COLLOCATIONS AND TEACHING

Investigating word combinations in two English textbooks for Norwegian upper secondary school students

Lívia Cerqueira de Souza Hodne

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor, Professor Aud Solbjørg Skulstad, for her support, guidance, and for valuable advice. I also thank my co-supervisor, PHD candidate Gard Buen Jenset, for all the constructive feedback, practical tips, and help with corpus work.

I also owe much of what I know today as a teacher and as a student of English to Susan Harris de Mello, first my teacher, then my boss, and always a friend. She was the one who first introduced me to the phenomena of collocations and showed me their importance in English teaching and learning. I am grateful to my students for being patient during my busy weeks and for being positive to my new 'experiments', especially when I wanted to try out the ideas of the present study in my classes.

I thank my dear family and friends in Brazil, who have accompanied me and encouraged me during the writing of this thesis even from a long distance. I would also like to thank my Norwegian family and friends who have followed closely each step of the process. I wish to extend a special thank you to my father, who has always shown me unconditional love and has taught me everything I know.

Finally, I thank my husband, Arnstein, for his love and patience through the unsociable weekends and extra housework, and for helping me 'take a step back and think objectively' when I most needed it.

SUMMARY IN NORWEGIAN

Denne masteroppgaven omhandler undervisning av kollokasjoner i videregående skole i Norge. Kollokasjoner er anerkjent som et viktig fenomen i det engelske språket, og derfor har mer og mer oppmerksomhet blitt gitt til temaet i undervisningssammenheng. Til tross for dette virker det som om lærebøkene i Norge enda ikke har tilpasset forskning om korpus og kollokasjon til klasserommet.

Studien omfatter en undersøkelse av vokabularøvelser i to lærebøker som er brukt i engelsk på videregående skoler i Norge (studieforberedende utdanningsprogram). Det første målet er å undersøke om ord representert som isolerte ord i vokabularøvelsene er en del av kollokasjonene i tekstene de er hentet fra. To forskningsspørsmål har blitt formulert: Hvor mange av disse ordene dukker opp i kollokasjoner, og hvilke av disse kollokasjonene ville være nyttige for denne elevengruppen? Det andre målet er å komme med forslag til hvordan vokabularøvelser kan bli forbedret basert på Kunnskapsløftet og tidligere forskning. Det tredje målet er å foreslå andre måter hvordan kollokasjoner kan bli integrert i engelsk undervisningen.

Denne studien bruker et korpus og en kollokasjons-ordbok for å avgjøre om ordkombinasjonene funnet i tekstene er kollokasjoner. Korpusanalyse er også brukt som en metode for å se etter frekvensen og styrken på kollokasjoner, og for å finne kollokatorer av ord fra øvelsene.

Resultater viser at mer enn en tredjedel av ordene fra de valgte vokabularøvelsene er en del av kollokasjoner i tekstene de er hentet fra, og at de fleste av disse er nyttige for undervisningen. Til slutt har eksisterende kollokasjons-øvelser og forskning på fremmedspråklæring og korpuslingvistikk blitt brukt for å legge fram forslag til hvordan øvelsene kan bli forbedret og komplimentert. For å oppsummere, viser studien at kollokasjoner opptrer hyppig i skolebøker og de formidler relevant sosiokulturell informasjon. Den viser også at korpus kan brukes til å forbedre språklige øvelser, og at det både er mulig og hensiktsmessig å integrere kollokasjoner i språkundervisningen.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

BNC   British National Corpus
CALD   *Cambridge advanced learner's dictionary*
CLT   Communicative language teaching
COCA   Corpus of Contemporary American English
EFL   English as a foreign language
ELT   English language teaching
ESP   English for specific purposes
OCD   *Oxford collocations dictionary for students of English*
SLA   Second language acquisition

1. INTRODUCTION

Collocations appear to be widespread phenomena in English – expressions such as *highly qualified*, *a central feature*, and *receive severe criticism* abound in the language and their importance for teaching has been increasingly recognized. However, it seems that collocations have not yet been generally integrated into teaching material, and, as a consequence, not been given serious consideration in the English classroom. The present study intends to address some of these issues in connection to findings in corpus linguistics.

This chapter will first situate this study in relation to previous research, then the aims and scope will be outlined and the main methods explained. The chapter continues with a discussion of the salience of collocations in vocabulary teaching and it provides a definition of the term. Finally, the terminology used and the structure of the thesis will be presented.

1.1 Relevance of the present study

Since the advent of computerized corpora in the 1960s, research within corpus linguistics has demonstrated its potential not only for lexicography (e.g. Sinclair 1987; Hunston 2002) and language research (e.g. Biber et al. 1998; Carter & McCarthy 1999; Partington et al. 2004), but also as a resource in language teaching (e.g. Michael Lewis 1997; Hoey 2000; Osborne 2000; Yoon 2008). However, relatively few studies have associated corpora, teaching, and textbooks (e.g. Römer 2004), and to my knowledge there is no research which in addition treats vocabulary or, more specifically, collocations. Meunier & Gouverneur (2007) for example study phraseology in textbooks for English language teaching (ELT) but their focus is on more advanced learners. In fact, it seems that the use of corpora for teaching upper secondary school students have not yet received due consideration. According to Chambers (2005:121), most of the studies on corpora and teaching concentrate on the context of university education (e.g. Bernardini 2000, 2002). She also observes that there is scant evidence confirming that corpus consultation will become a complement to course books, hence the need for more research integrating corpora into language teaching (Cambers 2005:111).

In order to fill this gap, the present study investigates collocations in two textbooks used in upper secondary school in Norway and offers ideas on how corpora can complement vocabulary exercises and contribute to teaching practice.

1.2 Aims and scope

As a teacher of English in upper secondary school in Norway, I have come to notice that some textbooks tend to present isolated words in vocabulary exercises. Having learned about the importance of teaching and learning words in chunks, I pondered whether these words could have been taught in collocations instead.

Thus, my first aim is to investigate if words taught in isolation in vocabulary exercises are part of collocations in the texts they are taken from. Based on previous research pointing out the pervasiveness of collocations in the English language (Sinclair 1991; Michael Lewis 1993), I hypothesize that a high proportion of these words will appear in collocations in the texts. The following research questions will also be addressed: (a) What proportion of the words in the exercises appears in collocations in the texts? (b) Which of these collocations would be useful to teach Norwegian students in the first year of upper secondary school?

It has been attested by class experimentation (Michael Lewis 2000), textbooks (McCarthy et al. 2006a, 2006b), and linguistic research (Nesselhauf 2005) that including collocations in course material is both attainable and highly recommendable. On these grounds, my second aim is to suggest how the vocabulary exercises selected for the present study can be complemented based on existing teaching material on collocations and by applying previous research on corpora and foreign language vocabulary acquisition. On the same premises, the third aim is to suggest additional ways in which collocations can be integrated in vocabulary teaching. Thus, this study is situated within the field of applied linguistics. More precisely, it belongs to the field of English language education (didactics) and corpus linguistics in the sense that it attempts to contribute to language teaching material and practice through the use of corpus information.

The scope of this investigation is restricted to two English textbooks used by Norwegian pupils in the first year of upper secondary school. These books are *Passage* (Sørhus et al. 2006a) and *@cross* (Rodgers et al. 2006a), both adopted for the general studies program. The material consists of the *@cross* textbook and workbook, the *Passage* textbook and website, and the Corpus of Contemporary American English (COCA; Davies 2008).

1.3 Methods

Both quantitative and qualitative methods have been employed to achieve the aims outlined above. Regarding the first aim, the words from the vocabulary exercises selected from each textbook were initially checked in context to see if they were part of any word combinations. To determine if the combinations found in the texts are collocations, they were looked up in

the *Oxford collocations dictionary for students of English* (*OCD*; McIntosh et al. 2009) and their frequency and strength were checked in the corpus. Thereafter, the percentage of single words from the exercises that are part of collocations in the texts was calculated. Determining whether or not a collocation is useful for students is a more subjective matter and open to interpretation. For the purposes of this study, I have used the competence aims of the Norwegian national curriculum for the English subject (L06 2006b) and research on vocabulary acquisition and collocations to help suggest what types of collocations are useful for teaching students at the level in question. Regarding the fulfillment of the second aim, the corpus was searched to identify the most frequent combinations with the words from each exercise and their strength. The findings were then used together with previous research and teaching materials based on corpora to propose ideas on how to improve the vocabulary exercises. The same sources were used to give suggestions on how collocations can be integrated in language teaching. In sum, this study will hopefully show that vocabulary teaching can profit from corpus use and from placing greater emphasis on collocations.

1.4 Why collocations?

In order to situate collocations within the area of vocabulary teaching, I will take the concept of 'word' as a point of departure (see Carter & McCarthy 1988). What exactly does it mean to know a word? Nation (1990:31) proposes a list of the different aspects of knowledge needed, as presented in figure 1.1 below. The *R* and *P* stand respectively for *receptive* (listening and reading) and *productive* (listening, reading, speaking, and writing) knowledge:

Form	Spoken form	R	What does the word sound like?
		P	How is the word pronounced?
	Written form	R	What does the word look like?
		P	How is the word written and spelled?
Position	Grammatical Patterns	R	In what patterns does the word occur?
		P	In what patterns must we use the word?
	Collocations	R	What words or types of words can be expected before or after the word?
		P	What words or types of words must we use with this word?
Function	Frequency	R	How common is the word?
		P	How often should the word be used?
	Appropriateness	R	Where would we expect to meet this word?
		P	Where can this word be used?
Meaning	Concept	R	What does the word mean?
		P	What word should be used to express this meaning?
	Associations	R	What other words does this word make us think of?
		P	What other words could we use instead of this one?

Figure 1.1 What is involved in knowing a word (Nation 1990:31)

As illustrated in the table, Nation uses the four general classification criteria drawing from George (1983): 'form', 'position', 'function', and 'meaning'. Unfortunately, some of these aspects tend to be given more prominence in the foreign language classroom, like 'form' and 'meaning', while others are hardly ever mentioned, which seems to be the case of 'position' (grammatical patterns[1] and collocations). Nattinger affirms that the meaning of a word is closely related to the word's associations and argues that 'the whole notion of collocations is extremely important for acquiring vocabulary and has yet to be exploited to its full potential' (1988:70). As the table shows, collocations, as well as all other aspects of a word, can be taught in relation to both receptive and productive knowledge (Nattinger 1988:75), the difference between the two being in the type of activity dealt with in class.

Thus, the main reason why the present study lays emphasis on collocations is that although their importance is to a great extent acknowledged (Sinclair 1991; Cowie 1998a), not much attention has been given to them in the teaching context. In order to further elucidate the term, a literature review is provided in the next section.

1.5 What is a collocation?

McIntosh et al. define collocation as 'the way words combine in a language to produce natural-sounding speech and writing' (2009:v). However, there are various ways to define it and it seems that linguists and teachers have not yet agreed upon a common description. In line with Fontenelle, 'there does not seem to be any clear-cut, non-controversial definition of the term "collocation"' (1998:191).

As a means of clarifying the concept, a general distinction that needs to be made is, in Sinclair's terms, between *the open-choice principle* and *the idiom principle*. As Sinclair explains it, the open-choice principle is 'the normal way of seeing and describing language' (1991:109), usually referred to as a 'slot-and-filler' model where nearly any word could fill the slots. However, this principle does not explain the numerous constraints in language choice. For instance, there is no reason why we do not say **to put something on fire*, but English native speakers would agree that the appropriate collocation is to *set something on fire*. According to Sinclair, 'the principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments' (1991:110). This model of

---

[1] It is worth observing that the grammatical patterns of a word are not to be confused with what is generally described as 'grammar'. See Hunston (2002:137-169) for clarification and examples.

interpretation is illustrated not only by collocations but also by other types of fixed expressions such as idioms, proverbs, clichés, technical terms, jargon expressions, and phrasal verbs (Sinclair 1991:111). The idea regarding these principles is that they co-exist, but can not be employed simultaneously.

Although the term *collocation* is used in different senses by a number of authors (Halliday 1966; Sinclair 1991; Stubbs 1995; Moon 1998), two main views of the concept can be identified: the *phraseological approach* and the *frequency-based approach* (Nesselhauf 2005:12). According to the frequency-based definition, a collocation is said to be the co-occurrence of words at a frequency that is higher than expected if words were combined arbitrarily in a language. Its main representatives are Firth, Halliday, and Sinclair. In the phraseological definition, 'collocations are associations of two or more lexemes (or roots) recognized in and defined by their occurrence in a specific range of grammatical constructions' (Cowie 1994:3169). This tradition emphasizes the relation between lexical and syntactic patterning in collocations and it has its roots in the work of H. Palmer (e.g. 1933) and Hornby (e.g. Hornby et al. 1942). This view has been strongly influenced by Russian phraseology and some of its supporters are Hausmann (1989), Cowie (1998a), and Mel'čuk (1998).

Collocations are considered by both the phraseological and frequency-based approaches as combinations of one or more *lexemes* or *lemmas*. According to Carter's definition, lexemes are 'the basic, contrasting units of vocabulary in a language' (1987:6-7). He exemplifies this by saying that when we look up words such as *bringing* or *brought* in a dictionary, we look for the lexeme BRING. As a way to restrict my search, I have limited myself to collocations with only two lexemes. In the case of the collocations found in the textbooks, however, a few longer ones have been included (e.g. *get good grades*) since I am interested in the types of collocation from the texts that might be useful for students.

According to the phraseological approach, the lexemes of a collocation should be syntactically related and these relations usually fall into ten categories. The following are the ones set by Hausmann (1989:1010; in Nesselhauf 2005:22): adjective + noun (*bright light*), (subject-) noun + verb (*the wind howled*), noun + noun (*consumer goods*), adverb + adjective (*bitterly disappointed*), verb + adverb (*handle carefully*), and verb + (object-) noun (*run the risk*) (my examples). In addition, as suggested by Benson et al. (1997:ix), lexical words plus a preposition also constitute collocations, so that the other four categories would be: verb + preposition (*glance at*), noun + preposition (*interest in*), preposition + noun (*by accident*), and adjective + preposition (*angry at*) (original examples). The first six categories are usually

called 'lexical collocations' and the four others 'grammatical collocations' (Fontenelle 1998:192), the latter are sometimes also called 'colligations' (Bartsch 2004:24). These ten categories plus the lexical one adverb + verb (*simply vanished*) are used in the present study to classify the collocations from the textbooks and corpus (see chapter 3).

Still within the phraseological approach, Cowie arranges word combinations in four groups devised in terms of transparency and commutability. Nesselhauf (2005:14) summarizes Cowie's (1981) classification as follows:

> **Free combinations** (e.g. *drink tea*):
> - the restriction on substitution can be specified on semantic grounds [i.e. you can substitute *tea* by *coffee*, *water*, *juice*, etc.]
> - all elements of the word combination are used in a literal sense
>
> **Restricted collocations** (e.g. *perform a task*):
> - some substitution is possible, but there are arbitrary limitations on substitution [e.g. you can also say *do a task*, but not *make a task*]
> - at least one element has a non-literal meaning, and at least one element is used in its literal sense; the whole combination is transparent
>
> **Figurative idioms** (e.g. *do a U-turn*, in the sense of 'completely change one's policy or behaviour'):
> - substitution of the elements is seldom possible
> - the combination has a figurative meaning, but preserves a current literal interpretation
>
> **Pure idioms** (e.g. *blow the gaff*):
> - substitution of the elements is impossible
> - the combination has a figurative meaning and does not preserve a current literal interpretation

Authors adopting a phraseological approach often use the term *collocation* to refer only to 'restricted collocations' and other terms to label the other kinds (c.f. Benson et al. 1997).

Although pertinent, I find Cowie's classification of 'literal' and 'figurative' senses not sufficiently clear-cut. For example, are *constitutional monarchy* and *customs agent* (see chapter 4) free combinations or restricted collocations? If they are restricted collocations, which word is used in a non-literal sense? In contrast to Cowie, Conzett (2000:74) adopts a less rigorous approach, illustrating the idea of collocation as a continuum, as pictured below:

friendly dog	strong coffee	sibling rivalry	throw in the towel
←	**weaker**	**stronger**	→
old car	heavy smoker	mitigating circumstances	Stars and Stripes
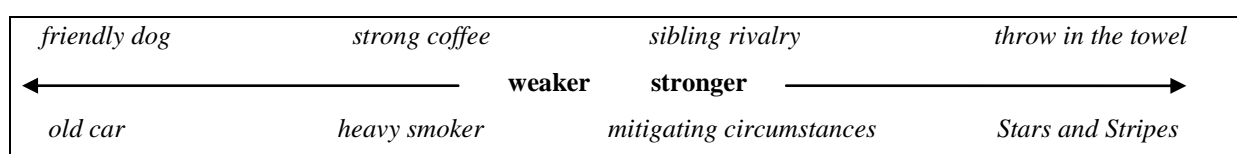
Figure 1.2 The concept of collocation as a continuum (Conzett 2000:74)

She excludes the two extremities (e.g. *friendly dog* and *throw in the towel*) and treats as collocations only the items in the middle of the continuum. Despite appearing simpler, this approach is also problematic: Where are the boundaries in this scale? Who or what decides where a combination should be placed? Carter (1987:55) suggests that collocational acceptability can be decided based on statistical measures of native-speaker intuitions. Church & Hanks (1990), on the other hand, recommend using mutual information (MI) with a score of 3.0 as a threshold, while McEnery et al. (2006:56) report that other statistical measures of association like *z*-score and *t*-score are also commonly used.

All the above considered, it is clear that linguists have not yet agreed upon the most accurate measure to define what a (restricted) collocation is. Therefore, I have decided to combine a number of elements to arrive at a definition appropriate for the present study. The *OCD* (McIntosh et al. 2009) is used as one of the tools to determine whether a word combination is a collocation or not. This dictionary was chosen due to its focus on usage (it is based on a 2 billion word corpus), its incorporation of syntactic relations (it classifies collocations in syntactical categories), and its concern with learners' needs (see McIntosh et al. 2009:v). Thus, it combines features of the phraseological and the frequency-based approaches and suits the pedagogical purposes of the present study. A disadvantage, however, is that this dictionary includes too wide a range of collocations, as defined in its aim:

> Give the full range of collocation – from the fairly weak (*see a movie, an enjoyable experience, extremely complicated*), through the medium-strength (*see a doctor, direct equivalent, highly intelligent*) to the strongest and most restricted (*see reason, burning ambition, blindingly obvious*) – for around 9,000 headwords (McIntosh et al. 2009:v).

