of the orthographic representation of the word in the mental lexicon. Further, Tamar Gollan, Kenneth Forster and Ram Frost claim that two cognate words may share the same orthographic representation in the mental lexicon of a bilingual person (ibid.).

In a study of words from Hebrew and Arabic, Raphiq Ibrahim (2014) found that cognates that have a phonological overlap are more easily recognized and learnt than non-cognate words. Ibrahim, therefore, terms cognates as resources in vocabulary learning. It should be noted that Semitic languages, in which words that have related meanings share the same three-consonant root, have a word formation structure that favours the formation of cognates. For example, the

consonant group *ktb* has the basic meaning of "to read". Different affixes are applied to create concrete words with related meanings, like *ki'taab* (book) and *'kaatiib* (clerk). Since Hebrew and Arabic share this three-consonant system, cognates between those languages are more likely to occur than is the case in language families which lack such a system, like e.g. the Germanic language family.

In their work with L2 acquisition, Susan Gass et al. (2013 pp. 207-208) state, not surprisingly, that the L1 plays an important role in language learning. They quote a review undertaken by Singleton of several studies that show a connectivity between the L1 and L2 lexicon. Although L1 and L2 lexis are separately stored, there is communication between them. The relationship between a given L1 word and its L2 counterpart in the lexicon varies from learner to learner, depending on acquisition factors and which formal and semantic connections the learner has made between the L1 and L2 word. It seems that the lexica of both the L1 and the L2 remain activated when a speaker uses the L2. In a study of bilinguals who spoke English and Spanish. Gretchen Sunderman and Judith Kroll found that similar words ("neighbours") in both languages – like English *gate*, English *game* and Spanish *gato* (cat) - had an influence on each other (in Gass et al. 2013, pp. 207-208). The influence of the L1 word decreased as the learners grew more proficient in the L2. Gass et al. also quote research that claims that a cognate effect is important only if the L2 is learned early. Furthermore, Gass et al. refer to research by Dijkstra and other scholars, that shows that there is a facilitation effects for cognates. When words were presented out of context, cognates were recognised faster than noncognates, while results were mixed for words that were presented within a context. Both nouns and verbs show a cognate effect. Processing a cognate activates both orthographic, phonological and semantic properties of the word, and this may be the reason for the cognate effect.

Nation found that the learning burden will be easier if words in the L2 are similar to words in the L1. Nation recommends that teachers should "point out connections between the second language and the first language" (2013, pp. 44-45). In the same vein, Nation states that "the burden of making the form-meaning connection is light if the word being learnt is a cognate or a loanword shared by the first language and the second language." (2013, p. 74). Nation refers to a study by Frank E. Daulton (1998), which showed that English loanwords helped Japanese students in their learning even though some of the loanwords have a narrower meaning in Japanese than they have in English. This lightening of the comprehension burden

is just what is needed for making the general language in vocational texts more reader-friendly.

Lengeling (1995, p. 18) mentions some disadvantages of using cognates. The cognates may appear to be too formal, stilted and unauthentic, and can lead to the wrong pronunciation. In Lengeling's case, Spanish was the L1 and the cognates were in English. The two first objections are not very valid for the situation in Norway, since the Germanic-based cognates generally will be "simple" and informal versions of English words. The situation is more or less the opposite. It may perhaps lead to an oversimplification of the learners' language. Pronounciaton will probably be little affected in the case of cognates like *hus / house* and *båt / boat*.

2.2.4 Psycholinguistic approaches to cognates

Studies of bilinguals show that both languages are being actively processed simultaneously. Judith Kroll et al. (2005, pp. 27-29) reviewed the literature about psycholinguistic approaches to the bilingual lexicon, covering both reading and speaking. Their research suggests that information about both languages is activated when bilinguals read and speak. There is no single mechanism to "switch off" one of the languages. When trying to recognise a word form, orthographic and phonological codes in both languages will be activated and information from both languages will be considered, until one word form is selected. Such an activation takes place not only when L2 is the target language but, more surprisingly, also when the word form is being activated in L1. Activation also occurs even if cognates in the two languages do not have the same orthography. Kroll et al. (2005) concluded that both orthography, phonology and word meanings of both languages are activated when bilinguals read an L2 text.

According to Kroll et al. (2005, pp. 28-29), frequent words will be more rapidly recognized. Words will also be more rapidly recognized when their spelling-to-sound relationship is unambiguous. When the same word can be pronounced in different ways, it seems to increase processing time because two alternatives compete with each other. The number of orthographic neighbours of the target word is also important. Cynthia Siew (2018) defines an orthographic neighbour as a word which differs from the original word by only one letter. Thus, an orthographic neighbourhood may be understood as consisting of a group of such

similar words. As an example, Adam J. Parker et al. (2021, p. 1) mention that the orthographic neighbourhood of *sleet* consists of the words *fleet, sheet, skeet, sweet, slept, sleek* and *sleep*. Kroll et al. (2005, pp. 28-29) maintain that in languages that have a "deep orthography" and many exceptions (like English), larger orthographic neighbourhoods tend to give shorter response times and thus make the words easier to process, while the opposite seems to be the case in languages with "shallow orthography" (like Spanish and Dutch). In languages with a "deep orthography" there is not a good correspondence between the words' written form and their pronunciation.

Working with cognates in English and French, Katherine J. Midgley et al. (2011) carried out an experiment of how the cognate status of words affects their comprehension by L2 learners. In their study, cognates are defined as words having a common etymological background – a "nonaccidental overlap of form in translation equivalents" (p. 1634). Some cognate words, like "fruit", have a complete orthographical overlap (sharing the same spelling both in English and French). Others have a "near complete" overlap, like "mask" in English and "masque" in French. Midgley et al.'s study took as a starting point that cognates are recognized more rapidly, and translated more quickly, than non-cognates in L2 acquisition. Such effects are not necessarily present in the processing of one's L1 (p. 1634). Generally, Midgley et al. explain the cognate effect with the increased exposure that a cognate word receives by being present in both languages. Most notably, the association between the cognates' form and their meaning receives an increased exposure (p. 1635).

Midgley et al. (p. 1636) used results from EEG studies, i.e. measurements of electrical activity in the brain applying electroencephalograms, on English native speakers who studied French as their L2. The test subjects had electrodes connected to their scalps, monitoring their brain activity. As the subjects received stimuli during the experiments, the event-related potential (ERP) was measured, recording aspects of the electrical activity in the subjects' brains. Among the EEG results, an ERP-component called the N400 is of particular interest. The N400 results have been found to reflect processes connected to word recognition. A long N400 reaction time (an increased amplitude) is a sign that a word is difficult to process, or difficult to integrate into its context. Such difficulties seem to reflect challenges with interpreting the connection between the form and meaning of the word. Midgley et al. (ibid.) hypothesized that cognates would be more easily processed by the brain than noncognates, since the cognates have a shared form-meaning connection in both L1 and L2. They also hypothesized that cognates in L2 would be more easily processed than cognates in L1, since

the L2 translation equivalent benefits from all the exposures the word form has experienced in the mother tongue.

The results from Midgley et al.'s study are mixed, but they confirm that cognate words are more easily processed than non-cognate words. Contrary to previous research in the field, it was found that cognates were more easily recognized not only in L2 but also in L1, possibly reflecting that the L1 cognates had played an active role in L2 vocabulary acquisition. In one context, there was a "reverse cognate effect" of L2 words, i.e. that cognates were more difficult to recognize. A possible explanation of this is differences of pronunciation, which can be quite significant between orthographically similar English/French cognate pairs. The word "fruit" has the same spelling in both languages, but its pronunciations in English and French, respectively, are quite dissimilar. In the case of Norwegian-English cognate pairs, pronunciation will probably create fewer problems.

Siew (2018) has used network science to study how orthographic word forms are organized in the mental lexicon and how that organization influences visual word recognition. This topic is interesting in connection with the question of students' recognition of cognates. In network science, the mental lexicon is seen as a language network, where words are related to each other semantically, phonologically and orthographically. Siew took as a starting point that previous studies have shown that the structure of the semantic and phonological language networks influences both language acquisition and spoken word recognition. Table 2 is an overview of the results of Siew's study of orthographic networks.

	Reaction time		Accuracy	
Tasks:	**Speeded naming** (Reading a word that appears on a screen)	**Lexical decision** (Deciding if an item is a real English word or a nonword)	**Speeded naming** (Reading a word that appears on a screen)	**Lexical decision** (Deciding if an item is a real English word or a nonword)
Network variables				
Degree (How many orthographic neighbours a word has)	Words with high degree were **more quickly** named	Words with high degree were **more quickly** recognized	Words with high degree were **more accurately** named	Words with high degree were **more accurately** recognised
Clustering coefficient (Shows whether a word's orthographic neighbours are also neighbours of each other)		Words with high clustering coefficients were **less quickly** recognised		
Closeness centrality (The inverse of the average distance between the word	Words with high closeness centralities were	Words with high closeness centralities were	Words with high closeness centralities were	

and other words in the network)	**more slowly** named	**more quickly** recognized	**less accurately** named	

Table 2 Orthographic form and word recognition (based on Siew 2018)

Some of Siew's findings may be summarized thus:

- The notion of degree is an expression of how many orthographic neighbours a word has in the network. An orthographic neighbour is a word that differs from the original word by only one letter. The results show that high-degree words are more easily recognized than low-degree words.
- The clustering coefficient is an expression of whether a word's orthographic neighbours are also neighbours of each other. Siew did not emphasize these results in her study.
- Closeness centrality is an expression of how central a word is in the network (how many links it takes to connect the word to all other words in the network). The results show that words with high closeness centrality were more quickly recognized in the lexical decision task, but they were more slowly and less accurately named in the speeded naming task. Siew explained the latter finding with the fact that words with a high closeness centrality have many words that resemble them and that they thus may "compete" with each other in the naming task. That degree and closeness centrality do not correlate with each other in this case may depend on the tasks used. Siew calls for more research in this field.

Even though Siew's work has been done on English language materials, the overall results seem to support the assumption that short, high-frequent Norwegian and English cognate words are easy to process. It is interesting to note Siew's observation that "[f]requent words tend to be short words that also tend to have several phonological and orthographic neighbours in the language" (p. 7). The Germanic-based words that may be used to facilitate Scandinavian students' learning of vocational terms are as a rule short and high-frequent. It could be interesting to investigate if the Germanic-based words in English are more high-degree than the Latin-derived words.

2.2.5 Word frequency and coverage

For people who are learning another language, "some words are much more useful than others" (Nation 2013, p. 9). This view is based on studies of word frequencies. Words like *the, a, and, said, he, she, I, me, was, want, with, who, house, must, went* occur very frequently

in texts. Learners will be able to say quite a lot, even with a small vocabulary, if they know the most frequent words (Nation 2013, p. 14).

Creating a system for determining word frequency, Nation used the British National Corpus to make 14 lists which each contain thousand word families. (A word family includes the inflected and derived forms of the base word, e.g. *map, maps, mapping*). The first list contains the thousand most high-frequent word families, the second list contains the next thousand word families, etc. In addition to the 14 lists there are three lists containing proper nouns, marginal words and compounds (see Table 3). The word family lists have been divided into high-frequency, middle-frequency and low-frequency word families. The high-frequent words cover a large part of all kinds of texts, and they are therefore crucial to learning. Nation included 2,000 word families in this category, while other researchers have suggested 3,000. Mid-frequency vocabulary occurs often enough to be worthwhile learning. That category includes the next 7,000 word families. There are large amounts of low-frequency words in the language. Nation recommends teaching them deliberately only when it is necessary to understand a text, and to otherwise leave them to incidental vocabulary learning (Nation 2013, pp. 20-29). See some selected examples of words in the different lists in Table 3.

Category of words	**Coverage**	**Word families**	**Examples** (the first 5 word families in each list)
High-frequency words (2,000 word families)	90%*	1st thousand	a, able, about, absolute, accept
		2nd thousand	above, abuse, accent, access, accident
Mid-frequency words (7,000 word families)	9%	3rd thousand	abbey, abroad, absence, accelerate, accordingly
		7th thousand	abate, abolition, abscess, abstention, absurd
		9th thousand	aback, abattoir, abdomen, aberration, abode
Low-frequency words (ca. 50,000 words)	1%	10th thousand	abet, abeyance, abject, AC, acclimatise
		12th thousand	aberrant, ablaze, abominate, abrogate, absolution
		14th thousand	abbess, ablation, abrasion, acanthus, actin
Proper nouns		'basewrd15'	Kwazulu, Firestone, Glynis, Hythe, Viner
Marginal words		'basewrd16'	aahs, ahem, argh, arrgh, urgh
Compounds		-	forever, aftershave, ashtray

*Table 3 Examples of words with different frequencies. * The 90% coverage of high-frequency words includes proper nouns etc.*

The examples in Table 3 give the impression that the least frequent words are predominantly of French/Latin/Greek origin, although there are exceptions. *Aback* and *abode* in the 9th thousand are of Old English origin. Some important words in vocational English are listed among the high-frequent words. The abbreviation AC (*alternating current*) in the 10th thousand is very frequently used in the electrical trades. Even in the 14th thousand, there

occurs a word which is concrete and will occur in descriptions of safety precautions in vocational texts: *abrasion*. It is of Latin origin (Merriam-Webster 2019).

How many words must learners know to understand a text sufficiently well? This factor is called coverage. Lexical coverage is "the percentage of running words in the text known by the reader" (Nation 2006, p. 61). Researchers - among them Nation, Hu and Schmitt - have given different estimates, but generally it is considered that a coverage of between 95% and 98% suffices for a learner to understand and be able to work with a text. 95% coverage means that one in every 20 words is unknown to the reader. A coverage of 98% means that one in every 50 words is unknown. 95% is considered to be the lower boundary for a lexical threshold, giving the learner a minimal comprehension of the text (Nation 2013, p. 14). Depending on the text genre, learners need to know about 4,000-5,000 word families to reach this lower threshold. 98% is the upper threshold, providing an optimal comprehension. Learners need to know between 6,000 to 9,000 word families to reach the upper threshold (Wenhua Hsu 2011, p. 248). If the coverage reaches 100% the learner understands every word in the text, so that reading it no longer entails any vocabulary learning effect. As mentioned, different genres have different degrees of difficulty. Figure 1 shows which vocabulary sizes a learner needs to have to attain a coverage of 95% and 98% within some text genres.

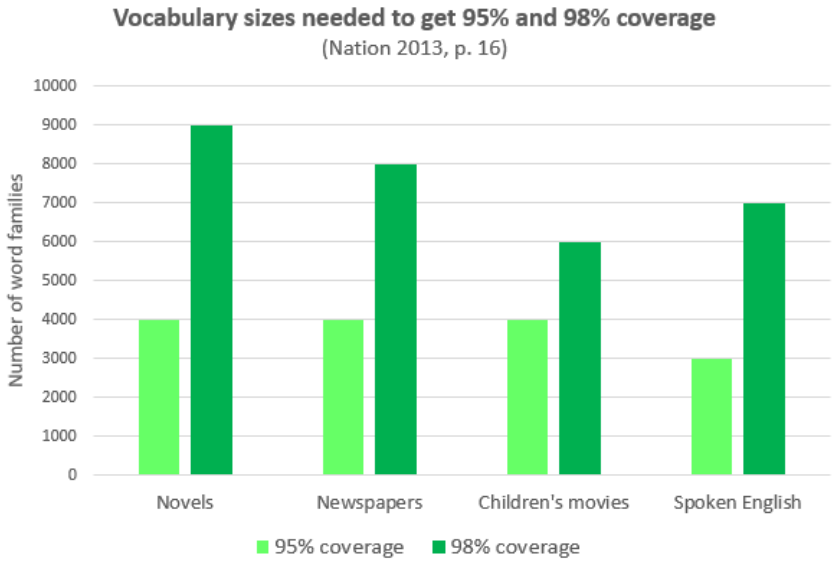


Figure 1 Coverage and genres. Based on Nation 2013, page 16.

Webb and Nation developed a manual of how to determine the vocabulary load of written texts (Webb & Nation 2008). Using the RANGE software, researchers and teachers can both assess the degree of difficulty in a whole text, and identify individual low-frequent words that may be challenging for students. RANGE, which was used to develop the Academic Word List, contains a set of 20 lists from the British National Corpus (BNC) with thousand words each. When analysing the text, RANGE indicates how many per cent of the words in the text belong in the first family of thousand words ("Base 1"), how many per cent belong in Base 2 etc. This gives a reliable measure of the degree of difficulty.

Figure 2 gives an impression of which coverage each of the BNC lists has. The high-frequency words in the first thousand-word list are ubiquitous, so they represent a very high vocabulary coverage. The words in the second, third, fourth lists etc. are less common, so they represent a smaller coverage. Low-frequency words in e.g. the "15th thousand" list, are still much less commonly used, and they therefore cover a very small part of the texts in the BNC. The cumulative coverage is the sum of the coverage of individual lists - e.g. the sum of the first four lists added together; which gives a coverage of a little more than 95%. The thresholds of 95% and 98% coverage, which are important in discussion of understanding texts, have been indicated in Figure 2. The word "token" stands for a single word in the corpus.

Webb and Nation have also developed the Vocabulary Size Test, which enables teachers to test the vocabulary sizes of their students. When knowing the average vocabulary size of a class, and taking into account the vocabulary sizes of individual students, teachers can use RANGE to determine whether a written text has a degree of difficulty that matches the level of the students. RANGE also identifies the level of difficulty of individual words, by indicating which of the thousand-word lists they belong to (Webb & Nation 2008, pp. 5-17).

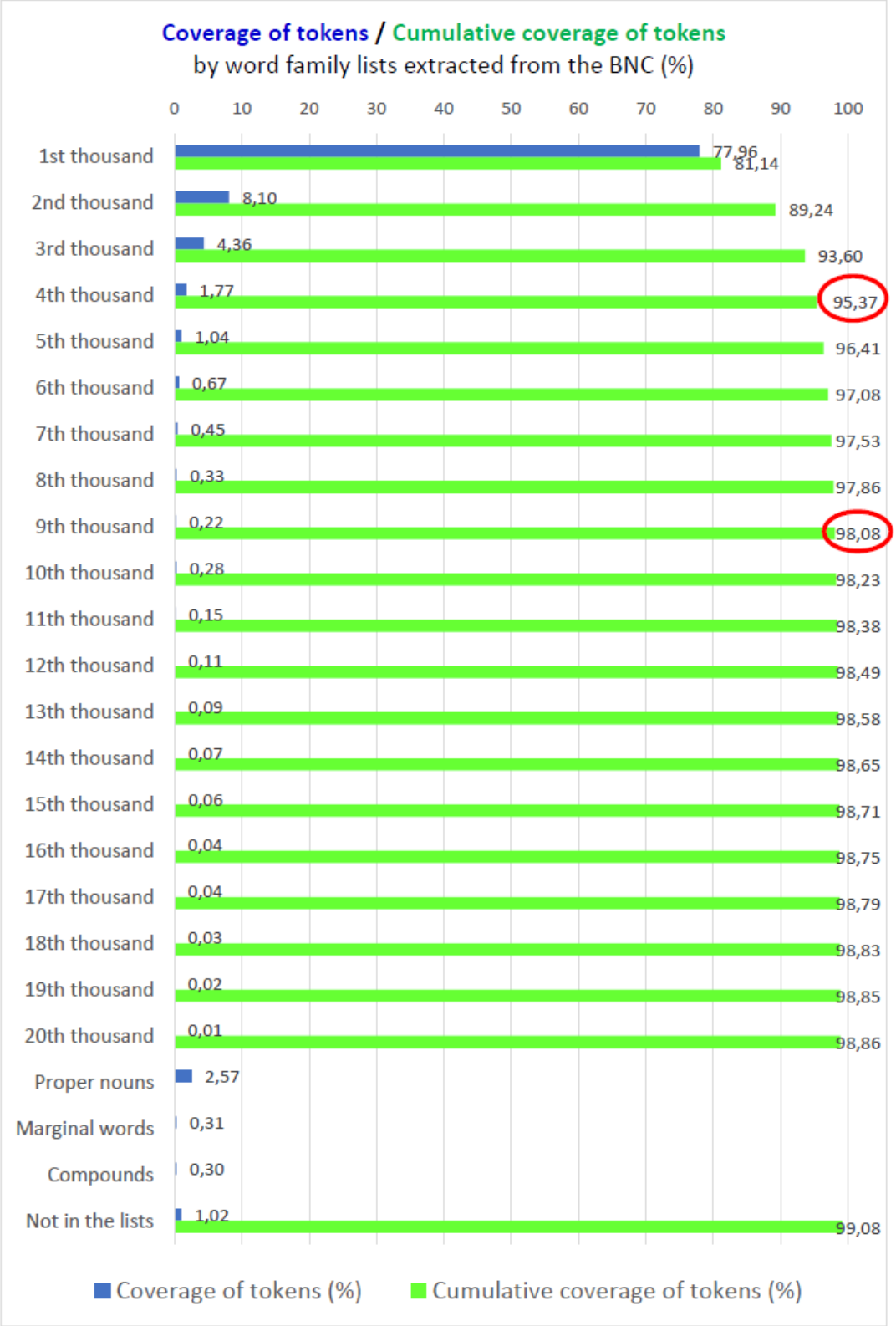


Figure 2 Percent coverage of tokens by word family lists made from the BNC (blue), and percent cumulative coverage (green). Based on Nation 2013, p. 21.

2.2.6 The number of syllables

Another factor that influences comprehension is the number of syllables in the words (syllable count). Fabienne Chetail (2014, p. 1249) investigated how the number of syllables influences word recognition. She found that in French, three-syllable words are processed more slowly than two-syllable words. This is true both for pseudowords and words of high and low frequency (pseudowords are artificial, meaningless words that follow the phonotactic rules of the language). At first, there was created an impression that this retardation effect occurs only for low-frequency English words. Some researchers attributed the retardation effect not to the number of syllables, but to the fact that English words have little consistency between the pronounciation and the orthography of vowels. That discrepancy might also confuse readers and slow down their processing. More recent research has, according to Chetail (2014, p. 1253), concluded that three-syllable words are recognised more slowly than two-syllable words in English, although this effect is more pronounced for low-frequency words than high-frequency words.

Eric Lambert et al. (2007) carried out experiments with handwriting in French which may shed light on why the number of syllables affect production. One reason for this seems to be that there is a limit for the amount of sequences of syllables that can be kept in memory simultaneously. When the sequences grow too many, some of them disappear from memory and must be retrieved anew, which slows down the handwriting process. Thus there seems to be indications that an increased number of syllables creates difficulties for both perception and production of written text.

2.2.7 Readability indexes and aspects of readability

As mentioned in 2.2 above, a number of different readability formulas and indexes have been developed over the years. Many of them were created before computers came in common use. Therefore they apply variables that are openly observable and quantifiable, like word length and sentence length. Table 4 gives an overview of some of the most central formulas / indexes and which variables they rely on. Note that some of them are made for specific uses, like the FORCAST formula for the army (DuBay 2004, p. 47).

Name of the index	Year	Variables used
"Classic" formulas		

Flesch Reading Ease score (1st part)	1948	Sentence length Number of syllables per word
New Reading Ease Score	1951	Sentence length Percentage of one-syllable words
Dale and Chall formula (original version)	1948	Sentence length Percentage of easy words (on the Dale-Chall 3,000-word list)
Fog Index	1952	Sentence length Percentage of words with two syllables or more
"New" formulas		
Coleman formulas	1965	Number of sentences per 100 words Percentage of one-syllable words Percentage of pronouns Percentage of prepositions
Bormouth Mean Cloze formula	1969	Letters per word Words per sentence Share of easy words (from the Dale-Chall list)
The Fry Readability Graph	1968	Number of sentences per 100 words Number of syllables per 100 words
The SMOG formula	1969	Number of polysyllabic words (more than two syllables)
The FORCAST formula	1973	Percentage of single-syllable words
ARI	1967	Words per sentence (Typewriter) strokes per word
NRI	1976	(A combination of formulas)
Flesch-Kincaid formula	1976	Sentence length Number of syllables per word
Hull formula	1979	Sentence length Percentage of prenomial modifiers

Table 4 Some of the traditonal readability indexes and the variables each of them used (DuBay 2004, pp. 21-50)

One of the most popular readability formulas is the Flesch Reading Ease score (DuBay 2004, p. 21). Its exact formula is:

Score = 206.835 – (1.015 x ASL) – (84.6 x ASW)

In the formula, ASL stands for "average sentence length" , i.e. the number of words divided by the number of sentences. ASW is the "average number of syllables" per word. The score will appear as a figure in the range of 0-100, where 100 is the easiest possible level. Texts with a score of 70 are suitable for adults, while the level of 30 is very difficult. (This is in fact only the first part of the formula; the second part aims at predicting the human interest in a text by quantifying pronouns, names, exclamations etc.).

Later the Flesch Reading Ease formula was modified further:

New Reading Ease Score = 1.599 nosw – 1.015 sl – 31.517

The abbreviation “nosw” stands for the “number of one syllable-words” per 100 words, while “sl” stands for the average sentence length, measured in the number or words. This new variety of the formula was adjusted to display its scores in grade levels, and was renamed the Flesch-Kincaid formula. Texts that are tested with this formula may e.g. be labeled as “suitable for 9th graders”. The factors in the two formulas were adapted so that their results correlated with McCall-Crabbs reading tests. The Flesch Reading Ease scale and the Flesch-Kincaid grade level formula have been integrated into Microsoft Word as tools that users can apply to test and adjust the difficulty of their texts while writing.

Quoting previous research in the field, DuBay (2004, p. 18) lists a wide range of different variables that have been considered relevant to include in readability formulas / indexes. William S. Gray and Bernice E. Leary published a work as early as in 1935, where they focused on the “style” aspect of readability. On the sentence level they included sentence length, the number of sentences per paragraph and the number of simple sentences. On the lexical level they pointed out the percentage of words that are easy (or difficult / unknown) for the reader group, the percentage of different words, the percentage of monosyllables (or polysyllables) and the number of personal pronouns (DuBay 2004, p. 16-19).

Much later, in his 1976 study (quoted by DuBay 2004, p. 38), George R. Klare listed a number of more abstract properties that may affect the readability of sentences. Is the sentence (or clause) long or short? Is the sentence active or passive? Affirmative or negative? Embedded or not? Simple or complex? Likewise, a row of new lexical properties was highlighted: Is it a content or function word? Frequent or not? Familiar or not? Long or short? Concrete or abstract? What do the readers associate with the word? Is the verb construction active? Klare’s properties clearly seem to enrichen the quality of a readability index, but some of them must be difficult to include even in computerised indexes.

Through time, the work with readability indexes has had considerable success but also faced criticism. Shesen Guo (2011, p. E103) has listed some limitations: The formulas rely too much on the quantifiable features of the text, while the readers’ prior knowledge and interests are not taken into account. Nonetheless, Guo stated that readability formulas give valuable information, are trustworthy and “are becomig more popular than ever”. According to DuBay (p. 19), the readability researchers themselves have been well aware of the limitations. They have urged users not to apply formulas mechanically but use them as “rough guides” along with other aspects of good writing.

2.3 Language relatedness and its relevance for Nordic students

Most of the languages in Europe, together with many Asiatic languages, have their common origin in the Proto-Indo-European language. Thus, English shares a common inheritance with many other languages. In addition to that, English has been influenced by other languages in the course of history, most notably by Old Norse, French, Latin and Greek. For educators in the Nordic countries it must be interesting to see how the language relatedness may help to facilitate the learning process.

2.3.1 The historical relationship between Norwegian and English

English and Norwegian are closely related languages. Both belong to the Germanic branch of the Indo-European language family. The Germanic languages of today (see Figure 3) consist of two branches: The North Germanic (Nordic) branch, comprising Danish, Swedish, Norwegian, Faroese and Icelandic – and the West Germanic branch, comprising German, Dutch, Frisian and English (James P. Mallory and Douglas Q. Adams 2006, pp. 22-23). English and Norwegian are historically related in two different ways. Firstly, they share a common origin, having inherited many linguistic features from their Germanic ancestor. Secondly, the invasions and settlement by Norwegian and Danish Vikings in the period between 793 and 1066 created a language contact situation where the English language was significantly influenced by the Nordic languages (Baugh & Cable 2013, pp. 87-93) .

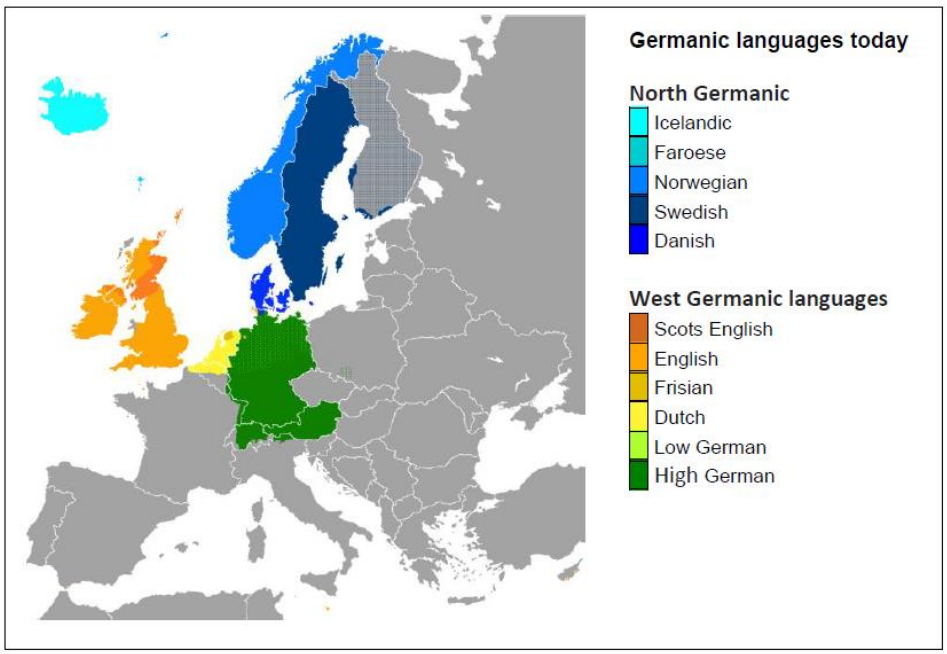


Figure 3 Germanic languages today.
Illustration by lenguas_germanicas.PNG on Wikimedia Commons

In the Viking Age people from mainly Norway and Denmark attacked the British Isles, many of them to settle permanently. Norwegian Vikings settled mostly in North East Scotland, while Danish Vikings established the "Danelaw" in the east of England. This movement of people had a great impact on the language situation in Britain (see Figure 4 for a map showing the distribution of languages in the early 10th century). According to Sandra Dögg Friðriksdóttir (2014) there are approximately 400 borrowings from Old Norse in the Standard English of today, and they are among the most frequently used words in English. Elly van Gelderen (2014, p. 102) mentions a far bigger number of 1,000 loans, stating that "[t]he influence of Scandinavian on the vocabulary of English is substantial". Some of the loanwords listed by van Gelderen are *anger, both, call, egg, get, give, guess, ill, mistake, nag, odd, rot, rugged, same, scrape, seem, sister, skill, sky, take, want, weak, window* and *wrong*. A possible reason for the diverging estimates may be the fact that Old English and Old Norse were closely related languages sharing a common origin, which may make it difficult to determine the etymological origin of every word. The Vikings added words to English that were of the same type as the already existing Anglo-Saxon terms: short, simple and concrete. Both Old English and Old Norse were syntactically complicated languages, but the language contact started the process of simplification that has led to the much simpler syntax that the English language has today (van Gelderen 2014, p. 103).

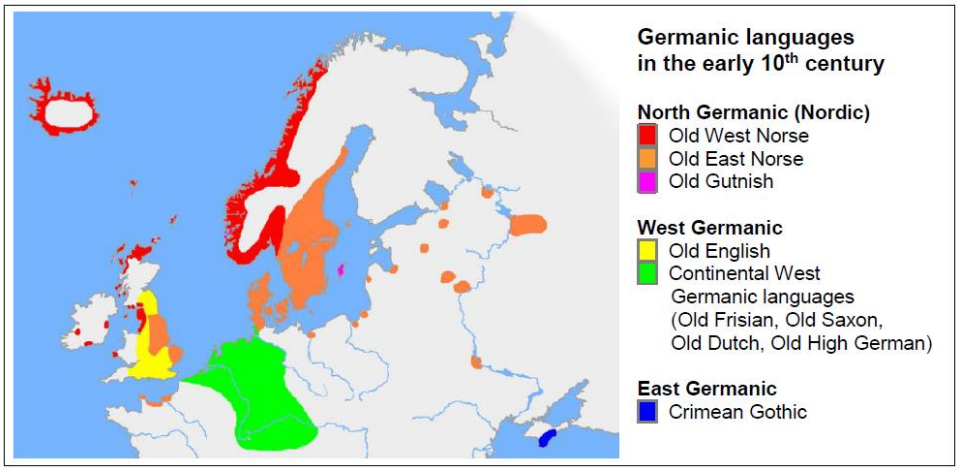


Figure 4 Germanic languages in the early 10th century. The areas in Britain where Vikings had settled are indicated with the same colour as the home countries of the settlers. Illustration by Wiglaf on Wikimedia Commons.

The Old Norse loanwords made the English vocabulary more nuanced. Dennis Freeborn (2006, p. 54) mentions that the initial consonant cluster sk- had gone into disuse in Old English. Many of the words beginning with sk- in modern English, like *to skip, skin, sky*, were introduced from Old Norse. In some cases, the Old Norse words were added to the English vocabulary rather than replacing words of Anglo-Saxon origin. Arna Rún Sésarsdottir (2015,

pp. 14-16) has explored the notion of "doublets" in English. A pair of such doublets consists of one Anglo-Saxon word and one word that was introduced from Old Norse. Examples: *heaven/sky, egg/edge, hide/skin, skirt/shirt, craft/skill, no/nay, from/fro, rear/raise, shatter/scatter*. Each word in the pairs took on slightly different meanings, thereby enriching the vocabulary. This relatedness in vocabulary makes the Norse-derived fraction of the English suitable for making the general language in vocational texts reader-friendly. The words that originate from Old Norse are short, simple, and often cognates with modern Norwegian words.

2.3.2. Linguistic purism – a German and Nordic tradition

In Germany and the Nordic countries there has been a certain resistance against the adoption of foreign loanwords. Starting in the 16th century, German linguists felt that the great influx of Latin and French imported words was alienating, and they feared that it would pose a hindrance for the education of ordinary people who knew only the vernacular. They began constructing German equivalents that they hoped would replace the foreign words. Such attempts at "cleansing" the mother tongue of foreign influences is called purism. The most active German purist was J. Heinrich Campe. Some of the words he constructed have been used in German to this day, while others were ridiculed by the contemporary intellectuals and soon forgotten (see Astrid Stedje 1989, pp. 133 and 150-151).

The idea of linguistic purism spread to Denmark and by extension to Norway, which was governed by Denmark at that time. The Danish linguists Jens Sneedorff and Frederik Eilschov constructed new "purely" Danish words, many of which followed the German pattern (see Ernst Håkon Jahr 1987, p. 61). That the Danish purists fetched inspiration from the Germans can be seen from the examples in Table 5. After Norway's union with Denmark was dissolved in 1814, both Knud Knudsen (the founder of the Bokmål variety of Norwegian) and Ivar Aasen (the founder of the Nynorsk variety) made active attempts of ridding the Norwegian language of foreign influences (Helge Sandøy 2004, pp. 131-132). In our time, Norwegian purists have tried to limit the impact of English influences. Even though the efforts of the German / Danish / Norwegian purists have been only moderately successful, the Norwegian language has numerous "home-grown" words where other languages have Latin and French-derived words. List 3 in Table 1 provides further examples of this.