Although it is explained further that 'totally free combinations are excluded and so, for the most part, are idioms' (McIntosh et al. 2009:vi), the dictionary includes combinations like *big difference*, *very important*, and *is impressive*. These, according to the substitution criteria set out by Cowie (above), would fit the category of free combinations and as such are not interesting for this study. As a means to exclude these combinations, I have used the MI threshold of 3.0 (see chapter 3), as provided by the corpus. The dictionary also includes under the label 'phrases' some idiomatic expressions of the kind Cowie calls figurative idioms. These were only included when the cultural element was considered highly important within the chapter, as the case of *member of parliament* in a chapter about government in the UK in @*cross* (see chapter 4). Otherwise they are not regarded as collocations in this thesis.

Another criterion established to decide whether a combination is a collocation was to check frequency. Drawing on Clear (1993:277), I have adopted the threshold of three occurrences in COCA to classify a pair as a collocation (see chapter 3). In addition, elements of the phraseological approach were included in that I have taken syntactic relations into account and have adopted Cowie's terminology of *(restricted) collocations* and *free combinations*. Thus, the definition of collocations adopted in this study is:

> Collocations are arbitrarily restricted lexeme combinations that are syntactically fixed to a certain degree, are included in the collocation dictionary, present an MI score higher than 3.0, and have a raw frequency of more than three tokens in COCA.

1.6 Terminology

In addition to the term *collocation*, some other terms will be used on a regular basis throughout this study. Here is a list of definitions of this terminology:

Collocate – a word which occurs in close proximity to the word under investigation without necessarily being part of a collocation (e.g. *very* is a collocate of *important* although *very important* is considered a free combination)

Corpus – a large amount of written and sometimes spoken material, usually in computerized form, collected to show the state of a language.

Concordance – a list of the uses of a word/lexeme in context provided by a corpus.

Foreign language – the language learned by students which is neither their mother tongue nor widely spoken in their country.

Word combinations – a general term used in this study to refer to all kinds of phraseological chunks (collocations, free combinations, idioms, phrasal verbs, fixed expressions, etc.).

A few other technical terms are also used, but are introduced and explained where relevant for the discussion.

1.7 Structure

This thesis is structured in six chapters. Chapter 2 provides some background for the teaching of vocabulary and presents previous research in the areas of collocation, corpora, and teaching. Chapter 3 describes the material and methods used, while chapter 4 presents the results from the analyses. In chapter 5 the results are interpreted in light of the background provided in chapter 2 and suggestions are given on how the exercises can be improved and how collocations can be included in vocabulary teaching. Finally, chapter 6 brings the study to a conclusion, summarizing the main findings and proposing ideas for further research.

2. BACKGROUND

2.1 Introduction

This chapter will first give an account of the teaching methods which have been most influential in Norway and how these methods treat vocabulary. Aspects of foreign language vocabulary acquisition and some of the previous Norwegian curricula will be considered. Collocations will also be discussed in relation to other areas of study such as cognitive linguistics, communicative language teaching, and culture. Lastly, relevant studies in corpus linguistics will be mentioned.

2.2 Teaching methods and the role of vocabulary

A number of authors have given surveys on the various teaching methods employed in the last few centuries (Larsen-Freeman 1986; Simensen 1998; Richards & Rodgers 2001; Howatt 2004) and on how vocabulary has been taught in these methods (Zimmerman 1997; Schmitt 2000). This section will give a description of the teaching methods which were most central in Norway in recent centuries and the role of vocabulary in each of them.

The first predominant method to be implemented in Norwegian schools was the *Grammar-Translation Method*. It was developed based on procedures used for teaching Latin and evolved out of the need to systematize foreign language teaching for school children (Howatt 2004:151). Zimmerman (1997:5) reports that this method was introduced in Prussia at the end of the eighteenth century and its main aims were to enable students to read classical literature and pass standardized exams. It was not expected that students would ever come to speak the languages they were studying, but that they would profit intellectually from the learning. Students were given extensive grammatical explanations in their own language and bilingual vocabulary lists, which were to be learned to help them translate classical texts. In this method vocabulary was introduced mainly as a way of illustrating a grammar point and when vocabulary difficulties arose, they were usually addressed by means of etymology. According to Steinberg & Sciarini, the grammar translation method 'has enjoyed and continues to enjoy acceptance in many countries around the world' (2006:141), principally in countries where language teachers are non-fluent and the classes are very large. Despite its advantages and the fact that it was widely used in Europe and in the United States well into the twentieth century, the method has drawn harsh criticism. The main objections have been the disregard of 'realistic, oral language' (Zimmerman 1997:6) and the focus on language

analysis instead of language *use* (Schmitt 2000:12).

As a consequence, by the end of the nineteenth century a new movement emphasizing listening and pronunciation appeared. This was known as the Reform Movement and one of its great achievements was the development of phonetics and the recognition of it as a science. Its members also defended the study of 'connected texts rather than unconnected sentences and lists of isolated words' (Simensen 1998:26). As a product of this movement, a use-based method emphasizing listening was introduced towards the end of the nineteenth century: the *Direct Method*. According to Zimmerman (1997:9), explicit grammar teaching and translation were set aside, while students were expected to learn English through the same process as native speakers do. They would listen first, and then speak, and only in later stages would they learn to read and write. Vocabulary was thought to be generally acquired through interactions in the classroom by asking and answering questions. Concrete words were taught using pictures, mimic, and *realia* (real objects) while associations of ideas were used to teach abstract vocabulary (Zimmerman 1997: ibid.). This method, however, also had its weaknesses. As Schmitt (2000:12) points out, teachers were not always proficient in the target language, failing to give students the input needed. Moreover, unlike L1 acquisition, in L2 acquisition students only have a few hours of exposure to the target language per week. This limited instruction time was one of the factors taken into account by the 1929 Coleman Report in the United States. This report concluded that the time of instruction was not enough to the development of comprehensive language proficiency and recommended that secondary students should instead be taught how to *read* in a foreign language (Schmitt 2000:13). According to Richards & Rodgers, this stress on reading 'led to the development of principles of vocabulary control, which were to have a major practical impact on the teaching of English in subsequent decades' (2001:37). It was at this point that lists of vocabulary based on frequency counts began to emerge (e.g. West 1953).

During the Second World War, it became clear that the approaches above did not manage to form competent users of the target language. The American military lacked people who were fluent in foreign languages and needed a teaching program that could train the soldiers rapidly in oral skills (Larsen-Freeman 1986:31). Structuralist linguist Charles Fries developed a method based on behaviorist principles (forming language habits) and on the Direct Method (emphasizing listening and speaking) which was first called the *Army Method* and later known as *Audiolingualism.* In this method new words were only introduced in drills, and vocabulary was thought to be acquired naturally through good language habits. It was also suggested that beginners should not learn too much vocabulary to prevent a false sense of

security (Zimmerman 1997:11). A parallel approach was used in Britain around the 1950s (although being developed as early as the 1920s) called *Situational Approach* or the *Oral Method.* The name derived from the idea of teaching language in sentence patterns replicating real situations. As Richard & Rodgers put it, 'language was viewed as purposeful activity related to goals and situations in the real world' (2001:40). Vocabulary was chosen to best illustrate and practice the sentence patterns and was presented as lists in substitution tables.

In spite of their appeal in language teaching, the behaviorist ideas of habit formation were severely attacked in the 1950s by Chomsky, who claimed that language was partly innate and governed by abstract rules (see Chomsky's review of Skinner 1959). As explained by Simensen, he 'claimed that the stimulus-response and conditioning theory of the behaviorists could not explain the creativity involved in generating all kinds of *new* utterances' (1998:81; author's emphasis). Even though Chomsky brought back to attention the creative aspect of language, he gave vocabulary scant importance. In reaction to Chomsky's notion of an autonomous linguistic competence, Hymes (1972) advanced the concept of *communicative competence*, emphasizing the relevance of social interaction to language learning (Zimmerman 1997:12). A new approach was developed from this notion and became known as *Communicative Language Teaching* (CLT). Elements of this approach, such as communication and cultural knowledge, are clearly emphasized in the latest curricula in Norway (Simensen 1998:118-119). Even though it is a meaning-based approach, vocabulary is again given a secondary status. According to Sökmen (1997:237), in the communicative approaches the priority is *incidental* vocabulary learning, so that students are encouraged to guess from context, use monolingual dictionaries, and avoid translation. CLT will be treated in more detail in subsection 2.5.4.

In the 1980s *The Natural Approach* was developed by Krashen & Terrell (1983) to aid beginners in achieving the ability to communicate orally in the classroom. It is based on five hypotheses, and '[its] methodology emphasizes comprehensible and meaningful input rather than grammatically correct production' (Zimmerman 1997:15). It follows that vocabulary, as relevant language input, is considered essential to the acquisition[2] process. As for more advanced students, Krashen (1989) suggests reading as the best way to develop wider vocabulary knowledge.

Although treated separately in most teaching methods, recent evidence from corpora has revealed that grammar and vocabulary are fundamentally linked. Lexical patterning is

---

[2] The terms *acquisition* and *learning* as defined by Krashen are not differentiated in the present study, but used interchangeably.

shown to exist on a much larger scale than ever imagined, making it difficult to separate grammar from lexis. In line with Schmitt, 'one must conceptualize them as partners in synergy with no discrete boundary, sometimes referred to as *lexicogrammar*' (2000:14). This idea gives vocabulary a more central place and is one of the greatest contributions corpus studies have made to the field of language teaching. Some authors give lexis an even greater role, and are the advocates of a lexical approach. According to Richards & Rodgers, this approach is 'derived from the belief that the building blocks of language learning and communication are not grammar, functions, notions, or some other unit of planning and teaching but lexis, that is, words and word combinations' (2001:132). Some of the attempts to integrate this view into teaching can be seen in *The lexical syllabus* (Willis 1990), *Lexical phrases and language teaching* (Nattinger & DeCarrico 1992), and *The lexical approach* (Michael Lewis 1993). Albeit not yet diffused in Norway, these lexical approaches are closely related to the topic of the present study, i.e. the phenomenon of collocation.

2.3 Foreign language vocabulary acquisition[3]

In this section, aspects of foreign language vocabulary acquisition relevant to the present study such as implicit and explicit learning, the different steps in the vocabulary acquisition process, vocabulary learning strategies, and the influence of L1 in L2 learning will be presented. Although pertinent under this topic, the role of memory will be discussed in connection with cognitive linguistics (see subsection 2.5.2).

One of the contentious issues within the field of foreign language vocabulary acquisition is implicit (incidental) versus explicit (direct) vocabulary teaching and learning. In line with Sökmen, 'the pendulum has swung from direct teaching of vocabulary (the grammar translation method) to incidental (the communicative approach) and now, laudably, back to the middle: implicit *and* explicit learning' (1997:239; emphasis added). Although these tendencies have developed in general chronologically, different views still coexist: while some advocate the (implicit) acquisition of vocabulary mainly by guessing from context (Nagy et al. 1985; Krashen 1989), others have realized the importance of combining the two approaches (Coady 1993), particularly in recognition of learners' individual strategies (Hulstijn 1993; Nation 2001). Despite its great impact on language teaching since the 1970s, CLT has been criticized for prioritizing implicit vocabulary teaching. Some of the criticisms are whether beginners know enough vocabulary to guess from context (Coady 1997:227) and

---

[3] Although I acknowledge the difference between the terms *second language* (L2) and *foreign language*, they are used in this section in free variation since I quote authors who do so (e.g. Swan 1997).

the fact that incidental learning has not shown to be more effective than other learning strategies such as the keyword method (Moore & Surber 1992). Moreover, other studies have proved that although reading for meaning has increased L2 vocabulary acquisition, direct instruction has reached even better results (Paribakht & Wesche 1993; Zimmerman 1994). These studies suggest that promoting a mixture of direct and incidental learning combined with learning strategies is the most fruitful way of teaching vocabulary.

Concerning the way in which new words are acquired, Hatch & Brown explain the five steps in the vocabulary acquisition process (based on Payne 1988): (1) encountering new words, (2) getting the word form, (3) getting the word meaning, (4) consolidating word form and meaning in memory, and (5) using the word (Hatch & Brown 1995:374). Thus, as the first step of the acquisition process, learners will encounter new words for example on the internet, television, music, books, movies, magazines, etc. In the case of the present study, the words from the exercises selected are encountered by students in the course book texts or listening activities. Hatch & Brown (1995:417) provide examples of how the other steps can be achieved. For example, getting the word form (2) is the step when students connect the sound with the word form, which can be done by reading the word aloud. As a way to get the word meaning (3), students can use bilingual or monolingual dictionaries or guess from context. Consolidating word form and meaning in memory (4) can be achieved by doing a number of exercises such as matching or translating words, filling in gaps, crossword puzzles, memory games, etc. Finally, to learn the uses of a word (5), students might be asked to write sentences or texts, answer questions that require the use of the word, or even consult concordances to check how the word is used (Hatch & Brown 1995:417-419). Most of the exercises selected in the present study are concerned with the fourth step and only a few with the fifth.

A way of achieving all these steps is to make use of various learning strategies. Nation (2001:218) presents a table with different kinds of vocabulary learning strategies than can be encouraged in class, as reproduced in table 2.1 below (the explanation that follows is also based on Nation 2001):

Table 2.1 Types of vocabulary learning strategies (Nation 2001:218)

General class of strategies	Types of strategies
Planning: choosing what to focus on and when to focus on it	Choosing words
	Choosing the aspects of word knowledge
	Choosing strategies
	Planning repetition
Sources: finding information about words	Analysing the word
	Using context
	Consulting a reference source in L1 or L2
	Using parallels in L1 and L2
Processes: establishing knowledge	Noticing
	Retrieving
	Generating

In table 2.1, the first strategy regarding 'planning' is 'choosing words'. This is related to the vocabulary goals of the learner or of the syllabus being used; consulting frequency lists for instance might be useful to select words to be learned. The second strategy, 'choosing the aspects of word knowledge', means deciding which of the aspects listed in figure 1.1 (form, position, function, and meaning) should be given prominence. Examples of 'choosing strategies' can be to refer to a dictionary or make word cards. The fourth is 'planning repetition' and this is when learners create a system for revising old material.

Under 'sources' the first strategy is to analyze words. This can be done by looking at word parts, for example affixes and stems. The second is that learners should use context, drawing on background knowledge and guessing meaning from clues in the text. The third is 'consulting a reference source in L1 or L2', those can be dictionaries, concordances, teachers, native speakers, etc. The fourth is using parallels from other languages to help learning a new item (this will be developed further below).

Under 'processes' the first strategy is noticing words to be learned according to one's needs. The second is 'retrieving', i.e. revising previously stored items by using cues (students can for instance cover the word form and remember it by looking at its meaning). Finally, the third strategy under 'processes' is 'generation'. This type includes, among others, trying to visualize examples of the word, 'creating contexts, collocations and sentences containing the word, mnemonic strategies like the keyword technique, and meeting and using the word in new contexts across the four skills of listening, speaking, reading and writing' (Nation 2001:222). All of these strategies can be encouraged in class to facilitate student's vocabulary acquisition. However, it is worth recalling that this taxonomy is only one of the many ways vocabulary learning strategies can be classified; other examples are to be found in Schmitt (1997) and Takač (2008).

Another relevant aspect of vocabulary acquisition for the present study is the use of L1 in language teaching and learning. Natural methods like the Direct Method banned the use of L1 in class due, among other reasons, to the distrust in a one-to-one correspondence between languages (Howatt 2004:313). The idea that there are exact equivalents can indeed be problematic when we consider the complex relationships that exist between words in different languages. Swan (1997:157-160) lists some these relationships as follows: words in various languages can have different grammatical contexts or collocations; equivalent concepts can be assigned to different parts of speech; there can be false cognates and differences in style and levels of formality; and the notion of a 'word' might vary, i.e. a word in a language can be translated to three or four in another. One way to avoid mistakes caused by these variations is to follow Morgan Lewis' suggestion that an item (such as *widely available* or *catch up with the news*) should be translated 'not word-for-word but whole phrase to whole phrase, bearing in mind that the structure of the expression may be very different in one language from the equivalent expression in the other' (2000:16). At the same time, Takač (2008:9) defends the importance of the L1 to the learning of the L2 in the sense that, by establishing equivalents between the two languages, learners do not need to relearn how to categorize the world. Hence, translation in vocabulary teaching should be used when it enhances learning and avoided when it prevents it.

2.4 Vocabulary in Norwegian curricula

In 2006, a new national curriculum was introduced in Norway presenting revised guidelines in different school subjects. This curriculum is called *Læreplanverket for Kunnskapsløftet*, or Knowledge Promotion (2006a), and will be referred to as L06 from now on in the present study.

Compared to the previous curriculum (L97), L06 (2006b) includes five basic skills in the English subject to be focused on at all levels of schooling. These are:

- Being able to express oneself in writing in English
- Being able to express oneself orally in English
- Being able to read English
- Having skills in mathematics in English
- Being able to use digital tools in English

The introduction of the last skill is particularly important in that it was accompanied by the sponsoring of personal computers for all students in upper secondary school (the government

covers part of the cost). According to the Directorate for Education and Training (*Utdanningsdirektoratet*; Udir 2007), the plan is that from the fall of 2009 every Norwegian upper secondary school student will have a personal laptop (and internet) to be used in class. This development has opened new possibilities for the teaching of English in many ways since students now have access to online newspapers, blogs, encyclopedias, dictionaries, forums, corpora websites, among other language resources.

In addition to a description of the five basic skills, L06 provides a list of competence aims that should be achieved by the end of certain school years, among them the first year of upper secondary school (the list of aims is provided in appendix 1). These aims are divided under three main subject areas: 'language learning', 'communication', and 'culture, society and literature'. The only reference to vocabulary in the whole curriculum is under the area of 'communication' and it is simply stated that students are expected to 'master a wide vocabulary' (L06 2006b). This is a rather general statement and gives little or no guidance to teachers when it comes to this important part of the teaching and learning of a language. A comparison of L06 with previous curricula reveals that they have become less and less specific about vocabulary. Simensen reports that between 1950 and 1975, 'a systematic approach to the selection, grading, and repetition of vocabulary was […] given high priority' (1998:59). According to Daasvand (2001), the curricula from 1957 and 1974 contained word lists to be learned and the curriculum from 1960 recommended the teaching of vocabulary based on frequency lists. She also observes a decreased attention to vocabulary in the curriculum from 1987 (2001:33-34), while Vestre (1998:9) remarks that in L97 vocabulary work was mentioned only five times. The reason for this development might be that in a time when English is the main language of games, movies, TV-series, music, and information, deciding on a vocabulary list could limit this great window of possibilities. At the same time, it is left to the teacher to decide what kind of vocabulary should be highlighted, which might not be an easy task considering the amount of texts and new words students often encounter.