French / Latin	German replacement form	Danish	Norwegian (Nynorsk)	English
poet	**Dichter**	**dikter**	**Diktar**	poet
autor	**Verfasser**	**forfatter**	**Forfattar**	author
passion	**Leidenschaft**	**lidenskab**	**Lidenskap**	passion
verosimilis	**wahrscheinlich**	**sandsynlig**	**Sannsynleg**	probable

Table 5 Linguistic purism: Examples of French / Latin words and their constructed equivalents

When the influx of French, Latin and Greek words into English was at its greatest around the middle of the 16th century, several English intellectuals objected to what they called "inkhorn" words. Complaining that the new words made the language obscure, they saw them as a mere expression of pedantry. Among these purists were Sir John Cheke and Thomas Wilson. The imported words had become so numerous, however, that a great amount of them were adopted into the English language on a lasting basis (Baugh & Cable 2013, pp. 215-220). Thus, the English language of today contains more French, Latin and Greek vocabulary than the German, Danish and Norwegian languages do. This situation is a good reason for adapting the general English of vocational texts, so that French-, Latin- and Greek-derived words do not unnecessarily hamper comprehension for Norwegian students.

2.3.3 French, Latin and Greek influence on English

A very significant part of the English vocabulary has been derived from the Romance languages French and Latin, as well as from Greek. Many words of these origins may create problems for Norwegian vocational learners, both because they are unrelated to Norwegian and because they belong to an academic language stratum that the students are not familiar with. On the other hand, some words of French, Latin and Greek origin are short and simple, and are well known to all English speakers.

The Norman French invasion of England in 1066 created an entirely new language situation. After William the Conqueror won England, the new ruling class there spoke French, while common people continued speaking English. In the Icelandic saga of Gunnlaug Ormstunge it is stated that the Vikings and the English could understand each other before the Norman invasion:

> … in those days there was the same tongue in England as in Norway and Denmark; but the tongues changed when William the Bastard won England, for thenceforward French went current there, for he was of French kin. (Morris and Magnusson 1901, Chapter 7).

French belongs to the Romance group of languages, which have all developed from Latin in areas that once belonged to the Roman Empire. The new rulers of England hailed from Normandy in Northern France (see Figure 5), and they therefore spoke a variety of French that differed from Paris French. The Norman French rulers of England kept their land areas in Northern France for a time. After they lost those possessions, however, contact with their old home country diminished and their use of French declined. After a couple of hundred years, English became the dominating language of England again, not only among the common people but also among the royals and nobility (van Gelderen 2014, pp. 104-106).

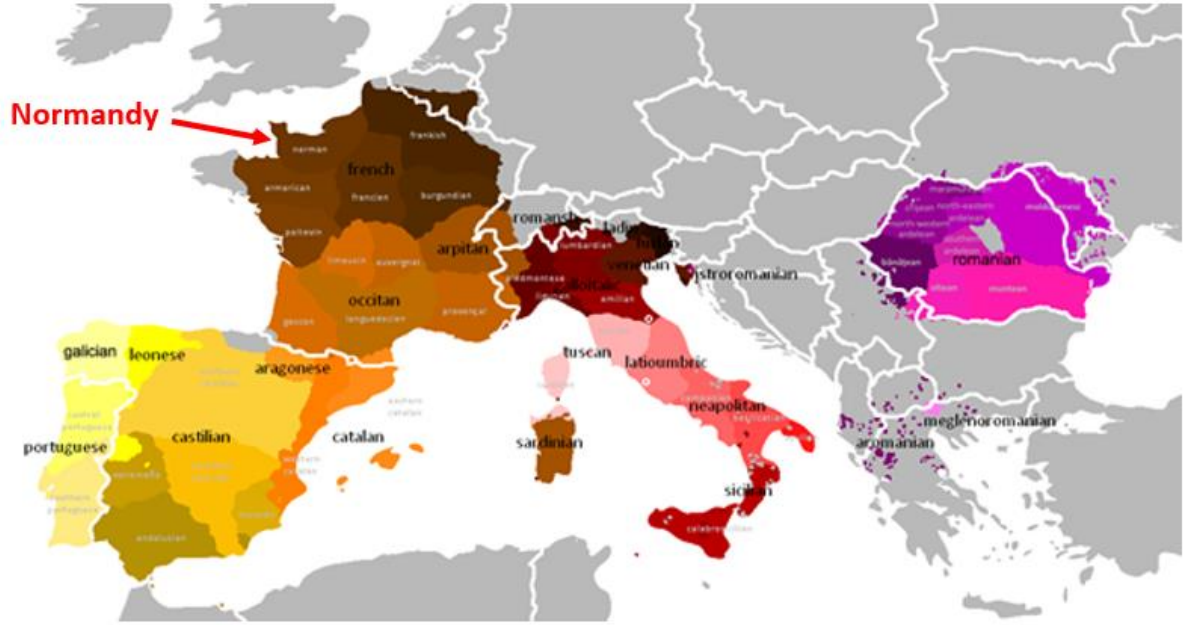


Figure 5 Map of the Romance languages, which have developed from Latin. In addition to the major ("national") languages Portuguese, Spanish, French, Italian and Romanian, regional varieties and dialects are also indicated on the map. In 1066, Norman French forces from Normandy in the north of France conquered England. Illustration by Servitje on Wikimedia Commons

The English language had now undergone radical changes. Many Latin words have come into English through its "daughter language" French. During the Renaissance and Enlightenment, a great number of Latin and Greek words were added to English, typically within fields like science, art and culture. Some researchers estimate that nearly 27,000 new words were introduced into English between 1500 and 1660, while others have arrived at a lower estimate of 10,000 words. Many, but not all, were foreign loans. In addition to actual Latin and Greek loanwords, many new words were created from Greek and Latin word elements like *re-, anti-, meta- and extra-* (van Gelderen 2014, pp. 179-181). There has later been a "paradigm shift"

where Latin has lost its importance and been replaced by English as the dominating language. Nevertheless, Latin has set its mark on the vocabulary of English and other languages. This linguistic heritage may be an advantage for speakers of Romance languages today, while it poses challenges for learners of English in Germanic-speaking countries.

Many of the new, intellectual loanwords belong to a different register than the words that are derived from Anglo-Saxon and Old Norse. In a discussion of productive and receptive vocabulary, Nation (2013, p. 48) quotes the argument of Corson that "Graeco-Latin" words only belong to the receptive vocabulary of many people, rather than being used productively. Graeco-Latin words are generally low-frequency, they have a structure that is "opaque" to some people, and some learners have a social background that has given them little opportunity to learn such words. The Graeco-Roman words constitute a part of what Corson calls the "lexical bar" (lexical barrier). The concept of the lexical bar reflects the fact that students must acquire a specialised academic vocabulary to succeed. In English-speaking countries, the academic vocabulary contains a high proportion of Graeco-Latin words which may make the studies unnecessarily difficult. Confirming Corson's impression, Nation (2013, p. 297) states that more than 90% of the words in the Academic Word List are Graeco-Latin. Around 80% of the words in the Academic Word List are cognate with Spanish words, which is not surprising considering that Spanish is a Romance language, descended from Latin. This language relationship makes it easier for students with a Romance language background to comprehend the academic vocabulary of English, while students with Germanic mother tongues, like Norwegian, will not benefit from the same recognition effect.

To determine how closely related two languages are to each other, a Swadesh list is often applied. A Swadesh list is a collection of words thought to express basic and universal concepts. The original Swadesh list, compiled by the US American linguist Morris Swadesh, contains 215 words. The Russian linguist Sergei Yakhontov selected 35 of those words to make a more compact list (World eBook Library 2019). If the Swadesh-Yankhtonov list is used to compare English, Norwegian and French, there are clearly most resemblances between English and Norwegian. On an intuitive basis, one finds around 16 English-Norwegian word pairs that "sound" similar to each other, while only three Norwegian-French pairs appear to be similar. Four of the words are – to a greater or lesser extent – similar in all the three languages. This simple comparison reinforces the impression that the Germanic-based part of the English vocabulary may be easy to comprehend for Norwegians.

Swadesh number	**English**	**Norwegian** (nynorsk)	**French**	Comment
1	I	eg	je	
2	you	du	tu	*You* in the singular form
7	this	denne	ce, cela	
11	who	kven	qui	
12	what	kva	quel	
22	one	ein	un	
23	two	to	deux	
45	fish	fisk	poisson	
47	dog (hound)	hund	chien	Eng. *hound* = hunting dog
48	louse	lus	pou	
64	blood	blod	sang	
65	bone	bein	os	
67	egg	egg	œuf	
68	horn	horn	corne	
69	tail	hale	queue	
73	ear	øyre	oreille	
74	eye	auge	œil	
75	nose	nase	nez	
77	tooth	tann	dent	
78	tongue	tunge	langue	
83	hand	hand	main	
103	know	vite, kjenne	savoir, connaître	
109	die	døy	mourir	
128	give	gje	donner	
147	sun	sol	soleil	
148	moon	måne	lune	
150	water	vatn	eau	
155	salt	salt	sel	
156	stone	stein	pierre	
163	wind	vind	vent	
167	fire	eld	feu	Norw. å fyre = to ignite
179	year	år	année	
182	full	full	plein	
183	new	ny	noveau	
207	name	namn	nom	

Table 6 Swadesh-Yankhtonov list of English, Norwegian and French. The words have the same numbers as in Swadesh's original 215-word list. Words that appear to be similar have been given the same colour.

2.4 The lexical bar

David Corson (1995) addresses the phenomenon of the "lexical bar", and observes that the English vocabulary is divided into two sections: Anglo-Saxon words that are typically short and simple and much used in everyday situations, and Graeco-Latin (Latinate) words that are typically long and sound learned and foreign. Corson does not see the Graeco-Latin words as redundant; rather they add nuances to the language (p. 90), but an overuse of such words may

cause problems for learners. Native speakers of English learn the Graeco-Latin words relatively late, through formal education, and it is generally these words that pose problems both for them and for learners of English as an L2. Corson argues that there is a mismatch between children's everyday discourse and the high-status language demands that the schools impose on them, and that this discrepancy makes many students struggle unnecessarily with their education (p. 1). Corson (pp. 81-82) points out that the lexical bar is not caused only by differences in vocabulary that stem from historical developments. There are both linguistic and social / intrapersonal factors in the language that maintain the effects of the lexical bar today.

Graeco-Latin words tend to be longer and differently structured than Anglo-Saxon words (p. 84). Their vowel-consonant and consonant-consonant structures are different from Anglo-Saxon words, and their written forms contain unfamiliar digraphs like *ph*, *pn* and *ps* (e.g. in *phonetic*, *pneumatic* and *psalter*). Latin and Greek affixes like *agri-, graph-, hemi-, intra-, meta-, quasi-, terti-* and *-logue* also contribute to giving these words a foreign look. Graeco-Latin words are more often encountered when reading, as opposed to when speaking or watching television for example (p. 88).

Corson perceives the divide between everyday Anglo-Saxon vocabulary and the more elevated use of Graeco-Latin words as a parallel to Pierre Bourdieu's concept of cultural capital (p. 4). Surveys have shown that readers experience Graeco-Latin words as "much more formal" than Anglo-Saxon words, regardless of whether the words are of high or low frequency (p. 87). Many learners from a working-class background are unfamiliar with the Graeco-Latin words. They have difficulties relating to them and learning them, and they often perceive them as formal, posh and/or ridiculous (p. 86). When students from more modest backgrounds come to school, they experience that only the "elite" vocabulary is praised and respected there, while their own linguistic skills are disregarded. The "linguistic capital" of working-class learners is not sufficiently appreciated (p. 95), which may bar many of them from developing their talents to the full potential. To which extent Norwegian students feel alienated to formal registers of English is difficult to know, but there are tendencies that Norwegians also react to and comment "snobbish" language use.

Corson (pp. 172-174) found that Graeco-Latin words have a much lower "imagery level" than Anglo-Saxon words. "Imagery" is the degree to which a word arouses sensory images. It is possible to define individual words' position on a "imagery scale" by asking subjects to rate to which extent the word expresses sights, sounds, smells or other emotions / sensations.

Research has shown that words with a high "imagery level" are concrete, easy to associate with other words, and easy to learn.

Graeco-Latin words often occur in formal and specialized contexts, and in their written form rather than in speech. This makes those words less accessible to learners (p. 174). Graeco-Latin words with affixes can be stored as whole-word entities in the mental lexicon of learners who have little knowledge of the affix system, while more experienced language users can recognize the affixes and thus process the words more efficiently (p. 175). The meaning of many Graeco-Latin compound words may have been obvious to the people who introduced them into the English language, but this meaning is lost to most present-day language users (p. 176). The different parts of the word do not prompt separate semantic activations in the mental lexicon.

As a result of developments outlined in subchapter 2.3.2, most of the Germanic languages have a tendency towards semantic transparency. For instance, the Norwegian compound word *tannlege* can be understood by anyone who knows the meaning of its elements *tann* (tooth) and *lege* (doctor). The same is the case for *Zahnarzt* in German, *tandarts* in Dutch and *tannlæknir* in Icelandic. Many of the "Graeco-Latin" words in English, however, are opaque and cannot be deciphered in the same way. The English word *dentist* cannot be understood by a learner who knows the words *tooth* and *doctor* and those three words thus have to be learned separately. Corson (p. 177) found that a large majority of Anglo-Saxon words are semantically transparent. Some Anglo-Saxon words have become opaque through language development, but many of them (e.g. *honey, hungry, kitchen*) are learnt early in life and do not pose any big challenge for learners.

According to Corson (p. 178) there is psycholinguistic evidence that "transparent Anglo-Saxon compounds are activated more efficiently than complex [Graeco-Latin] words". Corson supports this claim by stating that the speed of activating a compound (e.g. *seaweed*) depends on the frequency of the first word in the compound (*sea*). Complex Graeco-Latin words of the same length must be retrieved from the mental lexicon as whole words, and they generally have a lower frequency.

2.5 Summary

This literature review has focused on the problem that vocational students can face a heavy vocabulary burden when working with vocational texts. This burden can be lessened by

simplifying the general vocabulary of the texts, while retaining the vocational terms. This simplification of the general language can be carried out partly by applying high-frequent words with few syllables, and partly by using words that are cognates with the students' L1. For Nordic students, using words of Anglo-Saxon and Nordic origins, rather than words of French, Latin and Greek origin, can enhance comprehension. Having some knowledge of language history will enhance the understanding of which challenges Norwegian students encounter when learning English, and make it easier for educators to adapt the English vocabulary to different situations.

3. Method and materials

In order to map different properties of the vocabulary of vocational students, a diagnostic vocabulary test was developed and carried out. The aim of the test was to establish an impression of which categories of vocabulary vocational students normally know. As a follow-up of the diagnostic vocabulary test, the participating students were asked to fill in a survey which charted their attitudes toward learning different categories of words. The results of both of these instruments were used to form a readability index which may help Norwegian educators select the English vocabulary which suits the best for different purposes, e.g. in vocational texts.

3.1 Diagnostic vocabulary test

The aim of this survey was to test students' understanding of words of different frequencies and origins. The students were simply supposed to translate the words from English to Norwegian. The survey was carried out at the beginning of the school year 2021/22 at a vocational upper-secondary school with several different vocational programmes.

Almost all of the Vg1 classes at the school took part. The rather large number of 150 respondents, and the mostly manual method for summarising the results, made it laborious to summarise the survey, but on the other hand the sizeable number of respondents has strengthened its validity.

40 words were included in the survey. The selection of words consisted of ten lexical sets of synonyms or near-synonyms, each of which consisted of four words of differing frequency.

The survey form that as handed out can be seen in Figure 6. The words in the sets were as follows:

Lexical set no. 1:	immense – large – gargantuan - huge
Lexical set no. 2:	worn out – fatigued – tired - exhausted
Lexical set no. 3:	multiple – many – numerous - a lot
Lexical set no. 4:	to continue - to go ahead - to proceed - to carry on
Lexical set no. 5:	rich – prosperous - well-to-do - affluent
Lexical set no. 6:	to postpone - to put off - to delay - to defer
Lexical set no. 7:	to begin - to commence - to start - to launch
Lexical set no. 8:	tranquil – quiet – placid - calm
Lexical set no. 9	small – minuscule – tiny - diminutive
Lexical set no. 10:	terrific – wonderful – glorious - marvellous

The lexical sets were built up of synonyms or near-synonyms, so that the results can give an indication of how many frequent or non-frequent synonyms the students know within each set. To pick out synonyms, dictionaries were used, in addition to browsing various sets of synonyms published online. An online thesaurus at https://www.thesaurus.com/ was also used. The survey form is shown in Figure 6. The 40 words are distributed in a grid of five by eight squares. In this way the lexical sets were separated from each other in a haphazard pattern.

In the selection of words for the survey, care was taken to avoid unnecessary misunderstandings due to words having multiple meanings. In lexical set no. 9, it was first considered to include the word minute, which can mean "(very) small" like the other words in the set, but it was abandoned since it also has the much more well-known meaning of "60 seconds". Nonetheless, by accident one word with multiple meanings was included, namely terrific. Below follows the answer form for the diagnostic vocabulary test (Figure 6).

DIAGNOSTISK TEST I SKULESTARTEN – ENGELSK ORDFORRÅD

Namn: _____ Klasse: _____

Dette er ein test for å finne ut kva slags ord ein bør satse på å lære.

Skriv den norske omsetjinga av ordet, på denne måten:

example
eksempel

Ikkje bruk ordbok eller andre hjelpemiddel.

a lot	affluent	to begin	calm	to carry on
to commence	to continue	to defer	to delay	diminutive
exhausted	fatigued	gargantuan	glorious	to go ahead
huge	immense	large	to launch	many
marvellous	minuscule	multiple	numerous	placid
to postpone	to proceed	prosperous	to put off	quiet
rich	small	to start	terrific	tiny
tired	tranquil	well-to-do	wonderful	worn out

Figure 6 The answer form that was used in the diagnostic vocabulary test

Procedure and participants

The forms for the diagnostic vocabulary test were distributed to the Vg1 students (15 to 16-year-olds) in the first weeks of the school start of 2021. The distribution of the test was discussed in advance at a meeting of the school's English teachers. It was agreed that each teacher should inform their students that the test was voluntary, and that it was possible to choose alternative activities instead. The students wrote their names on their form so that they

could receive individual feedback, but the results were anonymised in the further treatment of the data.

The individual students received a feedback form in Excel (see the example in Appendix 2). The form showed the student's total number of correct translations, and also the percentage of correct translations in different frequency classes. The feedback form also contained links to sites with information of vocabulary learning and practices.