One way to tackle these problems is calling students' attention to useful vocabulary appearing in the texts from textbooks. Conzett affirms that 'in-context study of vocabulary encountered in reading has the well-known advantages of point-of-need relevance to the student, and natural, real-life examples of usage' (2000:72). Woolard (2000:33-34) suggests that students should learn how to notice *collocations* in the texts they read, and, more importantly, they should be able to select the ones that they need. This is a learning strategy that can aid learners in widening their vocabulary knowledge.

2.5 Collocations

A definition of collocation and the reasons why it has been chosen as the topic of this thesis were provided in chapter 1. In the present chapter collocations will be considered in relation to other relevant fields of study.

2.5.1 Background

As pointed out in the introduction, the definition of collocation is a matter of heated dispute, and it seems like each scholar chooses to define it either in overly general terms or according to the specific aims of their study. What is nonetheless agreed upon is the value of collocational knowledge for linguistic research and pedagogical purposes (Ellis 1997:128-129).

'You shall know a word by the company it keeps' (Firth 1968:11). This is one of the most common quotations used when scholars attempt to define collocation. Firth brought the term to the linguistic arena in his paper 'A synopsis of linguistic theory, 1930-1955' in which he frequently alluded to 'the collocational level' of meaning. He believed that we obtain the meaning of a word or a text through a 'mutually congruent series of levels' (Firth 1968:30), these levels being contexts of situation, syntax, phonology, phonetics, and *collocation*. Although the term came to be known mostly through Firth, H. Palmer had already commented on it several years before, saying that each collocation 'must or should be learnt, or is best or most conveniently learnt as an integral whole or independent entity, rather than by the process of piecing together their component parts' (1933:4; in Nation 2001:317). In fact, Bartsch (2004:28) notes that the first recorded use of the term is a quotation from the year 1750, but that then it simply meant a grammatical relation between words.

In spite of its early origin, the term *collocation* has only been given more attention after the development of electronic corpora in the 1960s and with it the opportunity to observe combinational patterns in the English language. Since then, research on collocations has increased substantially. Some authors even defend the view that knowledge of language depends on collocational knowledge (Ellis 2001) or that for a speaker to be fluent and use a language appropriately collocational knowledge is required (Pawley & Syder 1983). However, even though the widespread use and importance of collocations in the language is recognized by many (Kjellmer 1984; Stubbs 1995; Nation 2001), few attempts have been made to integrate the teaching of collocations in the English learning curriculum. Some exceptions are Michael Lewis, who has developed the Lexical Approach (1993), and McCarthy et al. (2006), authors of the *Touchstone* material that uses corpus information and

includes collocations in the vocabulary work. Another important contribution is from Nesselhauf (2005), who studies collocations in a learner corpus and based on the results provides suggestions on how to select collocations for teaching.

The next subsection presents research on why it is conceded that teaching and learning words in chunks is more effective.

2.5.2 Cognitive linguistics, chunking, and memory

Cognitive linguistics is a large field of study with its own principles and theories. Since the present thesis is not based on one theory but takes an eclectic approach, cognitive linguistics will be specifically related to collocations and teaching.

A common assumption in cognitive linguistics is, according to Taylor, 'the belief that language forms an integral part of human cognition, and that any insightful analysis of linguistic phenomena will need to be embedded in what is known about human cognitive abilities' (2002:4). Some of these cognitive abilities are, as described by Schmid (2007:117), perception, memory, and attention allocation. Among these, memory is particularly interesting in connection with collocations, as explained by Nattinger:

> Finally, one of the earliest findings from memory research was that short term memory holds a fairly constant number of units (Miller 1956), units which later research has shown likely to be 'chunks' of information, composed of several rather than single items. [...] Since a great part of the learner's task is to chunk unfamiliar material in meaningful ways, the teacher who makes this chunking easier increases the number of items the learners retain. (Nattinger 1988:64)

A point worth noting here is that Nattinger has a broad interpretation of collocations, including phrases such as *How do you do?*, which is different from the definition adopted in the present study. Nonetheless, since collocations are also part of the more general idea of 'chunks', his discussion of storage of chunks versus single items in memory is highly relevant.

Nation (2001) also talks about the role of chunking in the language users' memory. He explains the advantages and disadvantages of chunking and affirms that, 'the main advantage of chunking is reduced processing time. [...] Instead of having to refer to a rule or pattern to comprehend or produce the chunk, it is treated as a basic existing unit' (Nation 2001:320). This idea is closely related to Sinclair's idiom principle, which establishes that language users do not process all of the options available when uttering a sentence; they simply use ready chunks that have been heard and employed many times before. Nation and Sinclair's views

are directly connected to two of the various concerns of cognitive linguistics – *automatization* and *storage vs. computation*. According to Taylor (2002:13-14), these cognitive capacities can be perceived in language in that a great amount of what we say is automated, which is possible because stored chunks can be recalled rapidly and effortlessly, in contrast to the slow process of computation. He argues that the fact that rules can be applied in language to produce the correct input (computation) does not entail that they are always used by speakers, who often rely on ready-made language forms (Taylor 2002:14).

2.5.3 Language teaching and learning

'It is not true that our students necessarily learn what we teach them' (Morgan Lewis 2000:11). Morgan Lewis is categorical in his claim that teaching does not cause learning, which is also asserted by Larsen-Freeman (1997). According to him, while teaching is linear and step-by-step, learning is non-linear and cyclical (Morgan Lewis 2000:12). In consequence, students might only learn a taught item many weeks, months, or even years later, after he or she has encountered it a great number of times. Therefore, teachers should provide opportunities in which students can repeatedly meet the target vocabulary. One way of doing this, as suggested by Morgan Lewis and endorsed by Nesselhauf (2005:259), is that instead of prioritizing teaching new and rare words, teachers should concentrate on expanding knowledge of what is only 'half-known' by teaching students new collocates of a known-word. As he affirms, 'time spent on half-known language is more likely to encourage input to become intake than time spent on completely new input' (Morgan Lewis 2000:24).

Hill (2000) also promotes the idea of teaching a word with its most common collocates, explaining that most intermediate students know for example the words *hold* and *conversation*, but that they may not know the collocation *hold a conversation* (Hill 2000:64). He argues that 'the main thrust of vocabulary work in most classes should be to make students more collocationally competent with the words with which they are already partly familiar' (Hill 2000:67). What would then be this 'familiar language' and its collocations? Hill et al. (2000) recommend teaching the collocates of nouns that do not have very specific meaning and which are often accompanied by an adjective. These are also the nouns that present the longest entries in a collocation dictionary. Here is the list that they propose:

> *Account, action, answer, approach, argument, behaviour, change, circumstances, condition, consequences, decision, difference, discussion, effect, feature, idea, information, interest, issue, manner, method, move, performance, plan, policy, position, problem, programme, project, question, reason, relationship, result, scheme,*

situation, solution, state, story, style, system, theme, theory, use, view, vision, way, work. (Hill et al. 2000:101)

They choose the word *situation* from this list and exemplify how collocations can be used in a classroom activity presenting the following: *an embarrassing situation*, *a bewildering situation*, *a tricky situation*, *a unique situation*, *an extraordinary situation*, and *a tense situation*. Then they ask students to think of an example from their own life in which these collocations would apply. This idea is in accordance with Morgan Lewis's argument that a collocation 'can evoke a complex situation very precisely', as in the cases of *routine check-up*, *disperse the crowd*, and *widely available*, affirming that when we teach these items apart they lose their power to communicate (2000:15).

In a study of phraseology in learners' academic writing, Howarth (1998:181) has found that students consider de-lexicalized verbs like *get*, *put*, *take*, *do*, and *make* problematic and that they avoid word combinations with these verbs because of their uncertainty as to the appropriate collocates. This indicates that de-lexicalized verbs are another category that should be drawn attention to in class. If this type of collocations is critical for university students it is safe to assume that it is no less of a problem for upper secondary school learners.

In face of the various challenges in language teaching and learning, many claim that teachers should be more concerned about providing learners with strategies (see section 2.3 above) than using all classroom time on language items. This concurs with L06, which defines as one of the competence aims that students should 'exploit and assess various situations, work methods and strategies for learning English' (2006b). Michael Lewis emphasizes the same when discussing teachers' roles:

> Either teachers must select and teach a restricted lexicon – but on what criteria, for students of general English? – or they must adapt classroom activities so that, rather than teaching individual items, they provide learners with strategies which ensure that learners get the maximum benefit from all the language they meet in and, more importantly, outside the formal teaching situation. (Michael Lewis 2000:158)

As mentioned above, Woolard also recommends turning students' attention to collocations as a learning strategy. He reports from his teaching experience that in the beginning the teacher will have to point out useful language in texts or listening and get students to record it. In the long-run, however, students will be trained to recognize these chunks themselves inside and outside the classroom, thus becoming more independent learners (Woolard 2000:33-36).

2.5.4 Communicative language teaching and the Framework

After Hymes (1972) coined the term *communicative competence* a great number of methods and approaches were developed from its premise that knowledge of a language includes the 'competence as to when to speak, when not, and as to what to talk about with whom, when, where, in what manner' (Hymes 1979:15). Communicative language teaching (CLT) is the name given to these approaches and it has had significant repercussion all over the world.

There are, however, two distinguished versions of CLT, often called the 'weak' and the 'strong' versions. Simensen explains that 'the "weak" version implies learning to communicate in the target language' while 'the "strong" version implies communicating in the target language in order to learn it' (1998:113). In the strong version the texts used in class are to be authentic, i.e. 'prepared for native speakers of the language' (Hatch & Brown 1995:407). In fact, the use of authentic versus constructed texts has been a contentious issue in the last decades. With an increase in research pointing to the deficiencies of prepared materials, authentic texts have been given a more central role in language teaching (Simensen 1998:142). In Norway, authentic texts have been recommended for all levels since the 1987 curriculum. Nonetheless, this emphasis has been repeatedly criticized on various grounds. Howatt raises for example the question of whether native-like models are the most desirable for teaching, since:

> […] most current uses of the language are for international communication between people who are not native speakers – a trend that looks likely to continue – and effective communication in such lingua franca uses of English does not, it would appear, depend on conformity to native-speaker norms. (Howatt 2004:360)

In a time when English is used in such international contexts, some might criticize the teaching of collocations, inasmuch as it can imply that students should learn them to sound like native speakers (see McIntosh et al. 2009:v). However, the aim of teaching collocations is essentially to show how the language is commonly used so that students can learn to communicate fluently and accurately (Michael Lewis 2000:177). Thus, learning collocations should be about communicating effectively and not about sounding like a native speaker.

The use of authentic texts and their value in CLT is also questioned in relation to the relevance of using concordance lines in the classroom. Widdowson (2000) argues that they are decontextualized samples of the language and as such have lost their communicative purpose. On the other hand, as shown by Stubbs (2001b), this does not undermine the value of corpus consultation in the classroom to identify patterns and routines and show students what

is typical in the language. Michael Lewis (2000:196-199) defends the use of authentic examples from corpora, but warns that they should be selected by the teacher in advance to avoid overwhelming the learner by the great amount of data and number of unknown words.

Another important feature of CLT is the use of *communicative activities* in which students are supposed to complete a certain *task*. The Council of Europe issued in 2001 the *Common European framework of reference for languages*, which gives general guidelines for the teaching of English as a second or foreign language in Europe (referred to from now on as the Framework). According to these guidelines, *activity* is defined as 'the exercise of one's communicative language competence in a specific domain in processing (receptively and/or productively) one or more texts in order to carry out a task' (Council of Europe 2001:10), whereas *task* is defined as 'any purposeful action considered by an individual as necessary in order to achieve a given result in the context of a problem to be solved, an obligation to fulfil [*sic*] or an objective to be achieved' (Council of Europe 2001: ibid.). Despite not being much used in CLT, the term *exercise* is the one mostly used in the present study due to the type of material selected. This word is defined by *CALD* as 'a short piece of written work which you do to practice something you are learning' (2005:432).

The Framework has widely influenced both Norwegian textbook writers and L06. Chapter 5 of the book is particularly interesting regarding CLT and collocations, as it describes the user/learner's competences and shows that collocational knowledge is included within the linguistic competence. An overview of the main competences is given in table 2.2 below (summarized from Council of Europe 2001:101-130):

Table 2.2 An overview of the user/learner's competences in the Framework

User/learner's competences			
General competences	**Communicative language competences**		
knowledge of the world	linguistic competence	lexical	lexical elements (table 2.3)
			grammatical elements
sociocultural knowledge		grammatical	
intercultural awareness		semantic	lexical semantics (table 2.4)
			grammatical semantics
others		phonological	
		orthographical	
		orthoepic	
	sociolinguistic competence		
	pragmatic competence		

As presented by table 2.2, the general competences comprehend areas like 'sociocultural knowledge', 'intercultural awareness', 'knowledge of the world', among others, while the communicative language competences are divided into: 'linguistic competence',

'sociolinguistic competence', and 'pragmatic competence'. It is worth recalling that the Framework contains other subdivisions, which were not included in the table for not constituting relevant information for the discussion. As shown in table 2.2, the linguistic competence is subdivided into six different competences as shown in the table. The first one, the lexical competence, is divided into 'lexical elements' and 'grammatical elements' and it is defined as having knowledge of the vocabulary of a language and being able to use it (Council of Europe 2001:110). It is within 'lexical elements' that collocations are included. Table 2.3 below has been compiled based on the classification of these elements in the Framework (Council of Europe 2001:110-111):

Table 2.3 Classification of lexical elements compiled from the Framework

Lexical elements		Examples
a) Fixed expressions	sentential formulae	*How do you do?*
	phrasal idioms	*It's a long shot*
	fixed frames	*Please may I have…?*
	other fixed phrases such as phrasal verbs and compound prepositions	*to put up with* *in front of*
	fixed collocations	*to make a speech/mistake*
b) Single word forms		*tank*

Looking at this classification, we can infer that collocations are as important within the lexical elements as idioms and phrasal verbs, which usually receive greater attention in language teaching. Howarth, in his study of collocations in academic writing, notes the usual disproportionate concern with idioms:

> Of all categories of conventional lexical combinations, idioms have received the most attention in linguistic theory and description […]. Idioms have also been almost the only phraseological category to be recognized in ELT materials (though consider Rudzka *et al*. 1981, 1985 and McCarthy 1990). However, they are the least frequent category in the type of texts under discussion here, and arguably present less severe problems to learners. Far more significant is the central area of the spectrum – restricted collocations. (Howarth 1998:169)

Cowie (1998a:212) comments that H. Palmer showed reluctance to the type of idioms such as *put one's shoulders to the wheel* and *skate on thin ice* – that is, the ones that are structurally complex and culturally marked. Parallel to Howarth's findings, Cowie (1998a: ibid.) reports that computer-based studies show that restricted collocations are much more frequent in written texts than these types of idioms.

Collocations are mentioned once more in the Framework within the 'semantic competence' in the category of 'lexical semantics' (see the linguistic competences listed in

table 2.2). The elements of this category are compiled from the Framework (Council of Europe 2001:115) in table 2.4 below.

Table 2.4 Classification of lexical semantics compiled from the Framework

Lexical semantics	
Relation of word to general context	reference
	connotation
	exponence of general specific notions
Interlexical relations	synonymy/antonymy
	hyponymy
	collocation
	part-whole relations
	componential analysis
	translation equivalence

Again the Framework shows 'collocation' parallel to other relations that are usually given a greater role in textbooks and in the English classroom, as it is the case of synonyms, antonyms, and translation equivalence. These are practiced in exercises commonly used to teach learners the meaning of a word; however, they are not the only way to do so. Teaching students the collocates of a word can aid them in understanding the word meaning and in communicating more effectively. Hunston claims that 'one use of collocational information is to highlight the different meanings that a word has' (2002:76). Earlier in her book she explains this taking the word *shed* as an example:

> [The meaning of *shed*] is something like "lose" or "give", but the precise meaning of each phrase depends on the collocate:
> *shed light (on)* means "illuminate", usually metaphorically;
> *shed tears* means "cry" (literally) or "be sorrowful" (crying metaphorical tears);
> *shed blood* means "suffer" or "die", either literally or metaphorically;
> *shed jobs* and *shed staff* mean "get rid of people";
> *shed pounds* means "lose weight";
> in *shed skin* and *shed clothes*, *shed* means "remove";
> *shed cents* is used to indicate that shares or a currency become reduced in value;
> *shed image* means a deliberate changing of how one is perceived. (Hunston 2002:12)

As we can see from this example, simply teaching students a dictionary definition of *shed*, its possible translations, or asking them to match this word with a synonym, would neglect many of its varied and rich meanings and could give students a false idea that there is *one* equivalent of this word in their language. Morgan Lewis (2000:13) uses the synonyms *wound* and *injury* to exemplify the same view. He explains that teaching these words in collocations like *stab wound* and *internal injuries* would be more helpful for students than simply giving them dictionary definitions. As he affirms, 'it is the **collocational fields** of the two words which

reveal the difference of meaning, or rather more precisely, the difference between the ways the words are used' (Morgan Lewis 2000:13, original emphasis). What is meant by *collocational fields* here is the different contexts in which a word is used; for instance, some collocates of *injury* given by COCA (Davies 2008) are *brain*, *head*, *knee*, and *shoulder* and of *wound* are *gunshot*, *bullet*, and *stab*. This indicates that one of the collocational fields of the word *injury* is 'parts of the body' while a collocational field of *wound* is 'types of injury'.

Although collocations are recognized as part of the communicative competence of the learner in the Framework, little or no attention is given to it in Norwegian course material. One reason for this neglect, as Howarth (1998) concludes in his study, might be the teachers' limited knowledge of the phenomenon. He also observes that learners that are unaware of collocations might come to believe that knowing a language depends on their ability to syntactically combine items of vocabulary and memorize idioms. However, he deduces from his results that it is in the middle part of the spectrum, i.e. restricted collocations, where the difficulty lies and concludes that, 'such combinations are not optional stylistic adornments on the surface of text; they are essential for effective communication […]' (Howarth 1998:186).

2.5.5 Culture and communication in word combinations

Culture has had a central place in CLT. When explaining what communicative language teaching is, Berns stresses the cultural aspect by saying that 'culture is recognized as playing an instrumental role in shaping speakers' communicative competence, both in their first and subsequent languages' (Berns 1990:104). Thus, culture shapes the way in which we communicate. When advocating a communicative competence, Hymes (1979) emphasized the sociocultural features that influence the way in which people communicate in spoken and written discourse. Since then, sociocultural aspects have been taken more into account in language teaching; so much so that the Framework includes these under 'general competences' of the user/learner (see table 2.2) and it underlines the importance of 'knowledge of the society and culture of the community or communities in which a language is spoken' (Council of Europe 2001:102). This subsection will deal with the relationship between culture and collocations and its implications for the teaching of English in Norway.

Certain collocations can reveal changes in society and, as a consequence, changes in the language. Hunston comments on this matter affirming that 'newly emerging collocations can be used to indicate the growth of new concepts, and changes in the meaning of words' (2002:118). She uses Stubbs' examples of *single parent families* and *unmarried mothers* to show how new social structures can reflect on the language. Stubbs also studies the change in

meaning of the word *work* from 'doing something' to 'paid social relationship', so that nowadays saying that someone is a *woman who works* or a *working woman* does not indicate that she does *housework*, but that she has a paid job (Stubbs 1996:177).

In the same article, Stubbs analyzes discourse in the sense of 'recurrent phrases and conventional ways of talking which circulate in the social world, and which form a constellation of repeated meanings' (1996:158) and shows how collocations are used as a means of propagating pre-fabricated ideas. Some of the collocations he studies are *British heritage*, *deteriorating standards*, and *back to basics*, which were often used in the 1980s to convey conservative values. One of these values was the idealization of education in the past and skepticism to new teaching trends. To research how those ideas were conveyed in the language, he checked for example the common collocates of *trendy* in mainstream published language of that period and found among others: *lessons*, *methods*, and *teachers*. These were usually found in negative sentences such as 'Prince Charles is furious at *trendy teaching* which has axed Shakespeare from many schools' (Stubbs 1996:162-163, original emphasis). Thus, observing collocations in the English language today can help us identify important aspects of British and American culture and current events. We can bring to learners' attention for instance the expressions commonly used after the 9/11 attacks in the USA: *war on terror*, *anti-terror campaign*, *terrorist networks*, *counter-terrorist operations*, *terrorist attacks*, etc.

The idea of identifying sociocultural aspects in word combinations is also underlined by Teliya et al. (1998) in their study on phraseology and culture. They argue that 'linguo-cultural analysis […] is best suited for phraseology, and especially for restricted collocations. The latter abound in cultural information and can hardly be described at all as a class of denominations if their cultural meanings are not taken into account' (Teliya et al. 1998:56). Although their study was on Russian lexical collocations, their methods could be applied to investigations in any language since it analyzes stereotypes, cultural concepts, connotations, and conceptual metaphors that are behind certain word combinations. They note for example the Russian stereotype that women have low intellectual capacity and list a number of collocations that convey this idea (1998:63). One of the few examples they give of English is the case of the mule, saying that in English it seems that the mule is the stereotype for 'stubbornness', which is shown in expressions like *mulish stubbornness* and *as stubborn as a mule* (Teliya et al. 1998:67). The results found by Teliya et al. are confirmed by some of the findings in Bartsch's study of collocations in the BNC, where she has observed that certain collocations carry 'fixed and stereotyped embedded meanings which are deeply rooted in the

cultural background of a linguistic community (e.g. *age of consent*, *affirmative action*)'
(2004:177).

This is an interesting and apparently underexplored area that would be worth researching. There has been for example some studies on keywords (Said 1978; Williams 1985; Fairclough 1990) and fixed phrases (Moon 1994) in connection with culture, while not much has been done relating collocations, culture, and language teaching. Although it is beyond the scope of the present thesis to look into this matter in detail, I have tried to identify in the material sociocultural aspects that could be brought up in the English classroom. Besides being relevant regarding CLT, culture is also highlighted in L06. The curriculum sets under the subject area of 'culture, society and literature' the following competence aim: 'discuss social conditions and values in various cultures in a number of English-speaking countries' (L06 2006b), which points out the relevance of sociocultural aspects in the teaching of English in Norway.

2.5.6 Collocation dictionaries

According to Nation, from a vocabulary point of view, research into collocation is needed:

- to tell us what the high-frequency collocations are;
- to tell us what the unpredictable collocations of high-frequency words are;
- to tell us what the common patterns of collocations are […];
- to provide dictionaries (or information for dictionaries) that help learners deal with low-frequency collocations. (Nation 2001: 328)

All the above is possible through the use of corpora. However, in case corpus consultation is not possible in the classroom or students have difficulties in accessing the corpus directly, collocation dictionaries are an option. This type of dictionary is usually based on large corpora, including learner corpora, and provides invaluable information to learners. Three well-known collocation dictionaries which are currently available are *The BBI dictionary of English word combinations* (Benson et al. 1997), *The LTP dictionary of selected collocations* (Hill & Lewis 1997) and the most recent *Oxford collocations dictionary for students of English* (McIntosh et al. 2009). Such dictionaries can aid students in expressing their ideas more accurately, both in speaking and writing. One of the competence aims of L06 is the use of monolingual dictionaries (L06 2006b), so introducing collocation dictionaries in Norwegian schools could help enhance students' vocabulary knowledge.

2.6 Corpus linguistics

Hidalgo et al. state that 'corpus linguistics has played a powerful role in language research, grammar construction, dictionary making, natural language processing, cognitive studies, and language learning and teaching, among other fields' (2007:ix). This section will present some research in the area of language learning and teaching, especially in relation to phraseology, teaching materials, authenticity, English for specific purposes (ESP), and learner corpora. Some corpus limitations will also be mentioned.

A ground-breaking effort to bring corpus use to the classroom has been the work of Tim Johns (e.g.1994) on *data-driven learning*. The idea is to guide learners in using corpora to *discover* features and patterns in the foreign language similarly to how corpus linguists do (Bernardini 2004:16). This changes the focus from deductive to inductive learning and affects students and teachers' roles so that teachers become more like facilitators and learners become more responsible for their own learning. Other authors that favor this approach are Seidlhofer (2000a), Kennedy & Miceli (2001), and Chambers & O'Sullivan (2004).

Corpus linguistics has also been widely used in phraseology to analyze lexical patterning in language (Howarth 1998; Cowie 1998a; Stubbs 2001a). In relation to language teaching, Meunier & Gouverneur (2007) say that many authors have pointed up how crucial it is to address multi-word units (e.g. Michael Lewis 1993; Nation 2001), but they observe that studies on the treatment of these units in textbooks are rare (2007:121). In face of this lack of empirical material on the matter, they decided to investigate whether research findings in SLA have affected the design of five British EFL textbooks and found that the editors of these books are in fact aware of the importance of teaching phraseology. However, it seems that more research is needed to detect how much the findings in corpus linguistics have affected Norwegian EFL textbooks.

Another aspect of corpus linguistics related to teaching materials is on the question of authenticity. It is argued that textbooks that present invented examples and have descriptions based on intuition should instead use corpora as resources of 'real life language data', which, according to Collins (2000:52) has been happening gradually. Besides providing real life examples, corpora findings can also be used to overrule many of the traditional grammatical explanations presented in textbooks, as shown by a study carried out by Clear (2000:19-30). Johansson (2007) has also described how corpora can be used in teaching, emphasizing the role of the learner as a researcher and as such being able to have a more critical approach to textbooks.

Another area of corpus linguistics research is on the use of small corpora for teaching ESP (see Gaviolli 2005). By using specialized corpora (e.g. medical articles or business reports), specific expressions and technical terms commonly used in different fields can be pointed out to students. Learner corpora are another type of corpus that has brought new insight into language teaching and have been used to produce pedagogical material (e.g. the *OCD*).

However, using corpora also has limitations, and it is essential to be aware of what a corpus can and cannot do. Hunston lists what she sees as some of its main restrictions:

> (1) A corpus will not give information about whether something is possible or not, only whether it is frequent or not. (2) A corpus can show nothing more than its own contents. […] Thus conclusions about language drawn from a corpus have to be treated as deductions, not as facts. (3) A corpus can offer evidence but cannot give information. […] The corpus simply offers the researcher plenty of examples; only intuition can interpret them. (4) Perhaps most seriously a corpus presents language out of its context. […] These factors all show the need for a corpus to be one tool among many in the study of language. (Hunston 2002:22-23)

Since a corpus is used in the present study as part of the material and methods, these constraints have been taken into consideration. Nonetheless, despite its limitations, a corpus should not be discarded as a source of information about a language, even more importantly when looking for what is common usage. As Sinclair states, 'one of the principle uses of a corpus is to identify what is central and typical in the language' (1991:17).

2.7 A summary: Vocabulary teaching, collocations, and corpora

This chapter has attempted to show how vocabulary teaching, collocations, and corpora are related, emphasizing their importance for Norwegian students at the upper secondary school level. Since CLT, the Framework, and cultural awareness have been central to language teaching in Norway, they have also been given attention in the present chapter. Other topics which give a background to the remainder of this thesis are foreign language vocabulary acquisition and language teaching and learning.

3. MATERIAL AND METHODS

3.1 Introduction

This chapter will first introduce the primary material selected for investigation, i.e. *Passage* and @*cross*, and the criteria for the choice of this material will be presented. Next, a detailed description of the two textbooks and the corpus will follow. Finally, the methods that have been used to achieve the aims will be explained and some limitations of the material and methods will be discussed.

3.2 Material

The core material used in the present study is vocabulary exercises from @*cross workbook* (Rodgers et al. 2006b) and from the *Passage* website (Sørhus et al. 2006b), both of which are used in the first year of upper secondary school in Norway.

The textbooks @*cross* and *Passage* are written for pupils in the *studieforberedende program,* which is translated as 'programmes for general studies' in the English version of the L06 (2006b). As the name indicates, the English taught in this course is more general, different from programs such as 'health and social studies' in which the vocabulary is more specific to their line of work. Although I consider knowledge of collocations extremely relevant for the teaching of ESP, I have chosen to research these in connection with the general studies program since it is the one that I am currently teaching.

The textbooks usually adopted for this course in the upper secondary schools are the two mentioned above in addition to *Targets* (Berntzen & Bårtvedt 2005) and *Experience* (Heian et al. 2006). All the four books meet the requirements of the most recent English curriculum[4] (L06), which provides a list of competence aims to be achieved by the end of certain school years; among them, the first year of upper secondary school (see appendix 1).

3.2.1 Criteria for selection of textbooks and exercises

The first criterion established for choosing the textbooks was that the vocabulary exercises in the books should be presented separately from the grammatical ones. This can be challenging since, as Sinclair & Renouf maintain, 'it is almost impossible to teach grammar without in passing teaching some vocabulary. Vocabulary fleshes out the structures, introduces variety and promotes practice of the structure in question' (1988:143). However, considering that the

---

[4] Although *Targets'* edition is from 2005, it has been used in Norwegian schools until 2009 when a new edition was published. It has probably been adopted for so long for meeting many of the requirements of L06.

present study is concerned with lexis, the exercises used should work essentially with vocabulary. The @*cross* material, for example, is very clear in its categorization, placing the lexical and grammatical exercises in different sections ('building vocabulary' and 'learning grammar' respectively), which is why @*cross* was chosen. Below is an example of a vocabulary exercise from @*cross*.

---

190 **Write a paragraph describing American teenagers and their situation using the following words.**
Sports, prom, weird, wired, bullying, committed, frustrated, happy, shopping, trend, cliques
Add other words from the texts which you think fit in.

Figure 3.1 Example of a vocabulary exercise from @*cross* (Rogers et al. 2006b:111)

Passage, *Experience*, and *Targets* have vocabulary and grammar exercises mixed, which posed a problem for deciding on the second book. The solution was to look at their websites to see if the vocabulary exercises were clearly labeled as such there. Having found that in all three websites they were categorized as working on vocabulary, another criterion was needed as a means to select only one of them. Searching the three websites, it was found that *Targets* and *Experience* have too many vocabulary exercises, and it would not be possible to analyze all of them due to time restrictions. Therefore, the ones decided upon in the end were *Passage* and @*cross*. It is worth noting that only two textbooks have been selected in order to limit the scope of the study.

After choosing the two textbooks, certain criteria have been adopted to select the vocabulary exercises. Since the first aim involves finding out if words that are usually taught as separate items in textbook exercises are part of collocations in the texts they are taken from, the exercises should present lists of single decontextualized words, like the one from @*cross* in figure 3.1. Thus, collocation and gap-filling exercises, as exemplified respectively in figure 3.2 and figure 3.3, were excluded.

Make word pairs from the two columns

rate of	exchange
entrance	season
high	agency
round	trip
travel	fee
detailed	passport
valid	itinerary

Figure 3.2 Collocation exercise from *Passage* (Sørhus et al. 2006b)

Complete the text by filling in the gaps with words from the list on the right.

English started out as a Germanic [_____] spoken only in England. Today it has become a [_____] language with about 350 million [_____] speakers. What is more, it has become a [_____] franca spoken by many millions as a [_____] language. Experts [_____] that these non-native [_____] will have a dramatic [_____] of the language in the future.

| speakers |
| predict |
| impact |
| global |
| lingua |
| native |
| second |
| dialect |

Figure 3.3 Gap-filling exercise from *Passage* (Sørhus et al. 2006b)

It is necessary at this point to highlight that the two types of exercise above are highly recommended by several authors for vocabulary practice (Brown 1974; Stevick 1976; Nattinger 1988). Therefore, it is unfortunate that both textbooks offer remarkably few of these in its vocabulary section and that most of its exercises work with words in isolation. Regarding only collocation exercises, *@cross* includes none while *Passage* has only three. Nation explains that teaching vocabulary using lists of words out of context is probably effective for 'learners with a small vocabulary who wish to go on to academic study in a few months' time' (1990:2); which is not the case of Norwegian students at the level in question. At the first year of upper secondary school, students have been studying English for ten years (they usually start in first grade), so they have probably acquired a sizable vocabulary, and they still have at least two years until entering university.

Vocabulary exercises that work with proper names and names of geographical places or events (shown in figure 3.4 below) were also excluded. The reason is that the objective of these activities seems to be more to remember factual information from the texts than to test general vocabulary knowledge.

---

101 Word Search
 Find the words in the grid.
Arsenal, Covent Garden, Oxford Street, Big Ben, Ealing, Parliament, Camden Market, Globe, Tower of London, Chelsea, Greenwich, Tube, City, Heathrow Airport, Westminster, Downing St, Buckingham Palace, Soho.

Figure 3.4 Exercise with geographical names from *@cross* (Rodgers et al. 2006b)

After excluding the types of activities discussed above, there were exactly fourteen single word vocabulary exercises from each material, all of which were used in the present study. Table 3.1 below shows the types of exercises selected and the number of each type.

Table 3.1 Number and types of vocabulary exercises from each material

Type of exercise	@cross	Passage	Total
Matching synonyms, antonyms, or words and their definitions	4	13	17
Involving Norwegian translation	6	1	7
Using certain words to write a text	3	-	3
Finding words in grids	1	-	1
Total	**14**	**14**	**28**

All of these types of exercises, except the one to write a text, belong to the fourth step in the vocabulary acquisition process mentioned by Hatch & Brown (1995; see section 2.3); namely, consolidating word form and meaning in memory. The type in which students need to write a text with certain words requires that students *use* the words, which is the last step in the acquisition process. More examples of the types of exercise that were selected for investigation will be given in the *@cross* and *Passage* subsections below.

3.2.2 *@cross*

The *@cross* material consists of:

- textbook
- workbook
- teacher's book
- 8 CDs
- website

For the purposes of this thesis mainly the workbook and the textbook have been used. The material was developed by Drew Rodgers, Knut Skifjeld, Dilys Brown, & Martin Mulloy and published by Damm. *@cross* targets Norwegian pupils specifically, often contrasting Norway with the English speaking world. Regarding the variety of English used, it includes both American and British English, depending if the text is about the USA or the UK. The book also brings other varieties like Australian, Jamaican, and South African English.

@cross is divided into four parts: '@cross the UK', '@cross the English speaking Countries', '@cross the USA', and '@cross Literature'. Part 1 is subdivided into eleven chapters that cover different regions of the United Kingdom, including cultural and historical information about them. Part 2 has only three chapters, which talk about Jamaica, South Africa, and New Zealand. Part 3 contains twelve chapters that describe various states and

cities in the United States, and provide historical and cultural facts. With reference to genres, it is stated in the preface of *@cross textbook* that 'you will find a combination of two teaching traditions – storytelling and factual articles' (Rodgers et al. 2006a:5). Thus, the texts in each of these three parts are in the form of storytelling, with young characters experiencing the English speaking world. At the end of each chapter, there is a section called 'fact files' which gives extra factual information about what was presented in that chapter. The authors explain that the idea is to give pupils the opportunity to choose the type of text that suits them best, taking into consideration students' different learning styles (Rodgers et al. 2006a: ibid.). The other kinds of genre the book includes appear in Part 4 of the textbook. This last part is concerned with literary texts ranging from the 1500s until today, as required by the new curriculum. Whereas the texts in the three first parts are written for the purposes of teaching English to Norwegian students (i.e. constructed texts), the texts in the literature section are authentic (i.e. written for English speakers) and taken from a variety of genres, including poems, short stories, and excerpts from novels.

The textbook contains only texts, while the workbook and the website provide all the exercises. The authors state that they 'have included tasks covering communicative, cultural, numeric and digital skills' (Rodgers et al. 2006a:6). These kinds of activities are grouped in each chapter in the workbook under the headings: 'geography', 'fact file', 'the journey', 'learning grammar', 'building vocabulary', 'reading skills', 'writing skills', 'speaking skills', 'numeric skills', 'society and culture', 'making a presentation', 'listening skills', and 'wordbank'. All the vocabulary exercises selected in the present study are in the 'building vocabulary' section, which contains from one to three exercises in each chapter.

Since all the activities are in *@cross workbook*, this is my main source of material. The vocabulary exercises on the website are not clearly categorized and there are very few of them, which is why the website has not been used. In total, fourteen vocabulary exercises have been selected and they are mainly of the types:

1. Correlating English words with Norwegian translations
2. Translating a list of words
3. Linking English words with their synonyms or antonyms
4. Matching words with their definitions
5. Using a range of words to write a paragraph on a determined topic
6. Finding words in a grid (only those that provided a list of English words to be found have been chosen; vocabulary exercises with grids but no lists were excluded.)