AntWordProfiler

When developing and interpreting the results of the diagnostic vocabulary test, it was necessary to determine the frequency of each of the 40 words from the test and determine which frequency classes they belong to. AntWordProfiler (Anthony 2021) is a computer programme developed by the linguist Laurence Anthony. When AntWordprofiler is used together with BNC/COCA frequency lists (Nation 2021) it can identify which frequency class individual words belong to. In the BNC/COCA lists the words of English have been sorted in lists of 1,000 words each. The first 1,000 are the most frequent words, words 1,001-2,000 are less frequent, etc. Classes with high numbers thus contain low-frequent words.

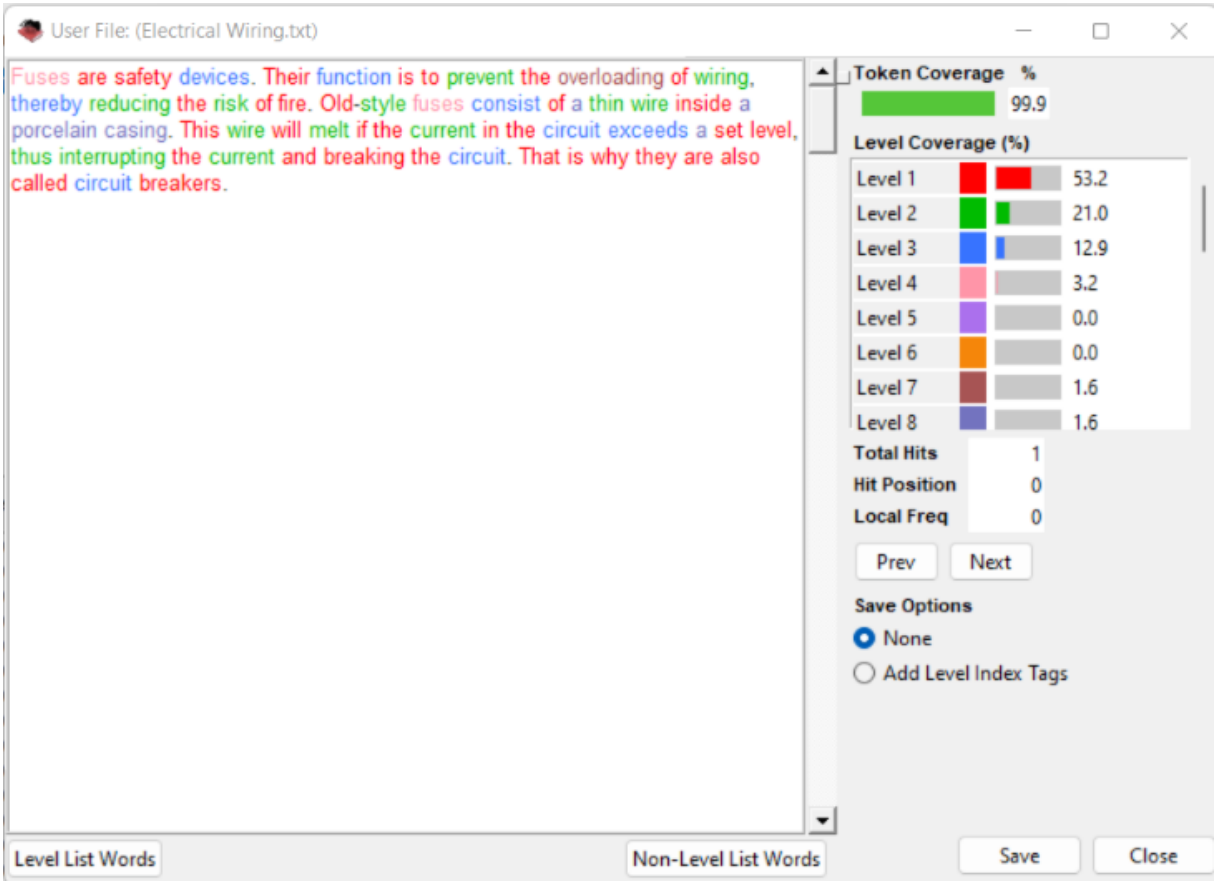


Figure 7 AntWordProfiler results for the text Electrical Wiring in Homes, which is one of the 14 example texts used with the readability index. The words risk and current, for instance, belong to Level 2 (frequency class 2), which has been given a green colour code in the AntWordProfiler display. Level 2 signifies that risk and current are among the second most frequent group of 1000 words in the BNC/COCA corpus and they should thus be commonly understood words.

Etymological dictionaries

Douglas Harper's etymological dictionary (Harper 2019) and Charles T. Onions' etymological dictionary (Onions 1966) have been used to trace etymology. A main point of this study has been to clarify the dichotomy between words of Latin and Greek origin on one side and Germanic words on the other. Thus, whether a word originates directly from Latin, or rather from Latin via Norman French or Paris French has not been deemed important. Where the etymology is listed as «Latin», the word has often been transferred from Latin to English via Old French and/or Anglo-French. This is the case with e.g. the words «approve» and «respond».

Syllables

Determining the number of syllables in each word is generally easy. In cases of doubt, an online syllable dictionary was consulted (How Many Syllables, 2022). There is a dispute about whether «tired», which occurs in the diagnostic vocabulary test, has got one or two

syllables. For the purposes of analysing the diagnostic vocabulary test, “tired” has been regarded as having one syllable, in tune with its designation on the How Many Syllables site.

Phrasal verbs

For the phrasal verbs, the number of syllables has been set in parentheses. It expresses the highest number of syllables in any of the words within the phrasal verb. Phrasal verbs and other complex word combinations clearly stand out as units that deserve to be treated separately. This is, however, beyond the scope of this thesis and must be left for the further development of readability indexes.

Diagnostic vocabulary test – list of etymologies etc.			
Word	Frequency class	Number of syllables	Etymology
immense	4	2	Latin
large	1	1	Latin
gargantuan	12	4	probably Spanish / Portuguese
huge	1	1	Old French
worn out	1	(1)	both elements from Old English
fatigued	5	2	Latin
tired	1	1	Old English
exhausted	2	3	Latin
multiple	3	3	Latin
many	1	2	Old English
numerous	3	3	Latin
a lot	1	(1)	Old English
to continue	1	3	Latin
to go ahead	1	(2)	both elements from Old English
to proceed	3	2	Latin
to carry on	1	(1)	*carry* from Latin, *on* from Old English
rich	1	1	Old English
prosperous	3	3	Latin
well-to-do	1	(1)	all the elements are from Old English
affluent	6	3	Latin
to postpone	4	2	Latin
to put off	1	(1)	both elements from Old English
to delay	3	2	Old French
to defer	4	2	Latin
to begin	1	2	Old English
to commence	5	2	Latin
to start	1	1	Old English

to launch	3	1	Latin
tranquil	5	2	Latin
quiet	1	2	Latin
placid	8	2	Latin
calm	2	1	Old French or Old Italian
small	1	1	Old English
minuscule	10	3	Latin
tiny	2	2	probably Old English
diminutive	9	4	Latin
terrific	5	3	Latin
wonderful	1	3	Old English
glorious	2	3	Latin
marvellous	2	3	Latin

Table 7 List of etymologies in the diagnostic vocabulary test.

Later in the autumn each student received an overview of his/her own results, along with some advice about lexical learning. An example of this feedback document can be seen in Appendix 2. Together with the feedback each student received a link to a follow-up survey (see 3.2) which was conducted online by the means of the "Nettskjema" program.

3.2 A follow-up survey after the diagnostic vocabulary test

On the results form that each student received after doing the diagnostic vocabulary test, there was a link to a short follow-up survey. This survey focused on the students' opinions about different categories of words, and how important it is to learn each category. Some word examples from the diagnostic vocabulary test were used. The follow-up survey was conducted using the "Nettskjema" website (Østfold University College 2022). Nettskjema is a service for data collection, run by the University of Oslo and made available for use by students from several colleges and universities.

The students were first asked how important they think it is to learn these categories of words:

- Vocational words
- Words on difficulty level 1
- Words on difficulty levels 2 and 3
- Words on difficulty levels 4, 5 and 6
- Words on difficulty levels 7 and above

- Phrasal verbs

Each level was exemplified with words that had been used in the diagnostic vocabulary test. The words *exhausted*, *multiple* and *numerous* were e.g. used as examples of levels 2 and 3. The vocational word category was exemplified with words from the respondents' respective study programs, e.g. with construction words like *hacksaw, mortar, excavator* and *trowel* in the form that the construction students filled in. Finally, there was a question about which level of words the respondent uses when writing, and one about which level of words the respondent understands when reading. See Appendix 3 for the original questions in Norwegian, and an English translation.

3.3 Analysing texts to create a Norwegian-geared readability index

One of the aims of this study is to develop a readability index which can help adapting the usage of English vocabulary to different situations, e.g. for developing learning materials for vocational students. An approach to readability is to analyse and classify the occurrence of words in a selection of student and textbook texts, and compare the findings with the level classification that has previously been assigned to these texts – in the form of e.g. marks at an examination or assigned levels of difficulty. How do the values frequency, etymology and number of syllables match the level of the texts? (It must naturally be taken into account that the texts have been assigned levels on the basis of other criteria besides the word level parameters. It can perhaps be expected that the language of an examination text with the mark 5 or 6 is on an advanced level, but the good mark can have been given due to other qualities of the text). An attempt has been made to develop a readability index which can be used to adapt textbook texts and other school materials to suit the students' needs in various situations. When developing the index, the results from the diagnostic vocabulary test concerning frequency, etymology and number of syllables were used for guidance.

3.3.1 Developing a readability index for Norwegian students

It is valuable to have an index that can be used to determine if a word is potentially well readable, or "reader-friendly", for Norwegian students. Such an index can be used to select vocabulary to use in vocational texts, and it can show which vocabulary to avoid. In a vocational text it is important that new, important vocational words are not overshadowed by

difficult general words. The index can help classifying vocabulary words, and it can be instrumental in the process of replacing high-load words with low-load words, i.e. replacing difficult words by easier words. Based on the findings in subchapters 3.1 and 3.2, two types of words in the general language of vocational texts can be defined:

Low-load words ("facilitating" words)

These words are desirable in the general language, because they lessen the vocabulary load in the general language, and frees capacity to tackle, for example, vocational vocabulary. For Norwegian students words of Anglo-Saxon and Old Norse origin will generally be easy to understand. Such words may be cognates or near-cognates, or otherwise familiar in structure. One- and two-syllable words are easy to process, including the abovementioned one-syllable words of French / Latin / Greek origin. In addition, high-frequency words facilitate understanding, so they may be placed on the "plus side" of the readability index.

High-load words ("burdensome" words)

These words create a heavier vocabulary load, and they are therefore mostly undesirable in the general language of vocational texts. They make the language more difficult to understand, which can be detrimental because it may interrupt the learning of vocational words in the text. For Norwegian students, words of French, Latin and Greek etymology will often be unfamiliar, and thus high-load words. One-syllable words of French / Latin / Greek origin are exceptions from this, as they are generally central words in the language, high-frequent and easy to understand. Multi-syllable words can often be a barrier to understanding, as are low-frequency general words. Since these types of words may reduce the readability, they may be placed on the "minus side" of the readability index. Many phrasal verbs and other complex expressions may also contribute to a higher vocabulary load. It is, however, difficult to find a practical method to identify and classfy these expressions.

The favourable and unfavourable properties of a general word, seen from a readability perspective for Norwegian students, can be summed up in this way:

+ Anglo-Saxon
+ Old Norse
- French
- Latin
- Greek

+ monosyllable, high-frequent words of French / Latin / Greek origin

++ high-frequent

+ mid-frequent

- low-frequent

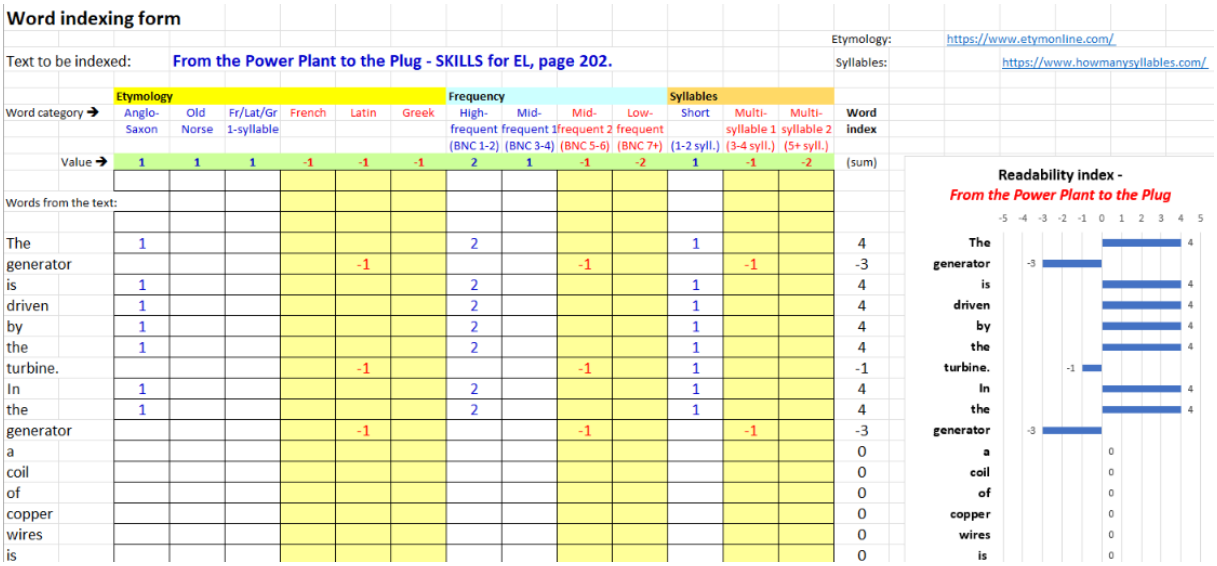
-- very low-frequent

+ monosyllable and bisyllable

- multi-syllable

-- very multi-syllable

In Figure 8 is an example in Excel on how the index can be put to practical use:



Word indexing form

Etymology: https://www.etymonline.com/

Text to be indexed: **From the Power Plant to the Plug - SKILLS for EL, page 202.** Syllables: https://www.howmanysyllables.com/

	Etymology						Frequency				Syllables			
Word category →	Anglo-Saxon	Old Norse	Fr/Lat/Gr 1-syllable	French	Latin	Greek	High-frequent (BNC 1-2)	Mid-frequent 1 (BNC 3-4)	Mid-frequent 2 (BNC 5-6)	Low-frequent (BNC 7+)	Short (1-2 syll.)	Multi-syllable 1 (3-4 syll.)	Multi-syllable 2 (5+ syll.)	Word index
Value →	1	1	1	-1	-1	-1	2	1	-1	-2	1	-1	-2	(sum)
Words from the text:														
The	1						2				1			4
generator					-1				-1			-1		-3
is	1						2				1			4
driven	1						2				1			4
by	1						2				1			4
the	1						2				1			4
turbine.					-1				-1		1			-1
In	1						2				1			4
the	1						2				1			4
generator					-1				-1			-1		-3
a														0
coil														0
of														0
copper														0
wires														0
is														0

Figure 8 Excel form for indexing texts. The text being analysed here has a high degree of readability, which is reflected in high values shown in the blue columns.

With the values that are used in this readability model, word frequency is emphasised more than etymology. Explanation:

- A 'Germanic' word origin (Anglo-Saxon or Old Norse origin) is valued (+1), because such words probably will be better understood by Norwegian students.
- Accordingly, a French / Latin / Greek origin will give the word a negative value (-1), but only if it is a long / low-frequent word.
- Short, high-frequent words of either Germanic or French / Latin / Greek etymology receive a positive value (+1), since students are likely to know them well. Those words are not given any "minus points" due to their origin.

- Frequency has been given much weight. Words have been assigned values according to which of the thousand-word families from the BNC they belong to. The most high-frequent words (frequency class 1-2) are given the value +2, and mid-frequent words of class 3-4 have the value +1. Among the less frequent words, mid-frequent words of class 5-6 count -1 and low-frequent words (class 7 and beyond) count -2. The RANGE software (Webb & Nation 2008) can be applied to determine word frequencies.
- Short words of 1-2 syllables have been given the value +1. If a word has multiple syllables it is generally a drawback for comprehension, so words of 3-4 syllables have been given a negative weight of -1, while words with more syllables than that count as -2.

Excel forms like the one in Figure 8 have been put to practical use in analysing the readability of 14 example texts. This model for readability will then need to be adjusted according to experiences from processing the example texts, and according to the results of the diagnostic vocabulary test.

3.3.2 Selection and indexing of texts

14 example texts were selected to be processed in the Norwegian-geared readability model. To reduce processing load, it was necessary to select random sections of texts. In the sample texts that have been analysed, 50 words from each have been selected, counting from the beginning of the second paragraph of each text. (An exception is the transcript of Obama's speech, which did not have any evident paragraph structure). The example texts can be seen in Appendix 4. The following categories of texts have been indexed:

- Five example examination papers published by the Directorate for Education and Training in 2014. The sample texts were marked by the Directorate (marks 2-6). For each examination paper a section of the longest answer text has been selected.
- Three vocational texts from old and new English textbooks, within mechanics and electrical trades.

- Four texts from the SKILLS English textbook currently used at the school in question. One of the texts is a simple literary text with a general topic. The other two are also literary but deal with vocational topics.
- The last two texts are very formal, written by eloquent authors.

3.3.3 Comparing the Norwegian-geared index to an established readability index

The same example texts that were indexed with the Norwegian-geared index, were also analysed with the Flesch Reading Ease index.

3.3.4 Indexing the words from the diagnostic vocabulary test

By indexing the diagnostic vocabulary test words, it is possible to compare the index results to the actual student answers. This is makes it possible to calibrate the accuracy of the index.

4. Results and discussion

This study has investigated the use of vocabulary, attitudes to the choice of words and different factors that may predict the degree of difficulty of individual words. In this chapter, results of the different tests, surveys and indexes will be presented. Further, the degree to which the Norwegian-geared readibility index that was introduced in 3.3 is well calibrated for practical use will be determined.