This list is more detailed than the one in table 3.1, where the categories were generalized. As stated in the criteria subsection, what all of these exercises have in common is that they present lists of single words to be worked with. A few of them had a collocation or two, in that case the collocation was not searched in the corpus. Below is an example of type 3:

<div style="border:1px solid">

36 Match the word in the left hand column together with the word which means roughly the same, from the right hand column. If you need help, look at pg [*sic*] 41 - 44 in @cross Textbook. There you will find the words used.

1.	coastal	a. national costume
2.	apply to	b. proof
3.	generally	c. along the seaside
4.	stunning	d. normally/commonly
5.	kilt	e. outskirts
6.	soundtrack	f. 100 years
7.	century	g. seek admission
8.	preserved	h. fabulous/fantastic
9.	suburbs	i. taken care of/looked after
10.	evidence	j. music to set the mood

</div>

Figure 3.5 Exercise to match synonyms from @*cross* (Rodgers et al. 2006b:34)

3.2.3 *Passage*

Passage is, as the author expresses in the preface, 'an all-in-one book with both texts and activities' (Sørhus et al. 2006a:7), so the material is formed only by:

- textbook
- CDs
- website

The authors are Theresa Bowles Sørhus, Richard Burgess, & Trond Christian Anvik and the material is published by Cappelen. *Passage* is also a book written for Norwegian pupils and it embodies both American and British varieties of English. The new curriculum has had significant consequences for this latest edition of *Passage*, as stated in its preface (Sørhus et al. 2006a:6). What is distinct from the previous edition is that to complement the material on the USA and the UK, the authors have added literature and information about other English-speaking countries, native peoples of the English speaking-world, and current global issues. Parallel to what we find in @*cross*, the writers have incorporated literary texts from the 1500s up to modern times; the difference is that these are included between chapters and not at the end of the book.

The *Passage* material is divided not by chapters but by general themes and each of these includes several texts. The themes/sections are:

- Starting Out
- Global English
- On the Road
- We Were Here First
- The Road to Freedom
- The Power and the Glory
- Global Village

These sections contain texts written specifically for Norwegian pupils and authentic texts taken from novels, short stories, literature, newspapers, etc. After each text there are activities distributed under the titles: 'understanding the text/story', 'talk about it', 'improve your language', 'role play', 'write about it', and 'research', among others which work on mathematical skills. Some texts are followed by all of these types of activities while others by only a few. All the vocabulary exercises are listed together with the grammar exercises under the heading 'improve your language'. The exercises on the website are listed under the title of each text of the textbook and there can be from none to three vocabulary exercises for each text. These are usually called 'pair up words', 'crossword', 'find words', 'find synonyms', or 'connect synonyms'.

As explained previously, the second half of the material is taken from the *Passage* website (Sørhus et al. 2006b), which offers a substantial amount of extra material, including vocabulary exercises. Among these, fourteen were chosen to be analyzed and they are of the following types:

1. Correlating English words with Norwegian translations
2. Linking English words with their synonyms
3. Matching up words with their definitions

Most of the exercises are to link words with synonyms (see table 3.1), as the one in figure 3.6 below.

Pair up the words on the left (as used in the text) with their synonyms on the right.

evidently ⭕	⭕ tired
commence ⭕	⭕ tablet
capsule ⭕	⭕ clearly
slack ⭕	⭕ begin
various ⭕	⭕ different
detached ⭕	⭕ distant

Figure 3.6 Exercise to match synonyms from *Passage* (Sørhus et al. 2006b)

In order to learn more about the collocational behavior of the words in the exercises selected, another kind of material was needed, which will be presented in the next subsection.

3.2.4 The corpus

The corpus chosen for this research is the Corpus of Contemporary American English (COCA) created by Mark Davies (2008). It is the largest and most recent corpus available online, being launched in 2008 and comprising more than 400 million words distributed among more than 160.000 texts. As described in its introductory page, COCA includes '20 million words each year from 1990-2009, and it is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts' (Davies 2008). What also makes it stand out among all the other corpora is that COCA is updated every six to nine months, and can therefore be used to record the changes in American English over time. The access to the corpus is completely free of charge. After trying a few query words, the user is required to register a username and a password; when logged in, you can track your previous searches and make your own search lists.

The main reasons for the choice of COCA are its size, its free access, and the search tools available, which will be explained in detail in the method section. Since the present study deals with students at upper secondary school level, a corpus that entails the latest trends in the English language seemed the most appropriate alternative. Some might argue that the British National Corpus (BNC) would be a better option since British English has always exerted great influence in Norway, especially because of geographical proximity. However, although being aware of the prestige of British English in Norway, I should point out that Norwegian teenagers are under even greater influence of American English on a daily basis. American bands, films, computer games, and TV-series are massively popular among teenagers in Norway. Moreover, the BNC contains 100 million words, 300 million less than

COCA, and it is mostly composed of texts written in the 1980s and 1990s, which limits the search and fails to represent English use today. Other important differences between these two corpora are exemplified in the section 'compare wordlists to BNC' in the 'more information' part of the COCA webpage, as follows:

> Some differences are related to culture, society, politics, or current events (e.g. Am *Republican, congressional, baseball, Iraqi*; Br *Tory, parliamentary, Victorian*), some are just different words for the "same" concept (e.g. *store/shop, attorney/solicitor, apartment/flat, mom/mum*), and some words in COCA (1990-2009) refer to things that are too new to have made it into the pre-1993 BNC (e.g. *web, Internet, high-tech, online*). (Davies 2008)

Turning to the question of representativeness, every corpus intends to be representative of a determined population. Large corpora like COCA and the BNC are built to represent the English language in general, or at least American and British English respectively. Nonetheless, to be able to make such a claim, these corpora have to follow certain criteria. In line with McEnery et al., 'the representativeness of most corpora is to a great extent determined by two factors: the range of genres included in a corpus (i.e. *balance* […]) and how the text chunks for each genre are selected (i.e. *sampling* […])' (2006:13, original emphasis). Hunston, however, adds a third factor – its change over time. She claims that 'any corpus that is not regularly updated rapidly becomes unrepresentative' (Hunston 2002:30). Davies has tried to meet these three criteria when designing COCA in that it covers a wide variety of genres, including spoken data, so it is *balanced*; the 160.000 texts are evenly divided among the five genres[5] (and again in various subgenres) and among the four time-periods, so it can be considered a *sample* of general English use; and it is constantly updated, showing *change over time*.

  Some might still question how comparable the texts in the textbooks and the texts in COCA are. Consonant with COCA, the textbooks aim to generally represent the English language, in particular these that are designed for the general studies program. Although neither the textbooks nor the corpus are able to show the whole of the English language, they can still give an idea of how English is being used through their selection of genres and registers. Even though the textbooks include both American and British English and the corpus comprises mainly American English, choosing a British corpus would also be restrictive and using two corpora would be outside the scope of this study.

---

[5] *Genre* is used here in the same way as it is in COCA, although it might not be the most appropriate term for this context.

3.3 Methods

As described in the introduction, the first aim is to find out if single words presented in vocabulary exercises are part of collocations in the texts they are taken from. The first research question is how large this number is and the second is which of these are useful to teach students. The hypothesis is related to the first research question and it predicts that a high proportion of collocations will be found. As a way to achieve this aim and address the research questions, first fourteen vocabulary exercises from @*cross workbook* and fourteen exercises from the *Passage* website were selected; then the words from each of these exercises were looked for in the texts they were taken from to see if they were part of any word combinations. Thereafter, these were checked in the *OCD* (McIntosh et al. 2009) and in the corpus and only the ones that were listed in both and which presented frequencies higher than 3 and MI scores higher than 3.0 in COCA were classified as collocations. Next, all the words which appeared in collocations in the texts were counted and a percentage calculated. Finally, the competence aims of L06 and research on collocation and vocabulary acquisition were used to discuss which types of collocations would be most useful to teach at this level.

The second aim is to suggest how the vocabulary exercises selected for the present study can be complemented based on existing teaching material on collocations and by applying previous research on corpora and foreign language vocabulary acquisition. Knowing that textbooks like *Touchstone* have been designed with the help of a corpus (see McCarthy et al. 2006a:iv), I have chosen to use information provided by COCA to complement the vocabulary exercises. Thus, the method was to look up the words from the exercises in the corpus to find their most frequent collocates with an MI above 3.0. Because of space and time constraints, only the ten most frequent collocates are listed and only the results for one word from each exercise are shown[6].The dictionary was only used to help select collocations appearing in the texts (related to the first aim) but not to determine if the ten combinations found in the corpus are indeed collocations. Thus, following the definition in chapter 1 accurately, these are simply referred to as 'combinations' or 'word combinations'. The reason for not using the dictionary in relation to the second aim is that it would be extremely time consuming and beyond the scope of this study.

The method used to achieve the third aim is purely qualitative. Given that complementing vocabulary exercises is not the only way collocations can be incorporated into

---

[6] The ten most frequent collocates of *all* the words from each exercise were searched; however, it was not possible to include them in the thesis also due to space limitation.

teaching, previous research and existing teaching material on collocations have been consulted to suggest other ways in which this can be carried out. All in all, the results achieved by using these procedures will hopefully give a rough idea of the range of collocations that appear in textbooks, show the potential of corpora for teaching, and emphasize the importance of including collocations not only in vocabulary practice but in English teaching in general.

3.3.1 The dictionary

One of the methods used in the present study was looking up words in a collocation dictionary. As mentioned above, the *Oxford collocations dictionary for students of English* (*OCD*; McIntosh et al. 2009) has been used to check whether the word combinations found in the texts are indeed collocations. This dictionary has been chosen due to the fact that it is the most recent and comprehensive one, containing around 9.000 headwords, 250.000 word combinations, and 75.000 examples. This dictionary was first edited in 2001 and this revised edition has added 100.000 collocations and 25.000 examples besides including a CD-ROM. The *OCD* comprises both British and American English and these variations are labeled in the dictionary when necessary.

The *OCD* covers thirteen types of syntactical combinations organized under three general entries: noun, verb, and adjective. Under the noun entries we find: adjective + noun (*bright light*), quantifier + noun (*a beam of light*), verb + noun (*cast light*), noun + verb (*light gleams*), noun + noun (*a light source*), preposition + noun (*by the light of the moon*), and noun + preposition (*the light from the window*). Under the verb entries: adverb + verb (*choose carefully*), verb + verb (*be free to choose*), and verb + preposition (*choose between two things*). Under the adjective entries: verb + adjective (*declare sth safe*), adverb + adjective (*perfectly safe*), and adjective + preposition (*safe from attack*) (McIntosh et al. 2009:vi-vii; original examples). Although the dictionary only provides the type adverb + verb in the list given in the introduction, when checking the entries it becomes clear that verb + adverb combinations are also included. Thus, the categories are fourteen in total and they correspond to the eleven that were listed in chapter 1 in addition to quantifier + noun, verb + verb, and verb + adjective.

The approach adopted by this dictionary pertains to learners' needs rather than to a particular theory and the editors have taken the following questions into consideration when deciding which collocations to incorporate (McIntosh et al. 2009:v):

- Is this a typical use of language?
- Might a student of English want to express this idea?
- Would they look up this entry to find out how?

To embrace what is typical, the editors have used the Oxford English Corpus, which comprises almost 2 billion words, to check for frequency and examples. The second question led the editors to aim for the kind of language students might need to write, such as essays and reports; hence, it takes account of non-specialist (science, history, sport, etc.) and specialist (law, medicine, politics, current affairs, etc.) areas in addition to informal language of the internet and spoken communication (McIntosh et al. 2009:vi). The third question led to the focus on noun entries since students are more likely to look, for example, for *rain* to find *heavy* than the other way around.

One downside of using this dictionary is that infrequent words that appear in the exercises might not be among the 9.000 headwords. Some might argue that this does not necessarily mean that they are not collocations, which is probably true. However, when not found in the dictionary, the combinations were excluded from this investigation to meet the definition established in chapter 1.

3.3.2 MI scores and frequencies in COCA

Since the aims of this study involve finding collocates of certain words and also checking frequency and strength of collocations, a corpus turned out to be the best method to access this kind of information. In addition to the advantages discussed in the material section, other positive aspects of COCA, which will be particularly useful for this study, are its user-friendly interface and its varied number of tools. As noted in the introductory text on the website, 'the corpus allows you to search for exact words or phrases, wildcards, lemmas, part of speech, or any combinations of these' (Davies 2008). Some of these tools are illustrated in figure 3.7 in subsection 3.3.3.

One of the important features of COCA is the possibility to measure Mutual Information (MI) scores. The MI shows how closely the words in a word combination are connected, i.e. it calculates their strength. Drawing from Fano (1961), Church & Hanks give a more detailed definition of MI:

> If two points (words), *x* and *y*, have probabilities *P(x)* and *P(y),* then their mutual information, *I(x, y),* is defined to be
> $$I(x,y) = \log 2 \ \frac{P(x,y)}{P(x)P(y)}$$

> Informally, mutual information compares the probability of observing *x* and *y together* (the joint probability) with the probabilities of observing *x* and *y independently* (chance). If there is a genuine association between *x* and *y*, then the joint probability *P(x, y)* will be much larger than chance *P(x) P(y),* and consequently *I(x, y)* $\gg$ 0. (Church & Hanks 1990:23; original emphasis)

Thus, according to this explanation, the higher the MI the stronger the combination is. Later in the same article, Church & Hanks note that 'as a very rough rule of thumb, we have observed that pairs with *I(x, y) > 3* tend to be interesting, and pairs with smaller *I(x, y)* are generally not' (1990:24). This means that pairs with scores above 3.0 can probably be considered collocations and below that, free combinations. Similarly, it is pointed out in the help section of COCA that 'typically, scores of about 3.0 or above shows [*sic*] a "semantic bonding" between the two words' (Davies 2008). On these grounds, I have used the 3.0 threshold to exclude free combinations. Illustrating how the MI works in practice, if we look for the collocates of the word *fancy*, we will find that *fancy restaurants* has an MI of 7.08, while *fancy house* has an MI of 2.21, which tells us that the word *restaurants* has stronger bonds to the word *fancy* than the word *house.*

However, according to Hunston (2002:72), the strength of a collocation can not always guarantee its relevance. McEnery et al. affirm that 'collocations with high MI scores tend to include low-frequency words' (2006:57), and infrequent collocations do not appear particularly interesting for language teaching. This is why the tables with the ten most common combinations are listed according to their frequency and not MI. When the collocates of *fancy*, for example, are listed in COCA in order of frequency, as displayed by table 3.2 below, the top combinations are *fancy cars*, *fancy restaurant*, and *fancy dress*; on the other hand, when shown in order of strength, the top ones are *fancy air-gulper*, *fancy antiseptics-soap*, and *fancy car-of-the-week.* This clearly elucidates why MI should not be used alone to determine what can be considered a collocation or what is relevant for teaching.

Table 3.2 Collocates of the adjective *fancy* in COCA

Collocates	Raw frequency	MI
cars	94	7.51
restaurant	79	7.73
dress	65	7.55
footwork	63	13.74
clothes	62	7.22
name	57	5.12
way	55	3.13
restaurants	48	8.08
car	43	4.82
hotel	40	6.31

Hence, preferably strong but also frequent collocations should be taught or students would be learning language items they would probably never use, which is why the raw frequency (i.e. the actual count of elements) of the combinations in the corpus is also checked. The use of raw frequencies might, however, be problematic and raise questions like: What should be the bottom line for a combination to be classified as a collocation? How can one determine if a certain number of occurrences is frequent or not? Clear affirms that 'the terminology is not well established yet, and it is by no means clear whether the observation of a single co-occurrence in a corpus should be ignored or whether the single instance should be taken to be one of many more which might have occurred' (1993:277). He decides then to adopt the threshold value of three, discarding pairs appearing fewer than three times in the corpus. It has been done likewise in the present study so that when a combination found in the textbook presents a raw frequency lower than three it is not considered a collocation. The issue of what number can be considered frequent, on the other hand, is only solved by using statistical tests like t and z scores. Clear does not rely only on frequencies, applying also MI and t scores in his search. In the present study, however, only the measures established in the definition of collocations are used and the collocations have been generally commented on in terms of usefulness for students and not in terms of frequency. The reasons why other statistical tests have not been employed are discussed in section 3.4 at the end of this chapter.

3.3.3 Description of corpus tools

This subsection will describe how the search in COCA has been carried out. At the same time, some of the corpus tools and technical terms will be explained.

The *node* is the term used to designate the word under investigation (Sinclair 1991:175) and *span* is the distance between the node and its collocates. The span of a word may vary considerably, ranging from an adjacent position to a gap that might even go over sentence boundaries (Carter & McCarthy 1988:34). The terminology used to define this exact distance between the main items of a collocation or combination is *gram*. Thus, when the node and its collocate occur as an adjacent pair they are called *bigram*; when there is a word in between, they are called *trigram* or *3-gram*; when there are two words in between they are called *4-gram*; and so forth. For instance, an adjective plus noun bigram can be illustrated by the collocation *heavy rain*; a verb plus noun 3-gram would be a collocation like *sips his coffee*, while *sipping his decaf coffee* would be a 4-gram collocation.

As to give a better idea of how the search tools of COCA work, figure 3.7 shows what the search interface looks like followed by an explanation of each element:
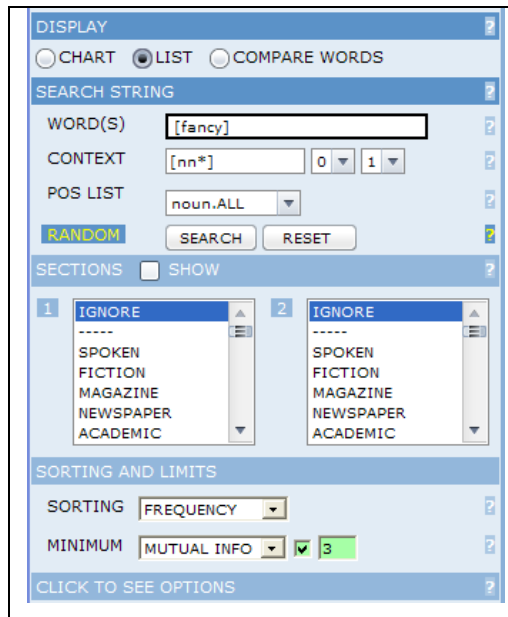
Figure 3.7 Search interface of COCA (Davies 2008)

Describing figure 3.7 from the top, the results of a query can be displayed in a chart or in a list. You choose 'chart' if you want to compare the use of a word in different genres (spoken, fiction, magazine, newspaper, and academic) and year spans (1990-1994, 1995-1999, 2000-2004, and 2005-2009). If you want the corpus to show only the number of occurrences (raw frequency) of a query word, you should choose 'list'. You can also choose 'compare words' to contrast the uses of any two words in the corpus. Moving to the 'search string' area, there is the 'word(s)' field where the node word should be written. It is also possible to look for common collocates with that word and choose from which part of speech you want these collocates to be; to do so, you click on 'POS list' (part of speech list). As figure 3.7 shows, the noun collocates of *fancy* were looked for, so by choosing 'noun.ALL' in the POS list, the label (or *tag*) [nn*] appears automatically in the 'context' field[7]. When you press 'search' the corpus lists the most frequent nouns used with the node or query word, in this case, *fancy*. (It should be noted that for verbs I have used 'verb.LEX', i.e. lexical verbs, to avoid auxiliary verbs like *be*, *have*, and *do*). The boxes showing '0' and '1' enable you to choose the distance between the query word and its collocates, i.e. the grams, and in which side you want the collocates to appear (you can have up to nine words on either side). This tool has been employed to find the ten most common combinations with the words from the vocabulary exercises. The result of the search illustrated in figure 3.7 is shown in table 3.2 above.

Under the 'POS list' there is a label called 'random'. This tool is meant to help those

---

[7] The tagset used by COCA is the same as the one used by the BNC (see *The BNC tagset* in the reference list).

who are unfamiliar with the corpus, giving an idea of how an entry can be written and how the results look like. Below 'random' you can choose to see the results in sections, which shows the same information as the chart, but without the graphs. You can choose to see one or two words in different genres and subgenres (e.g. you can choose education, history, humanities, etc. within the academic genre), year span, or a specific year. In the 'sorting and limits' area you can choose to sort the results by relevance (MI), frequency, or alphabetically. You can also determine the minimum of words the corpus should display and the minimum MI score. This tool has been used when looking for the ten most frequent combinations of a word to select only the ones with MI scores above 3.0.

When needing to check only the frequency of a collocation found in the text, I wrote the node word from the exercise in the 'word(s)' field and its collocate in the 'context' field, so in this case the POS function was not used (to add the MI, I chose the option 'relevance' in the 'sorting' field). The corpus offers other search tools which are designed for more detailed searches. Giving that they are not relevant to this study, only the most general ones have been described.

3.3.4 Criteria for corpus search

As a means to look for the ten most frequent combinations with the words from the exercises in the corpus, a few criteria had to be established. Some of the questions that needed to be addressed were: (1) If the exercise includes for example the verb *look*, should the corpus give results for *looking*, *looks*, and *looked* or only the infinitive *look*? (2) How many words to the left and to the right of the node *look* should the corpus find? (3) To what part of speech should the collocates of *look* belong to?

Addressing question (1), the definition adopted in this study states that a collocation is a combination of one or more *lexemes* so that various forms of the same word can be included (to find lexemes in the corpus the words have to be written in square brackets). For example, to find the most frequent collocates of *boundary*, the lexeme [*boundary*] is written in the 'word(s)' field and the POS for lexemes of lexical verbs (verb.LEX) is chosen. Then [vv*] appears automatically in the context field. This search gives results like: *cross the boundary*, *crossed the boundaries*, and *crossing the boundary*, which include form variations of the node and collocates.

When looking for combinations in the corpus, something that should also be decided upon was how many words should be allowed between the collocates, i.e. if the combination should be a bigram, a 3-gram, a 4-gram, etc., as explained above. This choice depended

directly on which part of speech the node belonged to. Thus, answering questions (2) and (3), the following criteria were established concerning grams and parts of speech:

Table 3.3 Criteria for finding combinations in the corpus

Node	Category	Abbreviation	Gram	Example
noun e.g. *issue*	adjective plus noun verb plus noun noun plus noun noun plus preposition preposition plus noun noun plus verb	Adj+N V+N N+N N+P P+N N+V	bigram 4-gram bigram bigram bigram bigram	*social issues* *discuss an issue* *security issues* *issue concerning* *at issue* *the issue arises*
adjective e.g. *weary*	adjective plus noun adverb plus adjective adjective plus preposition	Adj+N Adv+Adj Adj+P	bigram bigram bigram	*weary eyes* *increasingly weary* *weary of*
transitive verb e.g. *glance*	verb plus noun adverb plus verb verb plus preposition	V+N Adv+V V+P	4-gram bigram bigram	*glanced at the clock* *quickly glanced* *glanced at*
intransitive verb e.g. *vanish*	verb plus adverb adverb plus verb	V+Adv Adv+V	bigram bigram	*vanished completely* *simply vanished*
adverb e.g. *generally*	adverb plus adjective adverb plus verb	Adv+Adj Adv+V	bigram bigram	*generally accepted* *generally speaking*

The decisions relating to grams were based on how many words usually come between the two main items, if any. Hence, looking at the table above, we can see that in most cases bigrams were looked for since most of those lexemes are found adjacent to each other (e.g. there are no words appearing between *weary* and *of*). In the case of verbs, a distinction had to be made between transitive and intransitive verbs since the former takes a direct object and as such, it is common to be found in a V+N combination (e.g. **glanced** *at the* **clock**); the latter, on the other hand, is most commonly found in V+Adv combinations (e.g. *vanished completely*). Explaining the gram choice, the reason why the N+V combinations are 4-gram and not bigram like the others is that the direct object might include determiners, prepositions, adjectives, or other types of words (e.g. *discuss **an important** issue*). Summarizing, the following criteria have governed the search:

Table 3.4 Types of combinations looked for in the corpus

Node	Types of combinations		Grams
nouns adjectives verbs adverbs	Adj+N Adj+P P+N N+P N+V	N+N V+P V+Adv Adv+V Adv+Adj	bigram
	V+N		4-gram

As mentioned in chapter 1 and summarized in table 3.4, eleven categories were set up for the search in the corpus. The criteria above were established to organize and facilitate the search

in the corpus and it applies mainly to the search for the ten most common collocates with the exercise words. When encountering a collocation in the textbooks that combined some of the elements above, fit the definition established, and seemed useful for teaching, such as *get good grades* (V+Adj+N), that collocation was also included.

3.4 Limitations of the material and methods

Although the material and methods used in this study were chosen carefully and under established criteria to achieve certain aims, they do present some limitations. Firstly, the use of the textbooks *Passage* and *@cross* is not as widespread as that of *Targets* and *Experience*. Granted that I could have looked for this information in other ways, I have decided to use the Norwegian search engine *Kvasir* on the internet (www.kvasir.no; accessed on November 12, 2009) to find out what English textbooks are most used in the schools. The websites of the first sixty upper secondary schools that were listed when entering the words *boklista VG1* were checked, and their book lists examined to see what English textbook was adopted for the first grade in the general studies program. All of the sixty schools adopted one of the four books mentioned before, so I simply counted them and calculated the percentages. The results are presented in table 3.5 below:

Table 3.5 Textbooks used in a selection of 60 Norwegian schools

English textbooks	Percentage
Targets	40%
Experience	30%
Passage	21.6%
@cross	8.3%

As table 3.5 shows, *Passage* and *@cross* are the ones least used among the first sixty schools found by *Kvasir,* while *Targets* and *Experience* together comprise 70% of the schools. Based on these (approximate) numbers, some might claim that this study is not so across-the-board, and hence not so relevant in the Norwegian upper secondary school scenario. However, this thesis does not aim to encompass all books used in Norwegian schools, but to analyze some of them. Thus, the number of schools adopting these books is irrelevant as long as the books are still in use.

With regard to the corpus, there is an advantage that can turn out to be a limitation – namely that COCA is being updated, as remarked above, every six to nine months. Consequently, it is very likely that if someone tries to search for the same information a year after it was retrieved frequencies and MI scores will not be exactly the same. Although the

data collection started in 2008, all the results have been updated after the last revision of the corpus in the summer of 2009. However, there have probably been some adjustments a while after since some of the MI results are now slightly different in the corpus, despite the frequencies being the same. Another constraint pertaining to the corpus is the fact that whereas it is taken to represent the English language, it cannot present a completely faithful account of it, given that the corpus does not and cannot contain all the words that exist in the language. As McEnery et al. explain, 'a corpus is essentially a *sample* of a language or language variety (i.e. *population*)' (2006:13; original emphasis). Thus, some words or collocations that are used by English speakers might not be found in the corpus, which can happen also due to the fact that COCA comprehends mostly American English.

As noted earlier in this chapter, using raw frequencies without other statistical measures might be arguable since they are independent of the size of the corpus. McEnery et al. (2006:56) suggests two other statistical tests besides MI to identify significant collocations: the *t* and *z* scores. The *t* test takes corpus size into account but in contrast with MI, it tends to play up high-frequency pairs. A combination of MI and *t*-scores would then be the ideal, so that the intersection of the two would give good evidence for what is or is not a relevant collocation. Be that as it may, using statistical tests would be outside the scope of this thesis considering the great amount of calculations needed. Furthermore, applying statistical tests on very large frequencies, such as the ones from COCA, is a non-trivial task with many pitfalls. Bartsch (2004:103) has observed in her study of collocations in the BNC that the *t*-test highlights collocations with function words (such as phrasal verbs, which are not interesting for the present study) whereas the MI scores highlight content word collocations, also called lexical collocations. As Gabrovšek (1998:132-133) notes, this type of collocations is more unpredictable to students and the kind that is probably the most interesting for language teaching. In addition, McEnery et al. say that 'it is possible to use raw frequency (i.e. the actual count) where no comparison between corpora is necessary […]' (2006:52). Since no corpora comparison is required in the present study, the raw frequency results combined with the MI can be used as an aid to determine if a combination is indeed a collocation.

4. RESULTS

4.1 Introduction

This chapter will present the results from the investigation outlined in chapter 3 and a brief explanation will follow each exercise and table as they are introduced. These explanations have been included for matters of layout and convenience since the results constitute a too large amount of data to leave to comment on at the end. The interpretation of the results will be presented in chapter 5.

4.2 Presentation of the data

In the next two sections, the exercises taken from @*cross workbook* and the *Passage* website are presented exactly as they appear in the book and website. Each of them is followed by a table with the exercises' words displayed in alphabetical order together with their respective collocation(s) found in the texts (this type of table will be referred to as 'collocation table'). The collocations are written exactly as found in the texts (e.g. *marks the boundary* in exercise [1]) and the lexeme variations given by the corpus (e.g. *marking the boundary*, *marked the boundary*, etc.) are listed underneath. The total frequency of the collocation is given in parenthesis (see table 4.1 for an example). In case there were no collocations with a certain word, this word was left out of the table.

As mentioned in chapter 3, all the words from each exercise were also entered in the corpus to find their ten most common collocates. It should be kept in mind that these tables list the most frequent collocates *which present an MI above 3.0*. Nonetheless, to avoid being too repetitive, this information is not included in the table heading. Due to space constraints, the results for only one of the words from each exercise will be shown. This type of table (which will be called 'collocate table') is placed after the collocation table, so that each exercise will be followed by two tables. The choice of the words displayed is not arbitrary. They have been picked depending on if they are part of interesting combinations with respect to how corpus information can complement the exercises and to how the sociocultural aspects that they convey can be explored in class.

4.3 Results from @*cross*

In this section the fourteen exercises selected from @*cross workbook* are presented. The exercises are alluded to by their given number (1-14) and not their number in the textbook.

@*cross* has a total of fifty-six vocabulary exercises (listed under the 'building vocabulary' section of the workbook). Here are the types and number of most of them:

- translation (16)
- gap-filling (9)
- matching words to their synonyms, antonyms, or definitions (5)
- unscrambling words (3)
- working with other languages (Patois, French, etc.) (3)
- writing a text using certain words (3)

The other various categories (crosswords, odd-one-out, find words on a grid, play with words, etc.) contain only one or two exercises each, comprising a total of seventeen exercises. None of the fifty-six exercises work directly with collocations and most of them present essentially isolated words. As shown in table 3.1 and explained in chapter 3, the fourteen exercises selected fall within three of the categories above (translation; synonyms, antonyms, or definitions; write a text with certain words) in addition to one more in which students have to find words in a grid. All the texts on which these exercises are based are constructed and not authentic.

Below is the first exercise selected from @*cross*.

[1]   **22 - Berwick** *pg[8] 36-37*
Having read the text on Berwick, find the English words corresponding to the following Norwegian words by drawing lines:
1. nylig   a. attractive
2. grenselinje   b. trades
3. omgi   c. recently
4. forbløffende   d. excellent
5. arbeidsløshet   e. medieval
6. tiltrekkende   f. community
7. i følge   g. surround
8. middelaldersk   h. unemployment
9. utsøkt   i. amazing
10. næringsvei   j. boundary
11. samfunn   k. according to
(Rodgers et al. 2006b:28)

This is an example of an exercise matching vocabulary items in Norwegian with English translations. The text from which these words were taken describes a town called Berwick and it is in Part 1 of the textbook ('Across the UK'), in chapter 2, 'Arrival in England'. Table 4.1 below shows the English words from exercise [1] and, as stated above, those without any collocations in the text were left out of the table. The words listed as *not in the text* do not

---

[8] @*cross* seems to use *pg* as the abbreviation for 'pages' when the correct use should be *pp* (see *CALD* 2005:987).

appear in the Berwick text, although the heading affirms that they do. As seen from the table, the V+N (4-gram) collocations have been written like in *mark_boundary* to indicate that other words appear between the node and the collocate.

Table 4.1 Collocations from exercise [1] in @*cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
boundary	marks the boundary - mark_boundary - marks_boundary - marked_boundary - marking_boundary	V+N	(130) 42 36 28 24	 3.35 5.17 4.56 6.69
community	*not in the text*			
excellent	*not in the text*			
medieval	medieval art - medieval art - medieval arts	Adj+N	(54) 49 5	 5.52 3.61
recently	*not in the text*			
surround	*not in the text*			

As table 4.1 shows, some collocations are stronger and more frequent depending on the form of the verb or if the noun is plural or singular, which is the case of *medieval art* (MI 5.52, frequency 49) and *medieval arts* (MI 3.61, frequency 5), and also the case of the forms of *mark the boundary*. When any of the various forms of a collocation presents frequencies lower than 3 and/or MI scores lower than 3.0, they are still displayed in the table as long as one of its forms fits the category of collocation (e.g. if *medieval arts* appeared only twice or if its MI was 1.5, it would still be included because *medieval art* presents results above the threshold). Table 4.2 displays *boundary*'s ten most frequent 4-gram verb collocates (V+N). According to the table, *cross* and *push* are the most frequent verb collocates of *boundary*.

Table 4.2 Collocates of the noun *boundary* in COCA

Collocates	Raw frequency	MI
cross	157	6.23
set	113	3.16
crossed	74	6.22
crossing	66	6.61
pushing	65	5.68
push	65	5.68
transcend	60	9.02
crosses	57	7.24
define	53	5.84
draw	46	4.68

Exercise [2] below is also a matching exercise but of the type students have to correlate words with their definitions/synonyms. The text where the words (on the left) are taken from is in Part 1, chapter 3, 'To the North – Scotland'.

[2] **36 - Match the word in the left hand column together with the word which means roughly the same, from the right hand column. If you need help, look at pg 41-44 in @cross Textbook. There you will find the words used.**

1. coastal	a. national costume
2. apply to	b. proof
3. generally	c. along the seaside
4. stunning	d. normally/commonly
5. kilt	e. outskirts
6. soundtrack	f. 100 years
7. century	g. seek admission
8. preserved	h. fabulous/fantastic
9. suburbs	i. taken care of/looked after
10. evidence	j. music to set the mood

(Rodgers et al. 2006b:34)

Table 4.3 is shown below with words from exercise [2]. Again there are words that do not appear in the text indicated by the heading (pp 41-44).

Table 4.3 Collocations from exercise [2] in @*cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
coastal	coastal road - coastal road - coastal roads	Adj+N	(57) 51 6	 5.43 4.88
generally	*not in the text*			
preserved	*not in the text*			
suburbs	*not in the text*			

Table 4.4 lists the most frequent bigram adjective collocates with the selected noun *suburbs* (Adj+N). Some collocates of this word present sociocultural aspects that could be brought up in the classroom. The occurrence of *white*, *affluent* (the strongest one), *wealthy*, and *middle-class* could for example be discussed in relation to the meaning of *suburbs*. Maybe also the fact that *eastern suburbs* is not among the ten most common collocates while *northern, southern*, and *western* are the three most frequent ones.

Table 4.4 Collocates of the noun *suburbs* in COCA

Collocates	Raw frequency	MI
northern	134	7.98
southern	118	7.35
western	102	6.85
white	90	4.58
affluent	74	10.40
wealthy	50	8.23
middle-class	44	8.72
American	43	3.25
outer	36	7.84
industrial	32	6.14

Exercise [3] below is in Part 1 of @*cross* and the text where the words are taken from is in chapter 5, 'Over the Sea Again – to North Wales'. It is another exercise in which students have to match words with their synonyms and explanations.

[3]   **59 - Dinner and the Welsh Language** *pg 77-79*
        In the left hand column below you will find 10 words from the text. Match them with the words/expressions in the right hand column.
        1. supper   a. become different
        2. bright   b. not paying attention to
        3. neither   c. small village
        4. retire   d. evening meal
        5. changed   e. talk about something else
        6. afford   f. clever/intelligent
        7. mend   g. "call it a day"
        8. ignoring   h. not any
        9. change the subject   i. have the money to
        10. hamlet   j. repair/keep up
        (Rodgers et al. 2006b:45)

None of the words in exercise [3] appeared with collocates in the text; therefore, no collocation table is shown for this exercise. Some of the words were found in combinations like *mend walls*, *cannot afford*, *things changed*, and *fine supper*. Nonetheless, these pairs are either not listed in the collocation dictionary or have MI scores below 3.0 and thus do not constitute collocations.

Table 4.5 presents the ten most frequent bigram noun collocates of the adjective *bright* (Adj+N). The table shows that different combinations depict different meanings of *bright*, as the case of *bright colors*, *bright future*, and *bright side*.