4.1 Results of the diagnostic vocabulary test

The diagnostic vocabulary test, which aimed at establishing the charanteristics of the vocabulary which vocational students know, yielded results that are quite clear in some respects. The results can be seen in the Excel spreadsheet in Appendix 1. When correcting the diagnostic vocabulary test, 1 point was given for correct translations and 0 for wrong answers or blanks. Translations that were halfway correct received 0.5 point. In the result overview that each student received, there was both a sum of correct answers in plain numbers (with 1 point for each correct translation), and a sum that was adjusted according to the degree of difficulty of each word. On the plain-numbers scale 1 point per correctly translated word was given. On the level-adjusted scale each word yielded points according to its frequency class. If a student translated a word on BNC level 1 correctly it yielded 1 point, while a word on BNC level 4 yielded 4 points. In total, it was possible to score 40 points on the plain-numbers scale (which means that all 40 words were translated correctly), while it was theoretically

possible to score 122 points on the level-adjusted scale where the BNC levels were taken into account. An example of a feedback form that was given to a participating student can be seen in Appendix 2. The feedback form is not in itself a part of this study; it is rather a results overview that hopefully may inspire the participating students to focus on vocabulary learning.

With 150 participants, the test is rather comprehensive and should thus have a rather good validity. A couple of issues may challenge the reliability of this test. Firstly, it may be difficult for immigrant students to translate into Norwegian, even when knowing the meaning of the word. Secondly, it appears that one or two students have used a dictionary and thus achieved an artificially high score. Nonetheless, these are minor concerns that do not skew the results significantly.

Initially attempts were made to break down the results to section level, with each section covering one of the study programs at the school. However, it was found that a subdivision into sections would not show any major differences, and the validity would suffer from a lower number of participants. The results were therefore not subdivided according to sections.

The overall results from the diagnostic vocabulary test can be see in Figure 9. The percentages represent the ratio of correct translations of each word. Furthermore, there are three variants of the same diagram that illustrate different aspects of the translated words: In Figure 10 the translation results have been arranged according to BNC frequency class. In Figure 11 the translation results have been arranged according to the etymologic origin of each word, and in Figure 12 they have been arranged according to the number of syllables in the words.

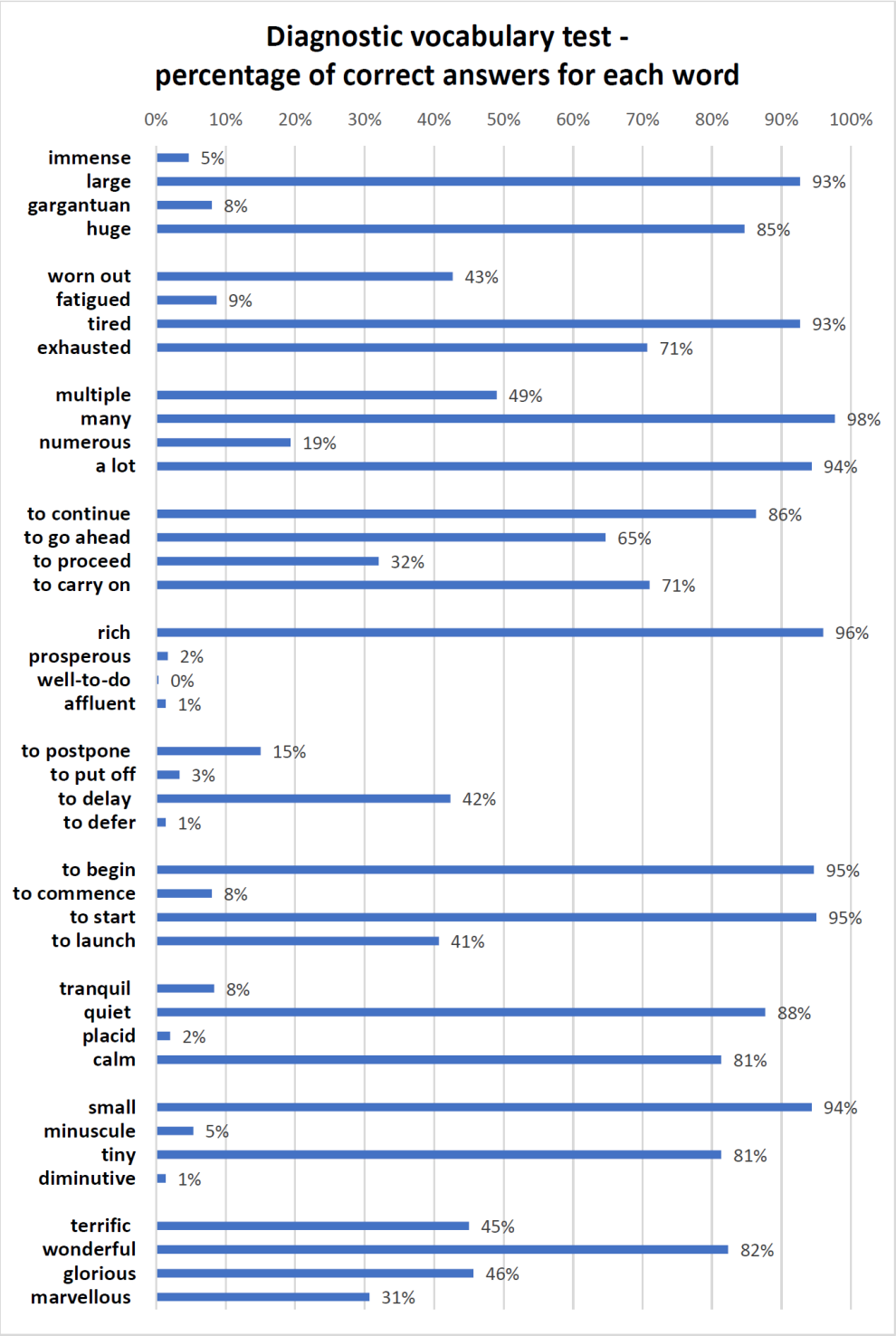


Figure 9 Unsorted results of the diagnostic vocabulary test.

In Figure 9 the results are presented in their unsorted form. The result for each word shows the percentage of correct translations into Norwegian.

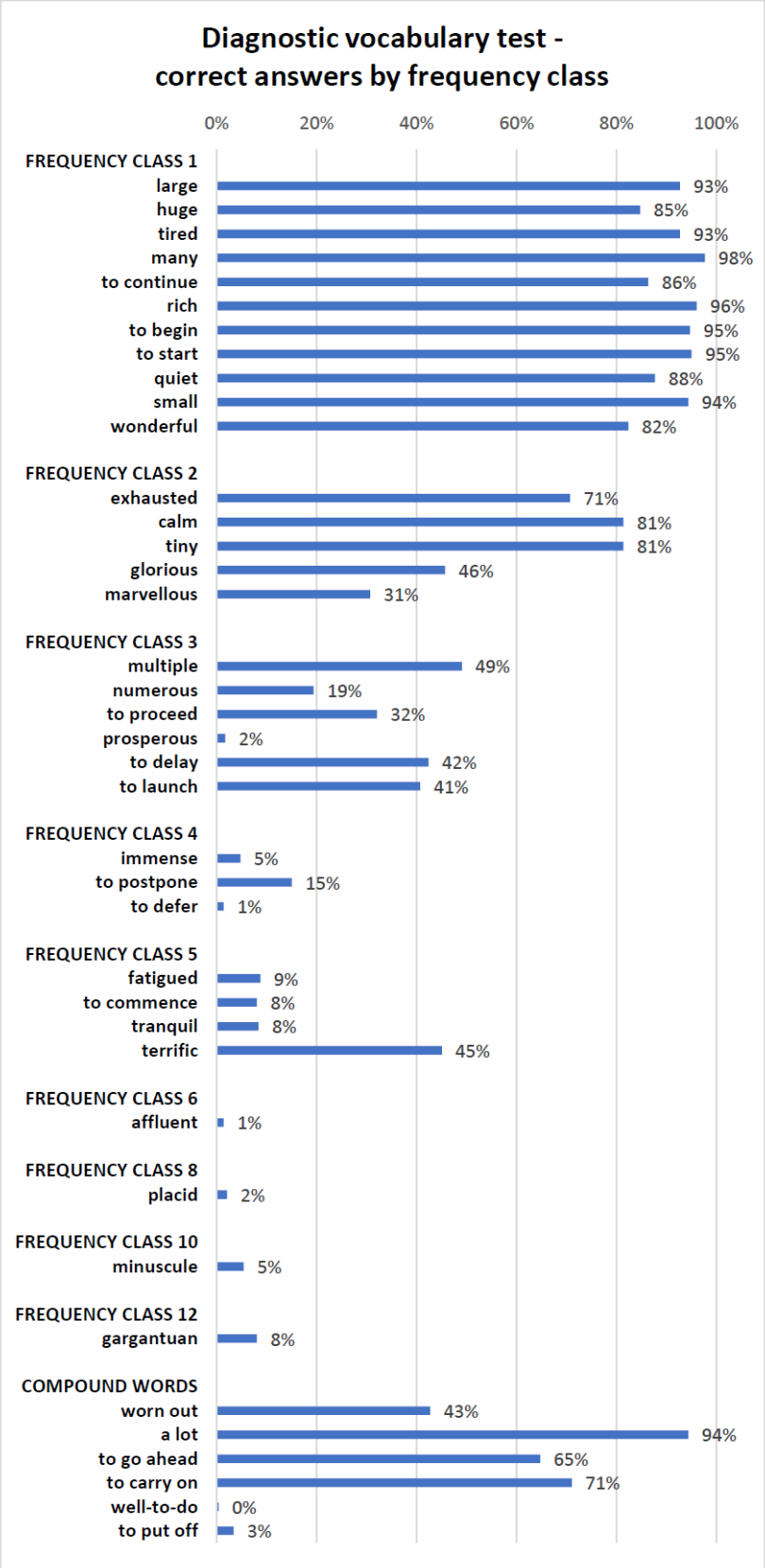


Figure 10 Results from the diagnostic vocabulary test, sorted by frequency class.

In Figure 10 the words in the diagnostic vocabulary test test have been sorted by frequency class. It seems clear that the most high-frequent words are translated correctly by most students, and that the amount of correct answers lessens by diminishing frequency of the words. The word «teriffic» stands out. It has a double meaning, and it may also have attained a greater frequency since the BNC/COCA list was created.

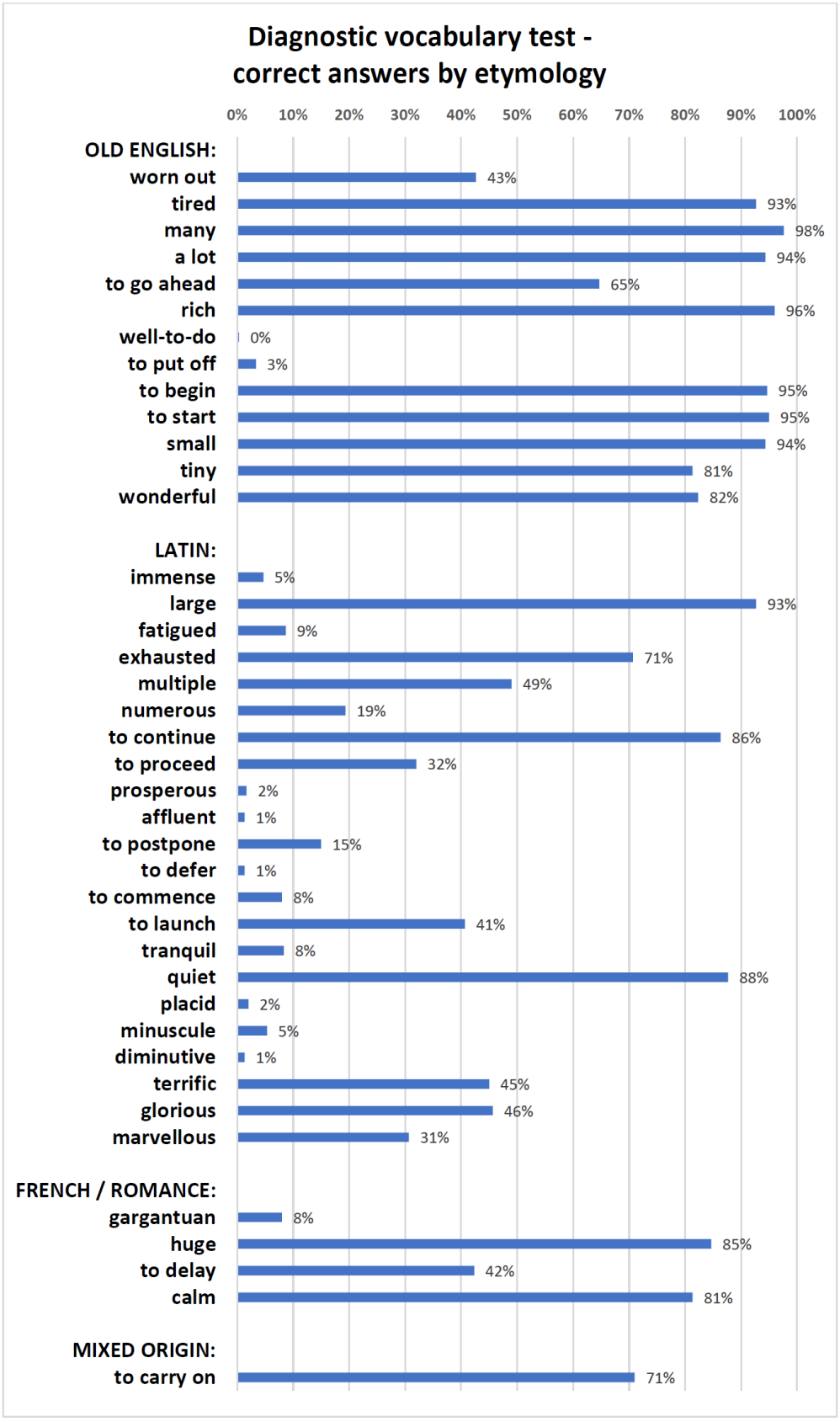


Figure 11 Results from the diagnostic vocabulary test, sorted by etymology.

Figure 11 shows correct translations into Norwegians of the words, sorted by the etymology of the words. Many words have come into English from Latin via French, and these words have been placed in the «Latin» category. A few words originate from other Romance languages, e.g. «gargantuan», which stems from Spanish or Portuguese. In the expression «carry on» the first element comes from Latin and the last from English.

The words of Old English origin were clearly best known to the students, with the exception of the phrasal verbs. For the words of Latin / French / other Romance origin the results are mixed. Some of these «Graeco-Latin» words, particularly the one- or two-syllable terms, are well known, while the multi-syllable words are unknown to many of the students

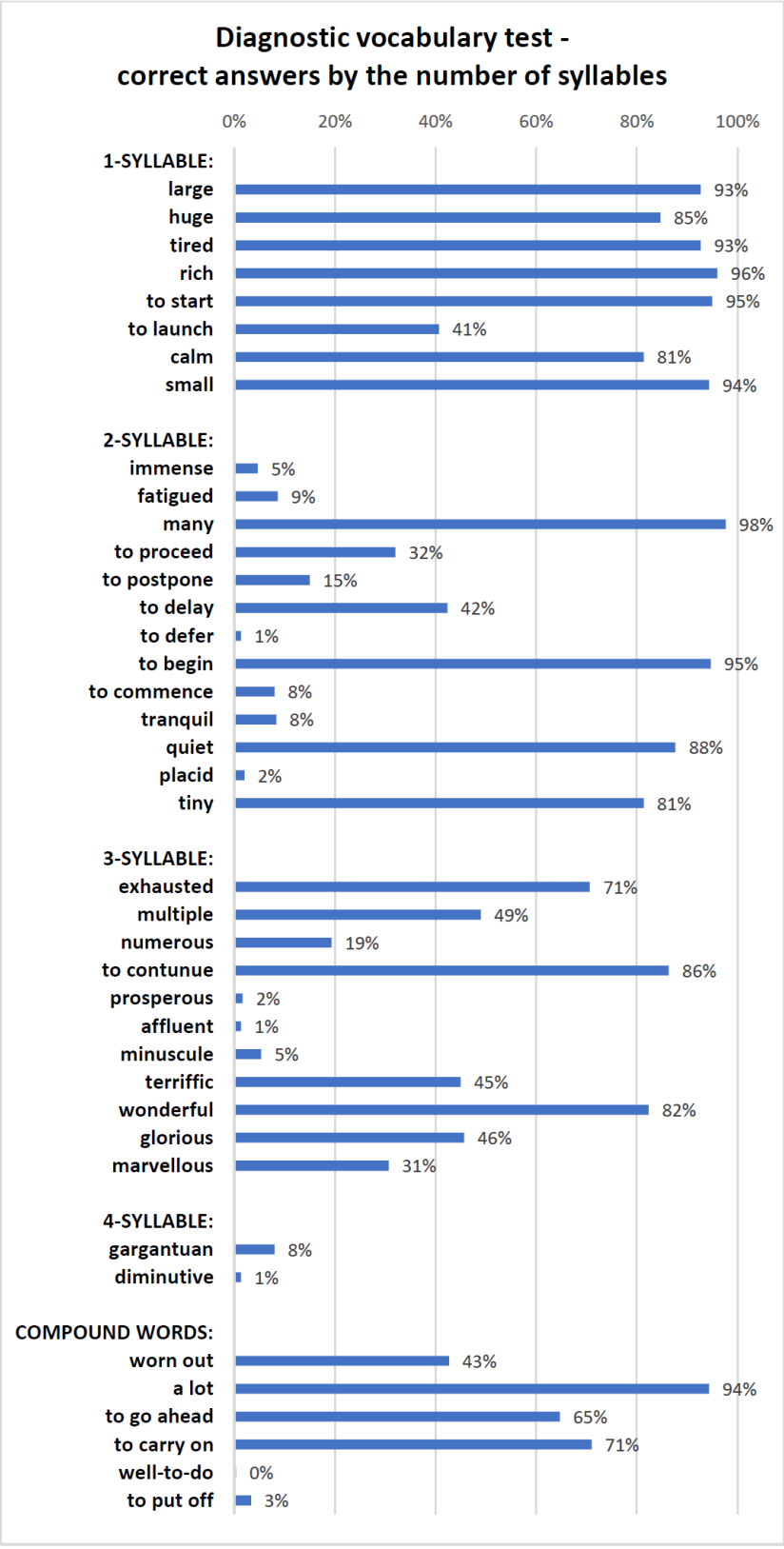


Figure 12 Results from the diagnostic vocabulary test, sorted by the number of syllables of the words.

In Figure 12 the translation results have been arranged according to the number of syllables in each word. The numbers of syllables in *How Many Syllables* (2022) have been used. The

compound words have been singled out as a specific group, since they cannot be categorised in the same manner as single words.

The monosyllabic words were the ones most correctly translated by the students, while the correct answers became more scarce as the number of syllables increased. The compound words in the test mostly consisted of monosyllabic words, but there is a striking difference between them in how well they were known to the students. While most students could translate «to carry on», «to put off» was unknown to students.

4.2 Results of the follow-up survey

The aim of the follow-up survey was to establish which categories of words the vocational students percieve to be worthwhile learning. The results can be seen in the Excel spreadsheet in Appendix 3. The number of respondents in the follow-up survey is 68, compared to 150 participants in the diagnostic vocabulary test. The numbers of respondents from each study program varied widely. This may influence the validity and reliability of the survey. There are fewer answers, and it is possible that the survey was answered by e.g. a disproportionate percentage of the students who scored well on the test. As with the test in 4.1, only the summarised answers from the whole survey are displayed.

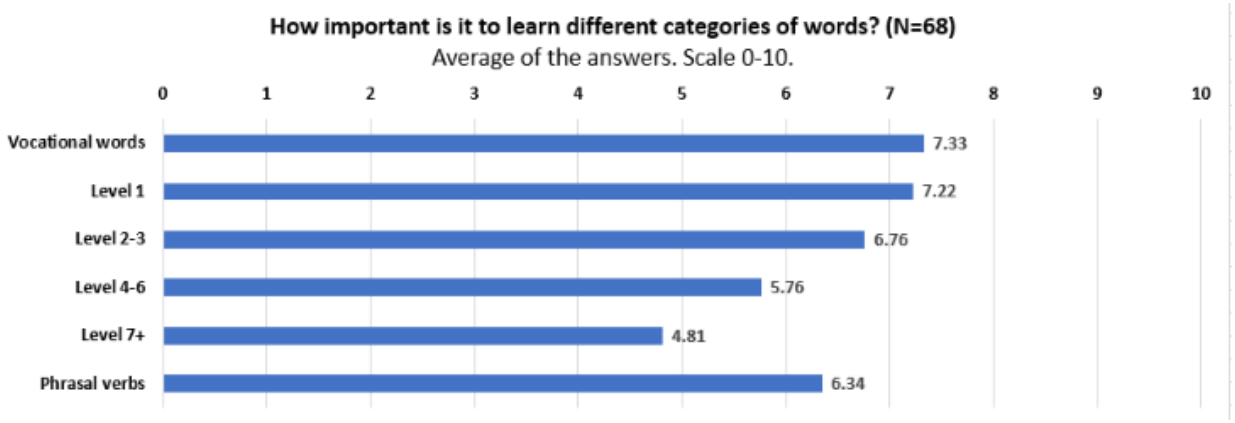


Figure 13 Results of the follow-up survey, concerning the importance of learning different categories of words.

Figure 13 suggests that the respondents generally prioritise to learn vocational words, together with the most basic words. Level 2-3 words and more advanced words are considered important to a degree diminishing in correlation with their level of difficulty. Phrasal verbs

are regarded as comparatively important to learn. The latter probably reflects the results from the diagnostic vocabulary test, where the same students scored poorly on some of the phrasal verbs. It may indicate that students struggle with phrasal verbs but wish to know them. On the other hand, low-frequent words may be equally difficult, but are not seen as so desireable to learn. All in all, the response to this survey is rather unambiguous, and the results are as expected in so far as the most frequent words are reguarded as being most important.

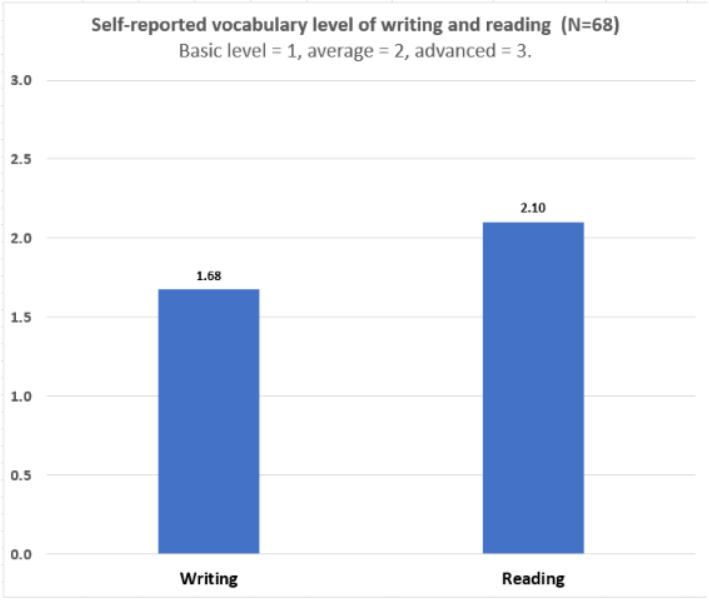


Figure 14 Self-reported mastery of vocabulary in writing and reading. The answers have been converted to numbers in this way: simple = 1, average = 2, advanced = 3.

On the questions about writing and reading, the respondents clearly reported that they understand a higher vocabulary level than they apply themselves their writing.

4.3 Results of the text analyses

The readability indexing of the 14 individual example texts revealed quite a span in the degrees of difficulty. The results can be seen in the Excel spreadsheet in Appendix 6. The indexing of four individual texts is shown in Figures 15a and 15b, and Figure 16 gives an overview of the results for all 14 texts. The visualisation with one column for each word

makes it easy to get an overview and spot which sections of the text may need to be edited for better readability. Only 50 words of each text have been indexed, due to time constraints. This may challenge the validity of the results, but on the other hand the vocabulary in each text is rather uniform. The results show clear differences between the texts, which is an indication that the readability index actually discriminates between different levels of difficulty and thus can be used as a tool when editing learning materials.

The results of the examination texts, written by Norwegian students, were surprising in so far as the best essay (marked 6) had the simplest language, even though there are many other factors than vocabulary in the grading of essays. The three vocational texts («Electrical Wiring», «From the Power Plant to the Plug» and «Welding») are clearly difficult to read, together with the two literary texts that have vocational topics (and vocational vocabulary), i.e. «It's a Wonderful, Digital World?» and «Adiós Hydraulics».

«Something About Me» belongs to the introductory chapter in *SKILLS*, and has a very simple language. The simplified «In Short» version of «Adiós Hydraulics» is much easier to read than the original text, as it ought to be. However, it gives food for thought that the vocational texts are on the same readability level as Obama's speech and «The Queen's Speech 2022». Texts aimed at vocational students must be expected to be easier to read than speeches on the highest levels of politics. The readability of the vocational texts does not seem to have improved over time (The text «Electrical Wiring» is from 2001, while «From the Power Plant to the Plug» and «Welding» are from 2020). Even though only a small selection of vocational texts has been studied here, the findings suggest that it is worthwhile to intervene in the vocational texts and make them more reader-friendly.

As stated above, Figures 15a and 15b show readability index results for four selected texts. Word frequency, etymology and the number of syllables have been taken into account in the index, as outlined in subchapter 3.3.1.

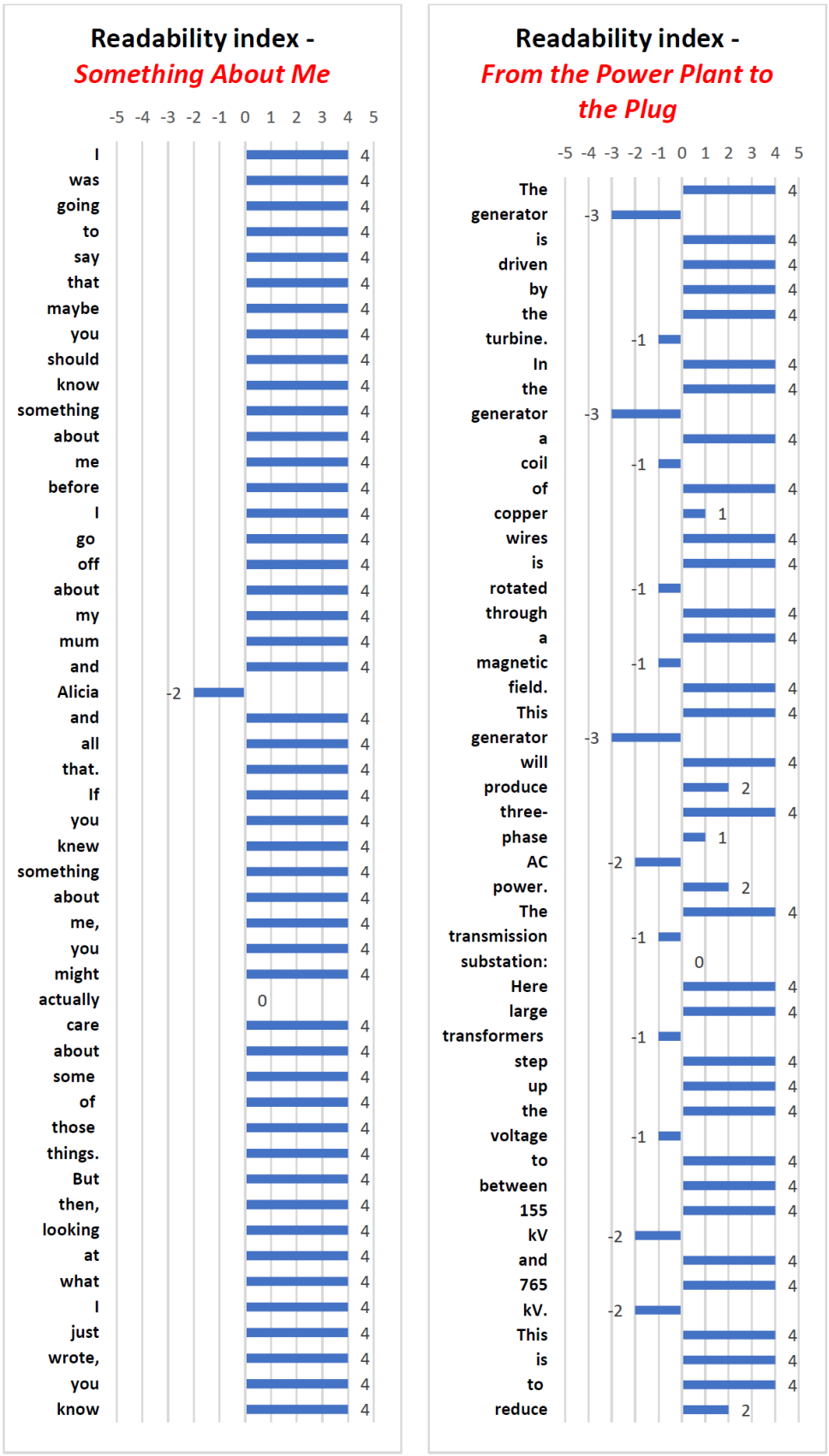
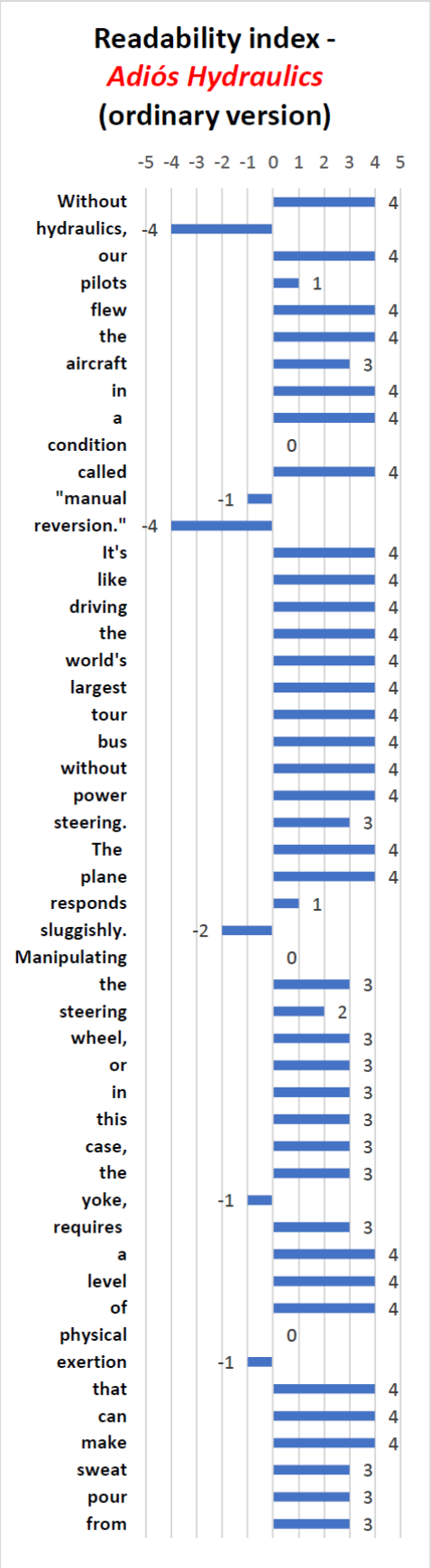


Figure 15a Readability index results.



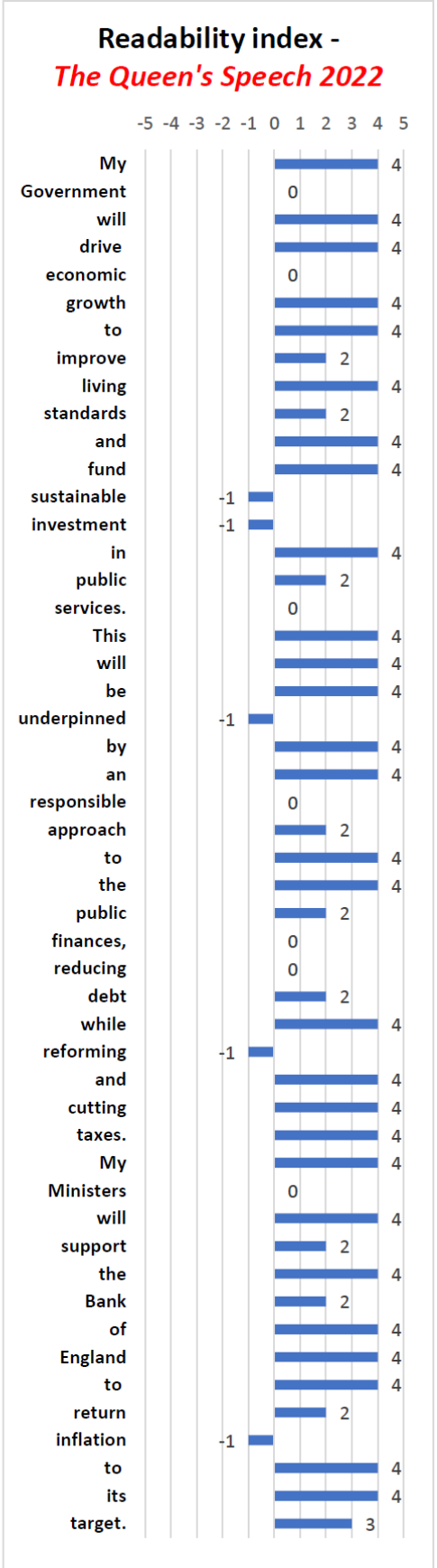


Figure 15b Readability index results.

Figure 16 shows the weighted average index results for all of the texts that have been indexed. The results are a function of frequency, etymology and the number of syllables in the words in each text. The results are «weighted» in the sense that desired language traits, e.g. few syllables and Germanic etymology, have been given additional points in the calculation of each word's result. The texts with the lowest figure have the most difficult vocabulary.

Figure 16 Results from the Norwegian-geared readability index – weighted averages

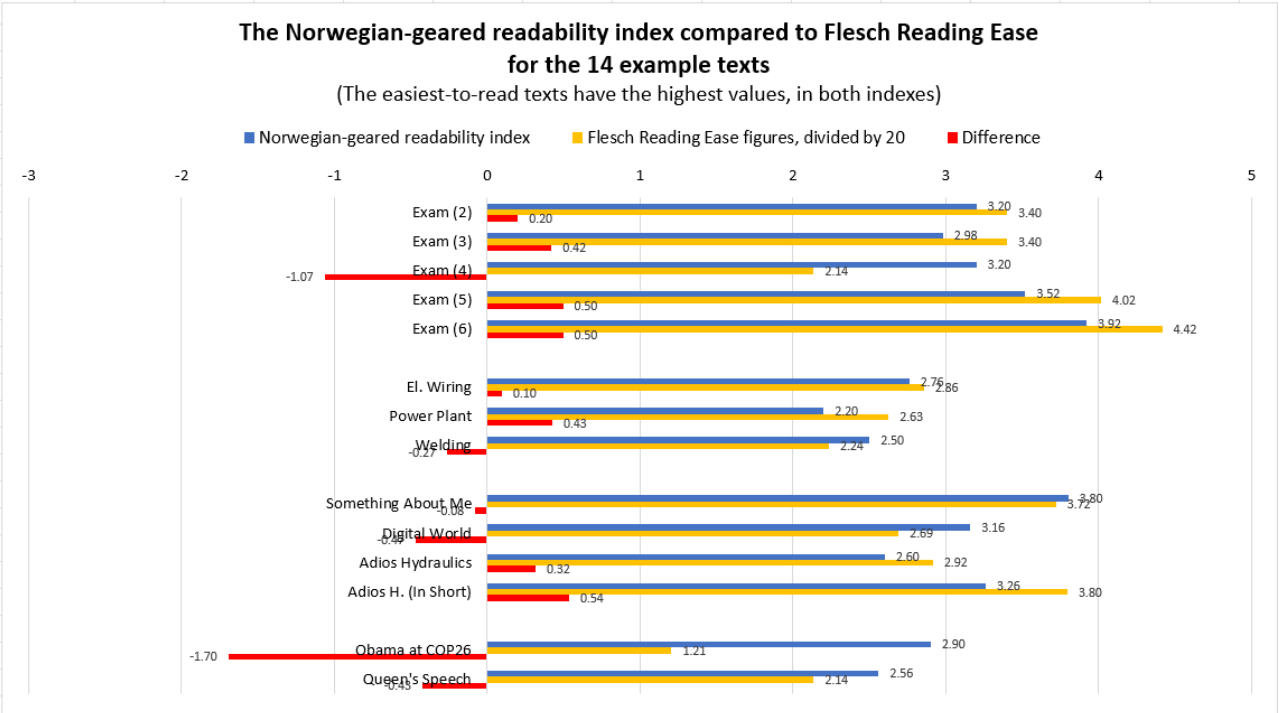


Figure 17 Results from the Norwegian-geared readability index, compared to Flesch Reading Ease results

Figure 17 shows the results from the Norwegian-geared readability index, compared to the percentage of words of frequency class 1 in the BNC corpus. Frequency class 1 contains the 1,000 most frequent words in the English language. The percentages are fetched from the processing of each text in AntWordProfiler. To display the two sets of values in the same figure, the frequency 1 results have been divided by 20. This has been done solely to make the results more visually comparable in the graph. Thus there is no logical connection between the two scales, and the comparison only gives a relative impression.