Table 4.5 Collocates of the adjective *bright* in COCA

Collocates	Raw frequency	MI
light	545	6.42
colors	520	8.83
red	518	6.69
lights	514	8.43
side	370	5.54
future	308	5.80
spot	297	7.49
stars	268	6.96
star	256	6.54
green	235	5.91

Exercise [4] below is also taken from Part 1 of @*cross* (the UK) and its words were taken from the text in chapter 6, 'The North-West'. This exercise is of the type that presents a grid where students have to find words. Since it was not necessary to include it here, the grid was omitted.

[4]   **74 - Find these words on the grid below.**

- lesson
- agenda
- amazed
- commodity
- incredibly
- keen
- pitch
- probably

- remains
- similarities
- suburbs
- trendy
- disease
- instantly
- obviously
- unemplyment [*sic*]

(Rodgers et al. 2006b:51)

Table 4.6 includes the words from exercise [4] with their respective collocations. As the table shows, there are again words that do not appear in the text from the textbook. The collocation *incredibly busy* is followed by *(2x)* in the table, which means that it appears twice in the text.

Table 4.6 Collocations from exercise [4] in @*cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
agenda	*not in the text*			
amazed*	amazed at	Adj+P	749	5.45
commodity	*not in the text*			
incredibly	incredibly busy (2x)	Adv+Adj	27	6.62
instantly	*not in the text*			
suburbs	surrounding suburbs	Adj+N	27	6.64
trendy	trendy restaurants	Adj+N	(77)	
	- trendy restaurants		46	9.89
	- trendy restaurant		31	8.26
unemployment	high unemployment	Adv+N	338	6.45

In table 4.6 *amazed* is marked with an asterisk because in this case the exact form *amazed* was looked for in the corpus and not its lemma AMAZE. The reason is that *amazed* (adj) and *amaze* (v) do not belong to the same part of speech and I have tried to adhere to the meaning intended in the exercise as closely as possible.

Table 4.7 shows the most frequent bigram adjective collocates with the noun *disease* in the corpus (Adj+N).

Table 4.7 Collocates of the noun *disease* in COCA

Collocates	Raw frequency	MI
infectious	1330	11.91
cardiovascular	701	11.19
transmitted	660	10.65
chronic	578	9.14
deadly	221	7.85
venereal	213	12.76
genetic	196	6.99
human	195	3.82
fatal	174	8.12
communicable	169	12.35

The collocates that appear in table 4.7 might be used in class to teach students the more specific, or technical, senses of *disease* like *venereal*, *chronic*, and *genetic*.

Exercise [5] is from Part 1 and the words are taken from chapter 8, 'Breakfast with the Royal Family'. In this exercise the *answer* column is not in the original activity; the answers are added here to show the words more clearly. Note that the word *Westminster* from the exercise is a proper name and therefore it was not included in the investigation.

[5] **100 - Here are a few words describing media and government. They are divided into two columns. Figure out which words and expressions these are and translate them into Norwegian in the 3rd column.**

		(Answers)	Norwegian
parlia	atives	*parliament*	
broad	lysis	*broadsheet*	
govern	archy	*government*	
cover	tion	*coverage*	
West	ment	*Westminster*	
ana	sheet	*anarchy*	
mon	ify	*monarchy*	
elec	ment	*election*	
rat	minster	*ratify*	
represent	age	*representatives*	

(Rodgers et al. 2006b:65)

Table 4.8 contains the words from exercise [5] and their collocations. Here there is a case of two collocations, *houses of parliament* and *members of parliament*, which fall in a different category, namely, N+P+N and appear in the entry of 'phrases' in the dictionary. The reason why they are included is that, as explained in chapter 1, these constitute highly relevant collocations for the theme of the chapter, which is about government in the UK. Due to the presence of the preposition, they were entered in the corpus as 3-grams.

As also shown in table 4.8, *general election(s)* and *elect_representative(s)* are written with an optional plural form. The reason is that when looking for the nodes *election* and *representative*, the corpus only shows variations, if any, of the collocates in the result list. However, when checking the actual sentences we find examples of *election* and *elections*, *representative* and *representatives*. The optional plural is shown every time this is the case. The collocation *election was held* was entered in COCA as *hold_election*, as listed in the dictionary. The combination *international coverage* was found in the text and also in the dictionary, but since its MI was only 2.44 it was omitted from table 4.8.

Table 4.8 Collocations from exercise [5] in @*cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
anarchy	*not in the text*			
election	general election(s)	Adj+N	1560	6.84
	election was held	V+N	(502)	
	- hold_election		277	3.54
	- holding_election		106	2.93
	- holds_election		28	1.82
	- held_election		91	1.71
monarchy	constitutional monarchy	Adj+N	73	10.44
parliament	Houses of Parliament	N+P+N (3-grams)	(185)	
	- houses of parliament		67	6.04
	- house of parliament		118	3.67
	Members of Parliament		(603)	
	- member of parliament		329	7.53
	- members of parliament		274	6.15
representatives	elect representatives	V+N	(129)	
	- elect_representative(s)		44	6.45
	- electing_representative(s)		10	6.40
	- elects_representative(s)		2	5.92
	- elected_representative(s)		73	4.53

Table 4.9 lists the ten most frequent bigram adjective collocates of the noun *coverage* in the corpus (Adj+N). Looking at the results in the table, some of the collocates of this word involve the topic of health insurance, as the case of *universal coverage*, *full coverage*, and *medical coverage*. These collocates could be used to talk about the health system in the US.

Table 4.9 Collocates of the noun *coverage* in COCA

Collocates	Raw frequency	MI
universal	528	9.75
live	373	6.25
special	177	5.59
full	171	5.48
medical	151	5.71
continuing	134	7.45
complete	116	6.04
extensive	91	7.34
political	83	3.56
additional	78	5.75

Exercise [6] is from Part 1 of @*cross* and the words are taken from the text 'Back to School at Eton College' in chapter 9. It is a matching and translation exercise[9] (there was probably a mistake by the editors in this exercise, since no word was written in number 8).

---

[9] The two forms of the Norwegian translations in the exercise are the two official languages in Norway, *bokmål* and *nynorsk* respectively.

[6]   **108 - Match the words on the left with the correct translation.**
1. pursue   a. læreplan
2. vocational   b. følge
3. curriculum   c. forbedre/forbetre
4. enhancing   d. yrkes-
5. evolve   e. rykte
6. reputation   f. fyrtårnskole
7. academically   g. veiledning/ rettleiing
8.   h. sette sammen/ setje saman
9. comprehensive   i. teoretisk begavet/ teoretisk gaverik
10. compile   j. høyt ansett/ høgt akta
11. Beacon school   k. allsidig, altomfattende
12. coaching   l. utvikle
(Rodgers et al. 2006b:70)

Table 4.10 shows the words from exercise [6] that present collocations in the text. All the collocations are pertinent to the area of education, which is the theme of the chapter. *Comprehensive school* seems to be especially relevant within this topic since it appears five times in the text. *Have a reputation* has only just been included as a collocation, having the highest MI as 3.19, as shown by the table. One of the variations of reputation appears with a negative MI, which indicates that it is definitely a free combination.

Table 4.10 Collocations from exercise [6] in @*cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
comprehensive	comprehensive school (5x)	Adj+N	(105)	
	- comprehensive school		95	3.79
	- comprehensive schools		10	2.36
curriculum	national curriculum	Adj+N	170	4.42
pursue	pursue an education	V+N	(179)	
	- pursue_educations		1	4.30
	- pursue_education		178	3.38
reputation	build (up) a reputation	V+N	(275)	
	- built_reputation		187	5.44
	- builds_reputation		5	3.84
	- build_reputation		35	3.20
	- building_reputation		48	2.69
	have a reputation		(2378)	
	- has		978	3.19
	- had		767	2.52
	- have		561	1.58
	- having		25	1.11
	- 've		37	0.18
	- 'd		10	-1.64
vocational	vocational education	Adj+N	303	8.95

Table 4.11 displays the ten most frequent 4-gram noun collocates of the verb *pursue* in the corpus (V+N). This might be a difficult word for students and its collocates can help clarifying its uses.

Table 4.11 Collocates of the verb *pursue* in COCA

Collocates	Raw frequency	MI
career	509	6.22
interests	311	6.22
policy	279	4.53
goals	258	6.10
careers	215	7.75
policies	202	5.61
strategy	181	5.36
dream	179	5.35
education	178	3.38
goal	148	4.86

Exercise [7] is in Part 1 of @*cross* and the words are taken from chapter 10, 'City of Dreaming Spires'. The words are taken from the main text about Oxford and from the Fact File text about education. This exercise is also a matching and translation exercise.

[7]   **119 - Match the words on the left with the correct translation.**
1. opinion   a. utlandet
2. scholarship   b. klubb for kvinnelig akademikere/ klubb for kvinnelege
3. abroad   akademikere
4. homesickness   c. mening/ meining
5. obtain   d. hjemlengsel/ heimlengt
6. fraternity   e. rådgiver/ rådgivar
7. host   f. stipend
8. counsellor   g. oppnå
9. sorority   h. student, akademiker/ akademikar
10. graduate   i. vert
                                          j. klubb for mannlige akademikere/ klubb for mannlege akademikarar

(Rodgers et al. 2006b:75)

Table 4.12 contains the words from exercise [7] with their collocations. Similarly to the previous exercise, there is a case of a collocation which appears more than once in the texts.

Table 4.12 Collocations from exercise [7] in @*cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
abroad	study abroad (8x)	V+Adv	(304)	
	- studying abroad		55	6.08
	- study abroad		174	4.50
	- studied abroad		23	4.37
	- studies abroad		52	3.58
graduate	graduate student	Adj+N	(3538)	
	- graduate student		1697	8.52
	- graduate students		1841	7.11
opinion	different opinions	Adj+N		
	- different opinion(s)		326	4.60
scholarship	receive a scholarship	V+N	(217)	
	- receive_scholarship		76	5.48
	- received_scholarship		109	5.10
	- receiving_scholarship		25	4.88
	- receives_scholarship		7	4.30

In relation to exercise [7] above, the word combination *apply for a scholarship* was found in the text and presented an MI above 3.0, but since it was not included in the dictionary it was omitted from the table.

Table 4.13 lists the ten most frequent 4-gram noun collocates of the verb *obtain* in the corpus (V+N). A discussion about the differences between *get* and *obtain* could be pertinent here.

Table 4.13 Collocates of the verb *obtain* in COCA

Collocates	Raw frequency	MI
information	802	5.10
data	298	4.41
permission	187	6.81
results	166	3.71
consent	159	7.02
approval	124	5.76
services	124	3.41
sample	114	4.76
copy	112	5.62
degree	104	4.15

Exercise [8] is from the last chapter of Part 1 of *@cross,* 'Glastonbury – Music, Myth and Magic' and it is a matching synonyms/explanations exercise.

[8]   **130 - Draw a line between words and phrases with the same meaning.**

1. legendary	a. mood
2. environmentally	b. moon covers sun
3. charity	c. offering
4. ambience	d. very famous
5. eclipse	e. pedigree
6. ancient	f. just ahead of you
7. sacrifice	g. not commercial
8. mysterious	h. think back on
9. remainder	i. ecological
10. descent	j. very old
11. reminisce	k. what's left
12. imminent	l. incomprehensible

(Rodgers et al. 2006b:80)

Table 4.14 presents the collocations found in the text. The combination *ancient ceremonies* appears in the text, has a raw frequency of 20 in the corpus, and an MI higher than 3.0, but since it was not included in the dictionary it is not shown in table 4.14.

Table 4.14 Collocations from exercise [8] in *@cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
imminent	imminent arrival	Adj+N	49	9.75
sacrifice	human sacrifice(s)	Adj+N	330	6.70

Noteworthy from exercise [8] and its correspondent collocation table (4.14) is the fact that *environmentally* does not appear with any collocates in the text (although *environmental issues* was found), which probably contradicts our intuitions about the use of this word. Table 4.15 lists the most frequent bigram collocates with the adverb *environmentally* (Adv+Adj). The collocates of *environmentally* shown in the table could be used to talk about what types of products or lifestyle would help prevent damaging the environment or even to start a discussion about environmental issues in the US and in the UK. Noticeable from table 4.15 are the high MI scores.

Table 4.15 Collocates of the adverb *environmentally* in COCA

Collocates	Raw frequency	MI
friendly	545	12.61
responsible	316	10.56
sound	286	9.25
sensitive	209	11.04
conscious	145	11.38
sustainable	138	11.73
safe	106	8.71
benign	91	12.08
destructive	52	10.82
aware	45	7.86

Exercise [9] is taken from the same chapter as exercise [8] and its words appear in the Fact File texts. It is the type of exercise in which students need to write a text using a number of words.

[9]   **132 - Charity**
        Combine as many as possible of these word [*sic*] in a text about aid work.
        Global issues, responsibility, poverty, debt, aid, trade, water, sanitation, health, dedication.
        (Rodgers et al. 2006b:81)

Table 4.16 below displays the results for the words from exercise [9]. Three of the words from the exercise were not found in the texts. The word *aid* was found in many combinations, like *aid projects*, *aid organizations*, and *aid initiatives*. All these present MI scores higher than 3.0 but are not included in the table for not being listed in the dictionary. They could, however be mentioned as examples in class to help students write this specific text. The combination *living in poverty* was found in the text and presents a frequency of 309 and an MI of 4.83 in the corpus, but did not appear in the dictionary entries (only in example sentences). On the other hand, *in poverty* appeared in the dictionary but its MI in the corpus was lower than 3.0. Thus, none of them were included as collocations.

The collocation *the burden of debt* is listed in the dictionary both as *burden of debt* and as *debt burden* and was looked for in the corpus as the latter since it fits better in the syntactic category of N+N. It should be noted that the results for *cancel the debt* include some British variations (*cancelled* and *cancelling*), which are less frequent than the American forms since this corpus is of American English.

Table 4.16 Collocations from exercise [9] in @*cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
aid	medical aid	Adj+N	65	3.96
debt	debt cancellation (2x)	N+N	19	8.90
	debt relief	N+N	277	8.21
	cancel the debt	V+N	(35)	
	- canceling_debt		8	6.63
	- cancelling_debt		1	6.50
	- cancel_debt		18	5.75
	- cancels_debt		1	5.14
	- cancelled_debt		2	4.55
	- canceled_debt		5	3.32
	the burden of debt			
	- debt burden	N+N	24	4.59
dedication	*not in the text*			
health	*not in the text*			
responsibility	*not in the text*			
water	drinking water	N+N	1309	7.38
	pure water	Adj+N	115	4.44

Table 4.17 shows the most frequent bigram noun collocates of the noun *trade* (N+N). The fact that the table shows the collocate *center* as the most frequent collocate can be used for a discussion on how current events (as the attack on the World Trade Center) can have an impact on the language.

Table 4.17 Collocates of the noun *trade* in COCA

Collocates	Raw frequency	MI
center	3615	7.63
agreement	1613	8.32
organization	822	6.99
deficit	811	8.77
commission	795	7.45
group	642	4.75
agreements	549	8.96
policy	546	5.40
unions	521	8.80
barriers	454	8.65

Exercise [10] is taken from Part 3 of @*cross*, called '@cross The USA', and its words are taken from chapter 1, 'An Invitation – Getting Ready'. Of the fourteen exercises selected from @*cross*, none was taken from Part 2, called '@cross the English speaking Countries'

(Jamaica, South Africa, and New Zealand). The reason is that none of them met the criteria established in the material and methods chapter.

[10]   **190 - Write a paragraph describing American teenagers and their situation using the following words.**
         Sports, prom, weird, wired, bullying, committed, frustrated, happy, shopping, trend, cliques
         Add other words from the texts which you think fit in.
          (Rodgers et al. 2006b:111)

Table 4.18 below shows the words and collocations of exercise [10]. The combinations *suffer from bullying* and *systematic bullying* were found in the text, but do not appear in the corpus or the dictionary. The word *clique* is not among the 9.000 entries of the dictionary and therefore is omitted here. However, it was found in the text in combinations like *member of a clique*, *belong to a clique*, and *clique identity*. The two first were found in the corpus with MI scores higher than 3.0 while the latter is not in the corpus. The word *prom* is not included in the collocation dictionary, but in the text it appears as *junior prom* and *senior prom*, which might be interesting for students within the topic of school in the US (their frequency and MI scores and were 33/ 9.13 and 148/ 9.00 respectively). The combination *wired teenagers* was not found in the dictionary or the corpus. One last note concerning table 4.18 is the appearance of the misspelled form *sports *heros* in COCA and how the MI score is affected by low occurrence, meaning that even words that do not exist might get an MI score higher than 3.0 (see subsection 3.3.1).

Table 4.18 Collocations from exercise [10] in @*cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
frustrated	feel frustrated	V+N	(203)	
	- feel frustrated		111	5.32
	- feeling frustrated		32	4.95
	- feels frustrated		13	4.79
	- felt frustrated		47	4.53
shopping	shopping lists	N+N	(413)	
	- shopping lists		68	8.12
	- shopping list		345	7.98
sports	sports hero	N+N	(157)	
	- sports heros		1	6.65
	- sports heroes		92	6.63
	- sports hero		64	5.18
trend	fashion trends	N+N	(82)	
	- fashion trend(s)		81	6.37
	- fashions trend(s)		1	3.92
	latest trends	Adj+N	58	6.68

Table 4.19 outlines the most frequent bigram adjective collocates with the noun *prom* (Adj+N). This is a case when sociocultural aspects are clearly conveyed by combinations, as

in *white prom*, *integrated prom*, and *separate prom*. The concordance lines could be checked for more information.

Table 4.19 Collocates of the noun *prom* in COCA

Collocates	Raw frequency	MI
senior	154	9.97
junior	33	9.07
white	17	4.50
integrated	12	8.39
high-school	8	10.15
separate	7	6.04
junior-senior	4	15.73
pink-chiffoned	2	17.90
frilly	2	10.47
fancy	2	6.32

Exercise [11] is taken from Part 3 of @*cross* and its words are from chapter 2, 'Arrival'. It is the type of exercise in which students have to write a short text using the given words.

[11]   **203 Use as many as possible of the following words in a paragraph about travelling:**
       Arrival, approach, immigration agent, relatives, stay (n), purpose, stamp, customs, exit, vegetables
       (Rodgers et al. 2006b:115)

Table 4.20 below presents the three words which appear with a collocate in the text. The collocation with *purpose* is usually *the purpose of*, but it was looked in the corpus without the article in order to follow the syntactic categories established in chapter 3.