The results of this comparison are not very clear. The relationship between the readability results and the class 1 results is approximately the same for most texts. This might be seen as a verification that the readability index compares well to an objective measure like the frequency results, but on the other hand it may just reflect that word frequency weighs much in the index. A few texts show a different pattern. That probably reflects that etymology and the number of syllables influence the readability results in those texts.

In Figure 18, the readability index has been applied to index the 40 words in the diagnostic vocabulary test. Doing this is a way of calibrating the index. Figure 18 shows that there is a good match for many words. For words like immense, prosperous and defer, however, there are great discrepancies. This propbably is a sign that the index underestimates the difficulty of low-frequent words. Such comparisons can be a good tool to adjust the readability index so that it reflects the real knowledge of the target group.

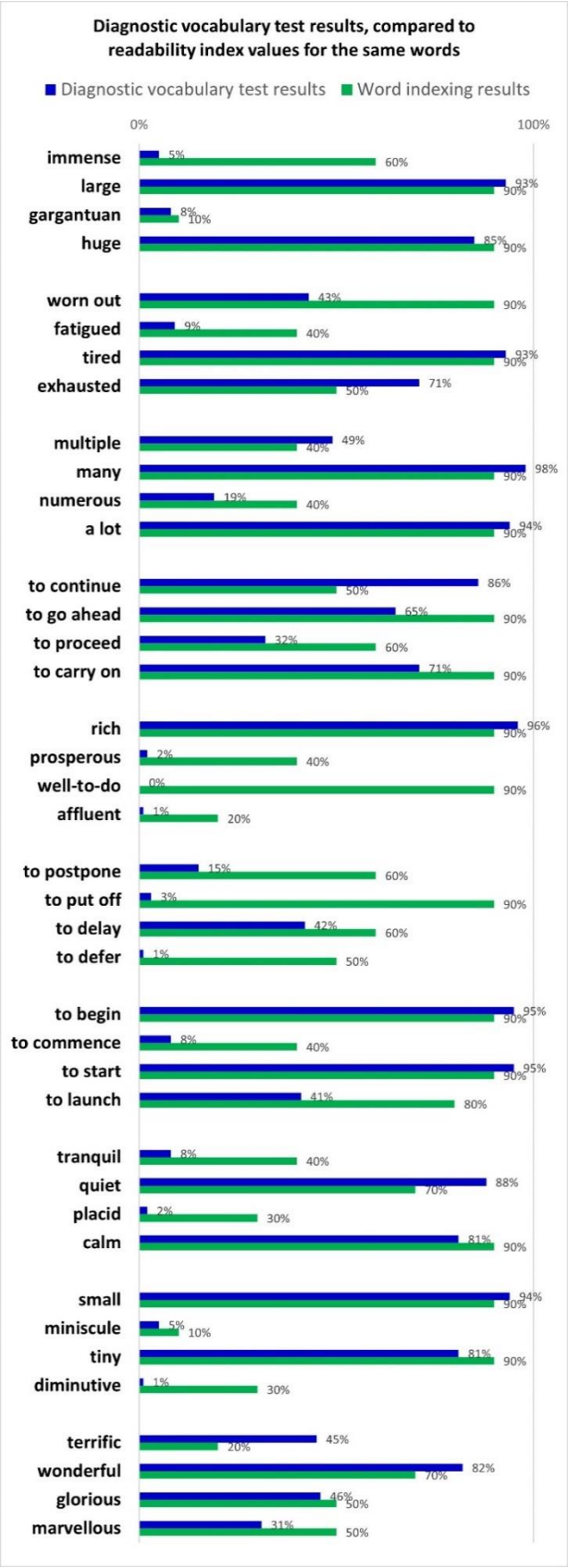


Figure 18 Results of the diagnostic vocabulary test, compared to readability index values for the same words

Limitations of this study

While this study has yielded some clear and evident results, e.g. about the lower readability of low-frequent and multisyllable words, some parts of the results are open to discussion and adjustments. It has become clear that the values used in the readability indexing form must be adjusted, so that they match the results from the diagnostic vocabulary test better. The BNC/COCA corpus that was used, is rather old and sometimes gives word frequencies that differ from modern English usage. The difficulties of comprehending phrasal verbs have not been addressed by this study.

An obvious weakness of the translation method in the diagnostic vocabulary test is the fact that some students have other mother tongues than Norwegian. There may be cases where a student knows the meaning of the English word but does not know its Norwegian equivalent. Given that most participants in the study are first-language Norwegian users this phenomenon does probably not have a very large impact on the results. To clarify this issue there might have been added a question to the form, where each student indicated his or her mother tongue.

When analysing the results of the diagnostic vocabulary test, it was interesting to see how the results differed according to the frequencies of each word, their etymological origin, and their number of syllables. When trying to see a pattern in the results, it was important to have in mind that the selection of words in the survey was done more or less at random. The words were selected to form four-word sets of synonyms. An approach without using synonyms might have opened for arranging a more systematic distribution of the survey words, e.g. with a more systematic distribution between word frequencies.

Recommendations for further research and further applications

In this thesis, indexing has been done manually in a very laborious manner. Using the affordances of today's computer software, it will be possible to draw indexing data from premade lists of words. Values for both frequency, syllable count, etymology and other relevant factors may be built into the software, enabling it to do the indexing of each word

automatically. Sentence length and complexity may be included, and it may even be possible to solve the challenges of processing phrasal verbs and other compound words. The user interface might be designed so that it gives a visualisation of the readability of each word as the user views and edits the text. Indexing made by existing computer programs already feature such functions and it is technically possible to construct software for the Norwegian-geared readability index. When designing such software it will also be possible to improve the readability index by applying a gradual scale of the values for word frequency and number of syllables, instead of assigning whole values like 1 or 2. These approaches will make readability work more practicable and the results more accurate.

5 Conclusion

The work with this thesis has shown many aspects of the importance of vocabulary and its properties. It is clearly necessary to be aware of the need to adjust vocabulary levels in everyday school work. A matter of particular importance is the need to simplify the general language in vocational texts, so that the students can focus on vocational terms that are new to them.

The methods applied have revealed that there are large, and to a great extent systematic, limitations in the English vocabulary of the students who took part in the diagnostic vocabulary test. Broken down to the word properties that were studied, the test showed the greatest knowledge of high-frequent words, and words of 1-2 syllables. For etymology there was a tendency towards a greater knowledge of words of Germanic origin. When asked in the survey of which words are the most important to know, the students prioritised having the knowledge of the most frequent words, vocational vocabulary and phrasal verbs.

The findings in the vocabulary test and the survey were used to construct a readability index where the settings were adjusted to favour «reader-friendly» parameters for the target group.

The index was tested by applying it for indexing a number of relevant texts; vocational textbook texts, general textbook texts and authentic texts. Comparing the different indexing results shows both that the index discriminates between texts in an adequate manner, and that the vocational texts have an unnecessary high level of difficulty.

The attempts at constructing the index have been partly successful, but the values must be adjusted and the index must be adapted further. Some special topics ought to be investigated further, e.g. the readability of phrasal verbs. The work with the readability index suggests that it can be developed into a useful tool for adapting learning materials. The goals for this study have for a large part been achieved, and the knowledge that has been acquired can be applied further to obtain more accurate results.

The work with vocabulary can be time-consuming. Instead of indexing texts manually, like in this study, it will be necessary to design software to make the readability index practically applicable. Indexing larger amounts of text will also enhance the validity of the results. Hopefully, the idea of adapting vocabulary work to match the students' mother tongue may be a contribution to the language teaching methods, particularly in connection with vocational texts.

List of references

Baugh, A. C. & Cable, T. (2013). *A History of the English Language*. Sixth edition. London: Routledge.

Chetail, F. (2014). Effect of number of syllables in visual word recognition: New insights from the lexical decision task. Language, Cognition and Neuroscience. 29. 1249-1256. 10.1080/23273798.2013.876504.

Corson, D. (1995). *Using English Words*. New York: Springer Dordrecht. DOI: https://doi.org/10.1007/978-94-011-0425-8

DuBay, W. H. (2004). *The Principles of Readability*. Retrieved from https://www.researchgate.net/publication/228965813_The_Principles_of_Readability on 22 August 2022.

Freeborn, D. (2006). *From Old English to Standard English. A Course Book in Language Variation across Time*. Third Edition. New York: Palgrave MacMillan.

Friðriksdóttir, S. D. (2014). Old Norse Influence in Modern English. The Effect of the Viking Invasion. B.A. essay at the University of Iceland, School of Humanities, Department of English. Kt.: 010889-3329. Retrieved from https://skemman.is/handle/1946/17234 on 17 February 2019.

Gass, S., Behney, J. and Plonsky, L. (2013). *Second Language Acquisition. An Introductory Course*. Fourth edition. New York and London: Routledge.

Gelderen, E. van. (2014). *A History of the English Language*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Gulbrandsen et al. (2001). *@t work: us. Elektrofag*. Oslo: Aschehoug.

Guo, S., Zhang, G., & Zhai, R. (2011). Integrating readability index into Twitter search engine. *British Journal of Educational Technology*, *42*(5), E103–E105. https://doi.org/10.1111/j.1467-8535.2011.01206.x

Harper, D. (2019, 30 December). Online Etymology Dictionary. https://www.etymonline.com/

Hestetræet, T. I. and Ørevik, S. (2018). Chapter 13. English in Vocational Studies. In A-B. Fenner and A. S. Skulstad (Eds.), *Teaching English in the 21st Century. Central Issues in English Didactics*. Oslo: Fagbokforlaget.

How Many Syllables. (2022, 17 July). How Many Syllables. https://www.howmanysyllables.com/

Horst, M., Cobb, T. and Meara, P. (1998). Beyond A Clockwork Orange: Acquiring Second Language Vocabulary through Reading. *Reading in a Foreign Language, 11* (2), 207-223.

Hsu, W. (2011). The vocabulary thresholds of business textbooks and business research articles for EFL learners. *English for Specific Purposes 30*, 247-257.

Ibrahim, R. (2006). Do Languages with Cognate Relationships have Advantages in Second Language Acquisition? *The Linguistics Journal 1* (3), 66-96. Retrieved from https://www.linguistics-journal.com/2014/01/09/do-languages-with-cognate-relationships-have-advantages-in-second-language-acquisition/ on 31 March 2019.

Jahr, E. H. (1987). Språkutviklinga etter 1814: Språkstrid og språkplanlegging. In E. B. Johnsen (Ed.), *Vårt eget språk. Bind 1. I går og i dag* (pp. 66-137). Oslo: Aschehoug.

Johnson, B. et al. (2022). The Queen's Speech 2022. Her Majesty's most gracious speech to both Houses of Parliament. https://www.gov.uk/government/speeches/queens-speech-2022 Retrieved on 27 July 2022.

Katamba, F. (2005). *English Words. Structure, History, Usage*. London: Routledge.

Kroll, J. F., Sumutka, B. M., and Schwartz, A. I. (2005) A cognitive view of the bilingual lexicon: Reading and speaking words in two languages. *International Journal of Bilingualism 9 (1)*, 27-48.

Lambert, E. & Kandel, S. & Michel, F. & Espéret, E. (2007). The effect of the number of syllables when writing poly-syllabic words.

Lengeling, M. M. (1995). True Friends and False Friends. *MEXTESOL Journal 19 (2), Convention Issue 1995*, 17-21. Retrieved from http://mextesol.net/journal/public/files/e3bf14f3334a6a1f26d048e08f9e47e7.pdf on 29 December 2019

Lightbown, P. M. and Spada, N. (2006). *How Languages are Learned*. 3rd edition. Oxford: Oxford University Press.

Lokøy, G., Hellesøy, S., Langseth, J. and Lundgren, H. C. U. *SKILLS, teknologi- og industrifag. Engelsk YF, vg1*. 2nd edition. 2020. Oslo: Gyldendal.

Lundgren, H. C. U, Langseth, J., Lokøy, G. and Hellesøy, S. *SKILLS, elektro og datateknologi. Engelsk YF, vg1*. 2nd edition. 2020. Oslo: Gyldendal.

Mallory, J. P. and Adams, D. Q. (2006). *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. Oxford: Oxford University Press.

Merriam-Webster online dictionary. (2019). Merriam-Webster, Incorporated. https://www.merriam-webster.com/ Retrieved on 9 Nov 2019.

Morris, W. and Magnusson, M.: *The Saga of Gunnlaug the Worm-Tongue and Rafn the Skald*. (1901). Translation from the original Icelandic 'Gunnlaugs saga ormstungu'. Retrieved from https://sagadb.org/gunnlaugs_saga_ormstungu.en on 12 August 2019.

Midgley, K. J., Holcomb, P. J., Grainger, J. (2011). Effects of Cognate Status on Word Comprehension in Second Language Learners: An ERP Investigation. *Journal of Cognitive Neuroscience. Jul2011, 23* (7), pp. 1634-1647. DOI: 10.1162/jocn.2010.21463

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening?. *Canadian modern language review*, *63*(1), 59-82.

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language*. 2nd edition. Cambridge: Cambridge University Press.

Nation, I. S. P. (2021). Range program with BNC/COCA frequency lists (Version 1.0.0). A set of Range language corpus files containing lists with 25,000 words from the BNC/COCA corpus, downloaded from Nation's website at the Victoria University of Wellington - https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs on 27 July 2021. The files were accessed through this link: *FILE*586.8KBRange with 25,000 words from the BNC/COCA lists (version 1.0.0)

National Partnership for Reinventing Government. Communicators Guide. From the webpage archive of the University of North Texas, https://govinfo.library.unt.edu/npr/library/review.html . Retrieved from https://govinfo.library.unt.edu/npr/library/papers/bkgrd/communicators.pdf on 17 September 2022

Obama, B. (2021). Barack Obama COP 26 Climate Speech Transcript. 8 November 2021. Retrieved from https://www.rev.com/blog/transcripts/barack-obama-cop26-climate-speech-transcript on 30 July 2022

Onions, C. T. (ed.). (1966). *The Oxford Dictionary of English Etymology*. Oxford: Clarendon Press.

Otwinowska, A. & Szewczyk, J. (2017): The more similar the better? Factors in learning cognates, false cognates and non-cognate words. *International Journal of Bilingual Education and Bilingualism*, published on 17 May 2017. DOI: 10.1080/13670050.2017.1325834 Link: https://doi.org/10.1080/13670050.2017.1325834

Parker, A.J., Egan, C., Grant, J. H., Harte, S., Hudson, B.T., Woodhead, Z. V. J. (2021). The role of orthographic neighbourhood effects in lateralized lexical decision: a replication study and meta-analysis. PeerJ. 2021 Apr 28;9:e11266. doi: 10.7717/peerj.11266. PMID: 33986993; PMCID: PMC8088209. Retrieven on 13 Nov 2022 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8088209/

Richards, J.C., Platt, J and Platt, H. (1992). *The Dictionary of Language Teaching and Applied Linguistics*. London: Longman.

Sandøy, H. (2004). Norvagisering og fornorsking. In H. Sandøy and J. Östman (Eds.), «*Det främmande» i nordisk språkpolitik. Om normering av utländska ord* (pp. 107-141, Oslo: Novus forlag.

Siew, C. S. Q. (2018). The orthographic similarity structure of English words: Insights from Network science. *Appl Netw Sci* **3,** 13 (2018) doi:10.1007/s41109-018-0068-1. https://appliednetsci.springeropen.com/articles/10.1007/s41109-018-0068-1 Retrieved on 20 October 2019.

Sésarsdottir, A. R. (2015). Doublets. A Study of Old Norse Influence on English Vocabulary. B.A. essay at the University of Iceland, School of Humanities, Department of English. Kt.: 280783-4719. https://skemman.is/handle/1946/22792 Retrieved on 17 February 2019

Stedje, A. (1989). *Deutsche Sprache gestern und heute. Einführung in Sprachgeschichte und Sprachkunde*. UTB für Wissenschaft, UNI-Taschenbücher 1499. München: Wilhelm Fink Verlag.

Torp, A. & Vikør, L. S. (2003). *Hovuddrag i norsk språkhistorie*. 3rd edition. Oslo: Gyldendal.

Utdanningsdirektoratet (The Norwegian Directorate for Education and Training). (2014). Vurderte eksamenssvar i engelsk fellesfag. Retrieved from http://www.udir.no/Vurdering/Eksamen-videregaende/ in 2014.

Vocabulary.com Dictionary. (2019). https://www.vocabulary.com/dictionary/ Retrieved on 1 Nov 2019.

Waring, R & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language. Oct 2003, 15* (2), pp. 131-163. ISSN 1539-0578.

Webb, S. & Nation, P. (2008). Evaluating the vocabulary load of written text. TESOLANZ Journal. 16.

Webb, S., Newton, J. and Chang, A. (2013). Incidental Learning of Collocation. *Language Learning 63* (1), 91–120 91. DOI: 10.1111/j.1467-9922.2012.00729.x

World eBook Library. (2019). "Swadesh List". Article from the World Heritage Encyclopedia. Retrieved from http://books.worldebooklibrary.org/articles/eng/Swadesh_list on 18 August 2019.

Østfold University College. (2022). Nettskjema. https://www.hiof.no/english/services/it/programs-services/nettskjema/index.html Retrieved on 7 August 2022.

Lists of tables

Lists of figures

Sources for software and language corpus files

Anthony, L. (2021). *AntWordProfiler* (Version 1.5.0) [Computer Software]. Downloaded from http://www.laurenceanthony.net/software/antwordprofiler/ on 29 July 2021.

Nation, I. S. P. (2021). Range program with BNC/COCA frequency lists (Version 1.0.0). A set of Range language corpus files containing lists with 25,000 words from the BNC/COCA corpus, downloaded from Nation's website at the Victoria University of Wellington - https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs on 27 July 2021. The files were accessed through this link: *FILE*586.8KBRange with 25,000 words from the BNC/COCA lists (version 1.0.0)

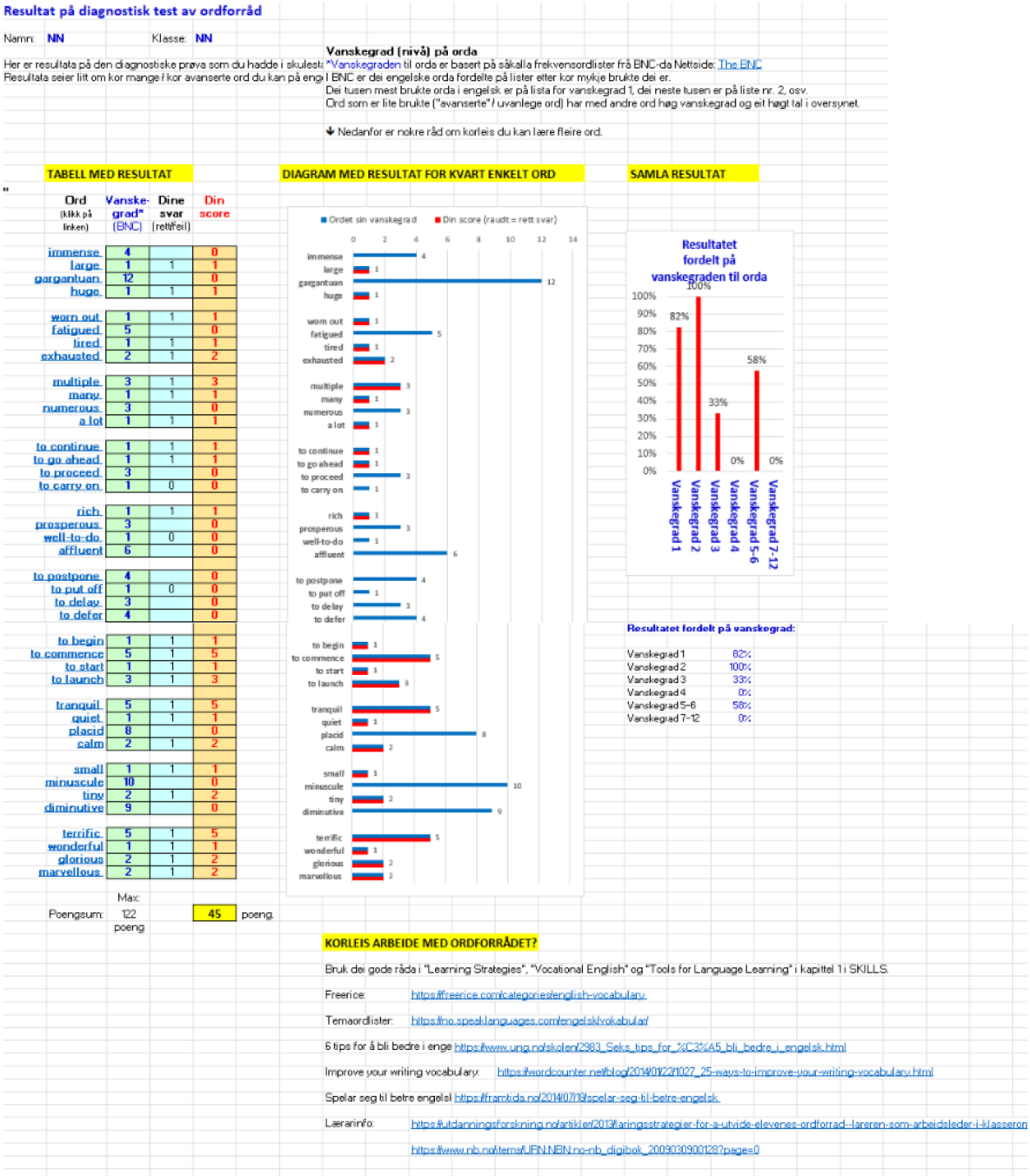
Appendices

APPENDIX 1

Excel document with results from the diagnostic vocabulary test.

APPENDIX 2

Example of a document with feedback on the diagnostic vocabulary test:



Resultat på diagnostisk test av ordforråd

Namn: NN Klasse: NN

Her er resultata på den diagnostiske prøva som du hadde i skulest *Vanskegraden til orda er basert på såkalla frekvensordlister frå BNC-da Nettside: The BNC
Resultata seier litt om kor mange *kor avanserte ord du kan på engI BNC er dei engelske orda fordelte på lister etter kor mykje brukte dei er.
Dei tusen mest brukte orda i engelsk er på lista for vanskegrad 1, dei neste tusen er på liste nr. 2, osv.
Ord som er lite brukte ("avanserte" / uvanlege ord) har med andre ord høg vanskegrad og eit høgt tal i oversynet.

Vanskegrad (nivå) på orda

↓ Nedanfor er nokre råd om korleis du kan lære fleire ord.

TABELL MED RESULTAT

Ord (klikk på linken)	Vanske-grad* (BNC)	Dine svar (rett/feil)	Din score
immense	4		0
large	1	1	1
gargantuan	12		0
huge	1	1	1
worn out	1	1	1
fatigued	5		0
tired	1	1	1
exhausted	2	1	2
multiple	3	1	3
many	1	1	1
numerous	3		0
a lot	1	1	1
to continue	1	1	1
to go ahead	1	1	1
to proceed	3		0
to carry on	1	0	0
rich	1	1	1
prosperous	3		0
well-to-do	1	0	0
affluent	6		0
to postpone	4		0
to put off	1	0	0
to delay	3		0
to defer	4		0
to begin	1	1	1
to commence	5	1	5
to start	1	1	1
to launch	3	1	3
tranquil	5	1	5
quiet	1	1	1
placid	8		0
calm	2	1	2
small	1	1	1
minuscule	10		0
tiny	2	1	2
diminutive	9		0
terrific	5	1	5
wonderful	1	1	1
glorious	2	1	2
marvellous	2	1	2
	Max:		
Poengsum:	122 poeng	45	poeng

DIAGRAM MED RESULTAT FOR KVART ENKELT ORD

SAMLA RESULTAT

Resultatet fordelt på vanskegrad:

Vanskegrad 1	82%
Vanskegrad 2	100%
Vanskegrad 3	33%
Vanskegrad 4	0%
Vanskegrad 5-6	58%
Vanskegrad 7-12	0%

KORLEIS ARBEIDE MED ORDFORRÅDET?

Bruk dei gode råda i "Learning Strategies", "Vocational English" og "Tools for Language Learning" i kapittel 1 i SKILLS.

Freerice: https://freerice.com/categories/english-vocabulary

Temaordlister: https://no.speaklanguages.com/engelsk/vokabular/

6 tips for å bli bedre i engel https://www.ung.no/skolen/2983_Seks_tips_for_%C3%A5_bli_bedre_i_engelsk.html

Improve your writing vocabulary: https://wordcounter.net/blog/2014/01/22/1027_25-ways-to-improve-your-writing-vocabulary.html

Spelar seg til betre engelsk https://framtida.no/2014/07/19/spelar-seg-til-betre-engelsk

Læraninfo: https://utdanningsforskning.no/artikler/2013/laringsstrategier-for-a-utvide-elevenes-ordforrad--lareren-som-arbeidsleder-i-klasseron

https://www.nb.no/items/URN:NBN:no-nb_digibok_2009030900128?page=0

APPENDIX 3

Excel document with results from the vocabulary follow-up survey.

APPENDIX 4

The questions in the vocabulary follow-up survey

Here are the original wording of questions in Norwegian, and a translation into English:

Original Norwegian version:

Nokre spørsmål om å lære engelske ord

1) [Spørsmål 1 hadde ulike døme for kvar studieprogram:]

BA: Kor viktig er det å lære yrkesretta ord (slik som hacksaw, excavator, mortar og trowel)?
EL: Kor viktig er det å lære yrkesretta ord (slik som switch, soldering, IC og pylon)?

HE: Kor viktig er det å lære yrkesretta ord (slik som CPR, surgery, syringe og diagnosis)?

NA, RM: Kor viktig er det å lære yrkesretta ord (slik som harrow, baler, spice og cuisine)?

TIF: Kor viktig er det å lære yrkesretta ord (slik som wrench, welding, piston og caliper)?

2) Kor viktig er det å lære ord med vanskegrad 1 (slike ord som large, many, rich, to begin)?

3) Kor viktig er det å lære ord med vanskegrad 2 og 3 (ord som exhausted, multiple, numerous)?

4) Kor viktig er det å lære ord med vanskegrad 4, 5 og 6 (ord som to postpone, tranquil, affluent)?

5) Kor viktig er det å lære ord med vanskegrad 7 og høgare (slik som placid, minuscule, gargantuan)?

6) Kor viktig er det å lære "phrasal verbs" (samansette uttrykk slik som to go ahead, to put off, worn out)?

7) Det ordforrådet som eg brukar når eg skriv engelsk er ...
enkelt (med ord som large, many, rich, to begin).
middels (med ord som exhausted, multiple, numerous).
avansert (med ord som tranquil, affluent, placid).

8) Eg kan forstå engelske tekstar som har eit ...
enkelt ordforråd (med ord som large, many, rich, to begin).
middels ordforråd (med ord som exhausted, multiple, numerous).
avansert ordforråd (med ord som tranquil, affluent, placid).

English translation:

Some questions about learning English words

1) [Question 1 had different examples for each study program:]

BA: How important is it to learn vocational words (such as hacksaw, excavator, mortar and trowel)?

EL: How important is it to learn vocational words (such as switch, soldering, IC and pylon)?

HE: How important is it to learn vocational words (such as CPR, surgery, syringe and diagnosis)?

NA, RM: How important is it to learn vocational words (such as harrow, baler, spice and cuisine)?

TIF: How important is it to learn vocational words (such as wrench, welding, piston and caliper)?

2) How important is it to learn words on level of difficulty 1 (words such as large, many, rich, to begin)?

3) How important is it to learn words on levels of difficulty 2 og 3 (words like exhausted, multiple, numerous)?

4) How important is it to learn words on levels of difficulty 4, 5 og 6 (words like to postpone, tranquil, affluent)?

5) How important is it to learn words on level of difficulty 7 and higher (words like placid, minuscule, gargantuan)?

6) How important is it to learn "phrasal verbs" (complex expressions such as to go ahead, to put off, worn out)?

7) The vocabulary that I use when I write English is ...
simple (with words like large, many, rich, to begin).
average (with words like exhausted, multiple, numerous).
advanced (with words like tranquil, affluent, placid).

8) I can understand English texts that have …
a simple vocabulary (with words like large, many, rich, to begin).
an average vocabulary (with words like exhausted, multiple, numerous).
an advanced vocabulary (with words like tranquil, affluent, placid).

APPENDIX 5

List of texts that have been indexed

50 words of each text have been indexed. Unless otherwise stated, the indexed section of the text begins at the start of the 2nd paragraph.

Title	**Description and source**	**The first 50 words of the 2nd paragraph in the text** (Hyphenated word are counted as two separate words)
Sample exam text marked 2	This text is taken from a set of sample exam answers that were published by the Norwegian Directorate for Education and Training (Utdannings-direktoratet) in 2014. This section of the text begins at the start of the 2nd paragraph of Task 2, which was the longest task. (Utdanningsdirektoratet 2014)	All since I was a child I was interested to work as a carpenter. The best thing to working as a carpenter is for me, that I created something. As carpenter you get possibility to be at one workplace from the building started, till it finished. When you worked as
Sample exam text marked 3	This text is taken from a set of sample exam answers that were published by the Norwegian Directorate for Education and Training (Utdannings-direktoratet) in 2014. This section of the text begins at the start of the 2nd paragraph of Task 2, which was the longest task. (Utdanningsdirektoratet 2014)	The employees' tasks and working environment have changed a lot. The new generations who will take over job as petroleum engineers will have flexible work hours or days. If you work at an oil platform for example in Norway, you have 2 weeks of work and 4 weeks of vacation.
Sample exam text marked 4	This text is taken from a set of sample exam answers that were published by the Norwegian Directorate for Education and Training (Utdannings-direktoratet) in 2014. This section of the text begins at the start of the 2nd paragraph of Task 2, which was the longest task.	To be an engineer it requires that you are focused all the time and think of different type of solutions when you are working with your tasks, for example if you are a petroleum engineer, so some of your task would be to find a solution to find oil or

	(Utdanningsdirektoratet 2014)	
Sample exam text marked 5	This text is taken from a set of sample exam answers that were published by the Norwegian Directorate for Education and Training (Utdannings-direktoratet) in 2014. This section of the text begins at the start of the 2nd paragraph of Task 2, which was the longest task. (Utdanningsdirektoratet 2014)	Before Junior threw that calculus book he was just another student. After he threw it, he was the kid who broke a teacher's nose. He will never be just another kid again. He can't return to the before part and not throw that book. Junior decided to throw it, and
Sample exam text marked 6	This text is taken from a set of sample exam answers that were published by the Norwegian Directorate for Education and Training (Utdannings-direktoratet) in 2014. This section of the text begins at the start of the 2nd paragraph of Task 2, which was the longest task. (Utdanningsdirektoratet 2014)	First off. Some of you might think that what we do in the IT-world is nothing else than sit on our behinds all day and write away on the computer. Although that might have been the case some 15 odd years ago, it could not be more wrong today.
Electrical Wiring in Homes	This vocational text describes how electrical installation is done in homes. (Guldbrandsen et al. 2001, p. 104)	Fuses are safety devices. Their function is to prevent the overloading of wiring, thereby reducing the risk of fire. Old-style fuses consist of a thin wire inside a porcelain casing. This wire will melt if the current in the circuit exceeds a set level, thus interrupting the current and
From the Power Plant to the Plug	This vocational text describes the generation of electricity in a power plant, and how the electricity is distributed to the consumers. (Lundgren et al. 2020, Elektrofag, p. 202).	The generator is driven by the turbine. In the generator a coil of copper wires is rotated through a magnetic field. This generator will produce three-phase AC power. The transmission substation: Here large transformers step up the voltage to between 155 kV and 765 kV. This is to reduce
Welding	This is a small chapter in a vocational text called *Technical Trades* in the English book SKILLS for TIF. It introduces the occupation of welder in contrast	Basically, welding is the process of permanently joining metal parts, using heat. TIG, stick and MIG welding are three common types of welding, recommended for different needs and circumstances. Welding is an important part of the

	to other jobs, e.g. crane operator or logistics worker. (Lokøy et al. 2020, TIF, p. 268).	construction of ships, automobiles and aerospace vessels. It is also used in large structures
Something About Me	This text is an excerpt of the novel *Slam* by Nick Hornby. It is printed in both the electrical trades and the mechanics (TIF) versions of the *SKILLS* English book. (Lokøy et al. 2020, TIF, pp. 91-92).	I was going to say that maybe you should know something about me before I go off about my mum and Alicia and all that. If you knew something about me, you might actually care about some of those things. But then, looking at what I just wrote, you know
It's a Wonderful, Digital World?	This is a text about positive and negative aspects of social media. It is printed in both the electrical trades and the mechanics (TIF) versions of the *SKILLS* English book. (Lokøy et al. 2020, TIF, pp. 116-117).	In a recent survey among US teens, 81 percent say they feel more connected to their friends when using online platforms to communicate and share content. As many as 68 percent say using social media makes them feel as if they have emotional support when times are tough. In addition,
Adiós Hydraulics (ordinary version)	This is a short story by Elliott Neal Hester, about a flight where the airplane has lost the power of its hydraulic system. A somewhat adapted version is included in the mechanics (TIF) version of the *SKILLS* English book. (Lokøy et al. 2020, TIF, pp. 272-274).	Without hydraulics, our pilots flew the aircraft in a condition called manual reversion. It's like driving the world's largest tour bus without power steering. The plane responds sluggishly. Manipulating the steering wheel, or in this case, the yoke, requires a level of physical exertion that can make sweat pour from
Adiós Hydraulics – In Short	This is a shortened and adapted version of the text “Adiós Hydraulics”. (Lokøy et al. 2020, TIF, p. 275).	Without hydraulics, the plane was hard to steer. The scariest part was that the landing gear didn't deploy automatically. The captain had to crank down the landing gear by hand. A silence swept through the cabin. The sixty passengers were scared. They started talking to each other for comfort. They
Barack Obama’s speech at COP26	From Barack Obama’s speech at the COP 26 climate summit in Glasgow on 8 November 2021 (Obama 2021) The speech transcript does not have an actual paragraph structure. The first sections contain greetings and introductory remarks, therefore	And on Paris, our goal was to turn progress into an enduring framework that would give the world confidence in a low carbon future, an agreement where countries would update their emissions targets on a regular basis, an agreement that would help developing nations get the resources they need to

	one of the following sections was used.	
The Queen's Speech 2022	The Queen's Speech is authored by the British Prime Minister and the Prime Minister's cabinet. It outlines the government's policies for the year to come. (Johnson et al. 2022).	My Government will drive economic growth to improve living standards and fund sustainable investment in public services. This will be underpinned by a responsible approach to the public finances, reducing debt while reforming and cutting taxes. My Ministers will support the Bank of England to return inflation to its target.

APPENDIX 6

Excel document containing the indexing forms for the 14 example texts.