Table 4.20 Collocations from exercise [11] in @*cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI

customs	customs agent
	  - customs agents
	  - customs agent
		N+N	(180)
			106
			74
				
				7.88
				7.19

purpose	(the) purpose of	N+P	13128	3.69

stamp	stamped the passport
	  - stamp_passports
	  - stamp_passport
		V+N	(44)
			17
			27
				
				8.34
				7.64

Table 4.21 shows the results for the ten most frequent bigram adjective collocates of the noun *vegetable* in the corpus (Adj+N). This is a word that is probably well-known by students but whose collocates might not yet be in students' productive knowledge. The collocates could be categorized into two lists: the quality of the vegetables (*fresh*, *green*, *leafy*, and *organic*) and how you eat them (*grilled*, *raw*, *frozen*, *cooked*, *steamed*, and *roasted*).

Table 4.21 Collocates of the noun *vegetables* in COCA

Collocates	Raw frequency	MI
fresh	336	7.63
green	187	5.98
leafy	140	11.14
grilled	132	9.56
raw	109	7.65
frozen	89	7.29
organic	88	7.43
cooked	87	7.89
steamed	86	10.07
roasted	86	9.11

Exercise [12] is in Part 3 of *@cross* and the words included are taken from chapter 3, 'A Day at Plimoth Plantation'. This is a matching antonyms exercise and only the words from the first column have been taken from the text in the textbook.

[12] **211 - Pair off words that have opposite meanings.**

1. Attractive
2. Celebrate
3. Remain
4. Protect
5. Innocent
6. Eager
7. To make it
8. Keep up
9. Independence
10. Equality
11. Delicious
12. Victim
13. Strict

a. Inequality
b. Expose
c. Indulgent
d. Awful
e. Ugly
f. Molester
g. Careless
h. Dishonor
i. Guilty
j. Fail
k. Stop
l. Dependence
m. Disappear

(Rodgers et al. 2006b:119)

Table 4.22 below lists the words which have collocations in the text. It should be noted that *to make it* and *keep up* were not checked in the corpus for being already word combinations.

Table 4.22 Collocations from exercise [12] in *@cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
delicious	delicious meal - delicious meal - delicious meals	Adj+N	(78) 49 29	 8.35 8.11
victim	victim of	N+P	8295	3.16

Table 4.23 displays the ten most frequent bigram noun collocates of the adjective *attractive* in the corpus (Adj+N). The table shows that this adjective is used over four times more frequently to characterize women than men among the combinations with an MI higher than 3.0 (*woman*, *women*, and *girl* comprise 394 tokens, against 89 for *man*). This indicates some sociocultural issues that could be brought to discussion. On the other hand, the table shows

that the strongest combinations are the ones in which the adjective *attractive* does not refer to people (e.g. *alternative*, *option*, and *feature*).

Table 4.23 Collocates of the adjective *attractive* in COCA

Collocates	Raw frequency	MI
woman	260	6.17
alternative	99	7.25
option	97	7.77
man	89	3.67
women	85	3.86
place	62	3.55
girl	49	4.80
features	37	5.57
feature	36	6.40
investment	31	5.23

Exercise [13] belongs to Part 3 of @*cross* and it is in chapter 4, 'Boston'. It is a purely translation exercise with words related to education.

[13]   **227 - Education Vocabulary**
      Translate the following words into Norwegian.

Grades	Subjects
Curriculum	Classes
Scholarships	Tuition
To graduate	Core courses
Electives	Degree

  (Rodgers et al. 2006b:129)

Two of the words presented in the exercise above are not in the chapter texts, as table 4.24 shows. As explained in chapter 3, the collocation *get good grades* (V+Adj+N) follow the definition of a collocation, it combines the syntactical categories devised in chapter 3, and can be considered relevant for teaching. For all these reasons it is included in the table. This collocation was written in the corpus as *[get] good grades* in the 'word(s)' field. The combination *tuition costs* was found in the text and, in the corpus, it presented a frequency of 81 and an MI of 7.51. However, it is not listed in the dictionary and so it is omitted.

Table 4.24 Collocations from exercise [13] in @*cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
classes	*not in the text*			
curriculum	national curriculum	Adj+N	170	4.47
electives	*not in the text*			
grades	get good grades	V+Adj+N	(177)	
	- getting good grades		51	9.09
	- gets good grades		12	7.99
	- gotten good grades		4	7.68
	- get good grades		82	7.59
	- got good grades		28	6.81

Table 4.25 lists the ten most frequent 4-gram verb collocates of the noun *scholarship* in the corpus (V+N). It shows that this word is commonly used together with the verbs *win*, *receive*, *offer*, *earn*, *award*, and *provide*, which could be categorized as verbs for giving a scholarship (*offer*, *award*, and *provide*) and verbs for receiving it (*receive*, *win*, and *earn*).

Table 4.25 Collocates of the noun *scholarship* in COCA

Collocates	Raw frequency	MI
won	146	5.29
received	109	5.04
offered	97	5.20
receive	76	5.42
win	74	4.32
earned	69	5.97
offer	62	4.13
awarded	58	7.09
provide	57	3.41
earn	37	5.61

Exercise [14] below is taken from Part 3, chapter 6 'Washington D.C.' of @*cross*. It is a translation exercise and it is the last one selected from @*cross workbook*.

[14]   **242 - Put the words in alphabetical order and translate them into Norwegian**.

Words	Alphabetical Order	Translation
Majority		
Bill		
Election		
Support		
Constitutional		
Consider		
Legislative		
Judicial		
Executive		
Law		
Agenda		
Minority		

(Rodgers et al. 2006b:136)

Many of the words from exercise [14] have verbs and nouns as collocates in the text and for this reason they present a great number of form variations, as provided in table 4.26 below. As shown in the table, there are two words presented in the exercise which do not appear in the text. It should be noted that the combination *in the majority* (P+Art+N) is the same case as *get good grades* explained above. In the case of the word *constitutional*, no collocation was found in the text and for this reason it is not included in table 4.26. On the other hand, the derivative

unconstitutional appeared three times in the text in the collocation *law_unconstitutional* (MI 6.16, frequency 112).

Table 4.26 Collocations from exercise [14] in *@cross*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
agenda	*not in the text*			
bill	pass a bill	V+N	(1280)	
	- pass_bill		566	4.24
	- passed_bill		561	4.08
	- passing_bill		98	2.74
	- passes_bill		55	2.49
	sign a bill		(725)	
	- signed_bill		346	4.30
	- signing_bill		63	3.67
	- sign_bill		266	3.05
	- signs_bill		50	1.40
consider	consider necessary	V+N	101	3.33
election	hold an election	V+N	(502)	
	- hold_election		277	3.60
	- holding_election		106	2.99
	- holds_election		28	1.86
	- held_election		91	1.77
	free election(s)	Adj+N	564	5.72
executive	executive branch	Adj+N	(1377)	
	- executive branch		1333	9.90
	- executive branches		44	5.43
judicial	judicial branch	Adj+N	(146)	
	- judicial branch		112	9.11
	- judicial branches		34	7.84
law	pass a law	V+N	(1642)	
	- passed_law		821	4.35
	- pass_law		632	4.12
	- passing_law		129	2.85
	- passes_law		60	2.33
legislative	legislative branch	Adj+N	(337)	
	- legislative branch		233	9.95
	- legislative branches		104	9.24
majority	in the majority	Prep+Art+N	638	3.24
support	*not in the text*			

Table 4.27 shows the ten most frequent bigram noun collocates with the adjective *constitutional* (N+N). The collocates from the table suggest that this word is restricted to politics and law and, as such, presenting them to students might be an aid to the understanding of the word's meaning.

Table 4.27 Collocates of the adjective *constitutional* in COCA

Collocates	Raw frequency	MI
amendment	1112	10.32
rights	782	7.63
right	650	4.62
law	598	6.53
convention	236	7.81
court	180	4.98
crisis	161	6.52
amendments	151	9.73
protection	147	6.45
issues	143	4.93

These were all the exercises from @*cross*, their collocations found in the textbook, and the combinations looked for in COCA. The total items of vocabulary included in these fourteen exercises are 156, from which only nine are chunks and not single words (5.9%). Here are these combinations and the exercise number in parenthesis: *according to* [1], *apply to* [2], *change the subject* [3], *academically gifted* [6], *Beacon school* [6], *global issues* [9], *immigration agent* [11], *to make it* [12], *keep up* [12], and *core courses* [13]. In the next section we turn to the results of the second part of the material, the *Passage* website.

4.4 Results from *Passage*

The exercises from *Passage* are taken from its website, where fourteen vocabulary exercises that present lists of single words were found. There are a total of seventy vocabulary exercises on the website and their types and number are:

- crossword (23)
- matching synonyms or definitions (15)
- word grid (13)
- gap-filling or completing a sentence (7)
- translation (6)
- collocation (3)
- working with compounds (2)
- setting words into categories (1)

As stated in chapter 3, the first aim requires exercises that work with single words. Thirteen of the match synonyms or definitions exercises and one involving translation (see table 3.1) have been chosen, all of them containing single words. I did not choose more of the type involving Norwegian translation because, as opposed to the ones selected in @*cross*, this type of exercise in *Passage* does not present the keywords in English, but in Norwegian. As in the previous section, the words that are already presented as combinations are not included in the

search, and the words which have no collocations in the text are omitted from the collocation tables. Pertaining numbering, the exercise numbers follow the order started in @*cross*, going in this section from 15 to 28 (they are not numbered on the website). Among the texts used for the exercises seven are authentic, four are constructed and three are mixed.

Exercise [15] is a matching synonyms exercise and it refers to the first part of the *Passage* textbook, which is an introduction to the other units. The text they are taken from is a short story by Ernest Hemingway, an authentic text.

[15]   **Starting Out - "A Day's Wait"**
          Pair up synonyms
          Pair up the words on the left (as used in the text) with their synonyms on the right.
          evidently   distant
          capsule   begin
          commence   different
          slack   tablet
          detached   tired
          various   clearly
          (Sørhus et al. 2006b)

Most of the words from exercise [15] have no collocations in the text they are taken from, which is the reason why table 4.28 only contains the collocation *detached from*. Concerning this collocation, the lemma was looked for in the corpus like all the other words and not the exact form *detached*.

Table 4.28 Collocations from exercise [15] in *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
detached	detached from	V+P	493	6.46

Table 4.29 displays the ten most frequent 4-gram verb collocates of the noun *slack* in the corpus (V+N). The table shows that the collocates of *slack* can demonstrate the different meanings of this word.

Table 4.29 Collocates of the noun *slack* in COCA

Collocates	Raw frequency	MI
cut	157	5.92
pick	116	6.79
take	99	3.56
wearing	59	6.19
went	53	3.58
dressed	46	6.60
wore	40	6.47
picking	32	6.73
picked	29	5.24
taking	29	3.58

Exercise [16] is also from the same introductory part of *Passage* and its words are taken from a newspaper article by Sarah Kershaw (an authentic text). The heading of the exercise points out that the words are taken from the text, which, in contrast with @*cross,* happens indeed in all the *Passage* exercises.

[16]   **Starting Out - "A Solitary Elephant Poses a Big Problem in Alaska"**
Link the words on the left (taken from the text) with their synonyms on the right.

debate	worry
anguished	discussion
acknowledge	worried
relocate	make friends
frigid	admit
consternation	move
socialize	suggestion
proposal	cold

(Sørhus et al. 2006b)

The words from exercise [16] appear in a number of free combinations in the text (*yearlong debate*, *frigid afternoon*, *socialize with*, and *unusual proposal*) but only in two collocations, which are listed in table 4.30 below.

Table 4.30 Collocations from exercise [16] in *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
consternation	cause consternation	V+N	(81)	
	- caused_consternation		48	8.78
	- causing_consternation		11	8.11
	- causes_consternation		10	7.09
	- cause_consternation		12	5.80
debate	a debate over (sth)	N+P	3728	6.05

Table 4.31 exhibits the ten most frequent bigram adjective collocates of the noun *debate* (Adj+N). The table shows that all the collocates of *debate* are associated with the topic of politics and current issues, the strongest ones being *heated*, *presidential*, and *congressional*.

Table 4.31 Collocates of the noun *debate* in COCA

Collocates	Raw frequency	MI
public	805	5.43
presidential	633	7.90
political	567	5.16
national	512	4.81
heated	344	9.36
current	219	4.99
ongoing	218	4.99
great	214	3.39
congressional	160	6.60
open	139	3.47

Exercise [17] is a matching synonyms exercise and it practices words taken from an autobiography called 'When I was Puerto Rican' from 'Global English', the second section of *Passage*. This text is written by Esmeralda Santiago and it is an authentic text.

[17]   **Global English - "When I was Puerto Rican"**
        Link up the words that mean more or less the same.
        notify   cloudy
        monologue   pronounce
        clasp   shout
        overcast   accompany
        yell   inform
        enunciate   hold
        escort   brow
        forehead   speech
        (Sørhus et al. 2006b)

Table 4.32 below includes the collocates of the words from exercise [17]. The combination *overcast morning* was found in the corpus with an MI above 3.0 (5.98) and raw frequency of 12, but it does not appear in the dictionary.

Table 4.32 Collocations from exercise [17] in *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
clasp	clasped my hands - clasp_hands - clasp_hand	V+N	(652) 524 128	 8.48 5.76
enunciate	enunciate (every) word - enunciate_word - enunciate_words	V+N	(50) 35 15	 7.06 5.51
yell	yell at	V+P	2584	5.28

Table 4.33 contains the ten most frequent 4-gram noun collocates of the verb *clasp* in the corpus (V+N). Looking at the table, we can see that the collocates of this word comprise parts of the body (mainly *hand*) as its collocational field. It should be noted that the word *prayer* here is part of the expression *hands clasped in prayer*, where the collocate *hand(s)* appears again. Furthermore, many of the occurrences of *knees* are also together with *hand(s),* like in the examples *his hands clasped between his knees* and *his hands clasped around his knees*. Also in the case of *back*, the concordance lines show that *back* and *hands* appear in the same sentences, most of them being close variations of *his hands clasped behind his back*. The same happens with the collocate *head*, where many of the examples are of the kind *hands clasped behind/above/under his head*. What becomes clear by looking at the concordance is the strong bonding between *clasp* and *hand* (also shown by the MI of 8.42). Most of the

occurrences of *clasp* together with *knife* are instances of the compound word *clasp knife* (US *pocketknife*) but there are occurrences of the collocation *clasp a knife* as well.

Table 4.33 Collocates of the verb *clasp* in COCA

Collocates	Raw frequency	MI
hands	524	8.42
hand	128	5.70
back	72	3.06
head	34	3.67
arms	29	5.10
knees	24	6.61
prayer	17	6.38
knife	17	6.38
shoulder	16	5.12
neck	15	5.14

Exercise [18] is also taken from the 'Global English' unit and the words are from a short story called 'The Shining Mountain' by Alison Fell, an authentic text. The exercise is again of the matching synonyms type.

[18]   **Global English - "The Shining Mountain"**
           Join words that mean the same (or nearly)!
           cast   slide
           weep   throw
           pant   tired
           weary   cry
           skid   burden
           shin   climb
           load   breathe heavily
           (Sørhus et al. 2006b)

Table 4.34 lists the words from exercise [18] with their collocations found in the text.

Table 4.34 Collocations from exercise [18] in *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
load	carry loads	V+N	(418)	
	- carrying_load		152	5.72
	- carry_load		180	5.35
	- carries_load		30	4.44
	- carried_load		56	3.85
weary	grow weary	V+Adj	(233)	
	- grown weary		90	9.19
	- grow weary		50	7.79
	- grew weary		44	7.59
	- growing weary		43	6.89
	- grows weary		6	6.80
weep	weep _ tears	V+N		
	- weep_tears		69	6.99
	- weep_tear		3	3.84

Three words from exercise [18] appear in free combinations in the text and therefore are not included in table 4.34. These are: *lungs panted*, *skidding and sliding*, and *shin rocks*.

Table 4.35 presents the ten most frequent 4-gram noun collocates of the verb *cast* in the corpus (V+N). The collocates of *cast* reveal its several meanings, as for example in *cast a shadow*, *cast a vote*, *cast doubt*, *cast iron*, and *cast of characters*.

Table 4.35 Collocates of the verb *cast* in COCA

Collocates	Raw frequency	MI
shadow	605	7.89
vote	574	6.22
shadows	500	8.17
characters	415	7.07
ballots	401	9.13
doubt	384	6.21
members	377	4.47
iron	374	7.11
light	308	4.06
votes	285	6.61

Exercise [19] is from the third section of *Passage*, 'On the Road', and the words are taken from the short story 'Paradise' by Matthew Kneale, an authentic text. This is another matching synonyms exercise.

[19]   **On the Road - "Paradise"**
    Pair up words that mean the same – or nearly the same.

snap	break
push	pull
sip	dislike
resent	look
glance	drink
tug	disappear
vanish	shove

    (Sørhus et al. 2006b)

Table 4.36 below contains the words from exercise [19] with their collocations. The collocation *glanced at*, as indicated in the table, was found five times in the text. Two other combinations were found in the text and in the corpus but not in the dictionary: *tugging his wrist* and *doubts vanished*.

Table 4.36 Collocations from exercise [19] from *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
glance	glanced at (5x)	V+P	7360	5.05
push	pushed through	V+P	1470	4.72
sip	took a sip	V+N	(1385)	
	- took_sip		820	5.67
	- takes_sip		254	5.18
	- taking_sip		133	3.88
	- take_sip		153	2.29
	- taken_sip		25	1.49
	sipping tea	V+N		
	- sip_tea		550	8.37

Table 4.37 arranges the ten most frequent 4-gram noun collocates of the verb *snap* in the corpus (V+N). These collocates reveal some distinct meanings of *snap*, such as in *snap your fingers*, *snap pictures/photos*, *snap peas* and *snap beans* (compounds), and *snap to attention*.

Table 4.37 Collocates of the verb *snap* in COCA

Collocates	Raw frequency	MI
fingers	686	7.87
pictures	205	6.01
picture	199	5.21
peas	161	8.70
attention	143	4.23
photos	131	6.15
photo	86	3.20
beans	76	6.02
neck	65	4.51
shot	56	3.22

Exercise [20] below belongs to the fourth part of *Passage,* called 'We Were Here First' and the text where the words are taken from is a factual text, as the book defines it, about the Indians in America. Unlike all the previous exercises, this one is based on a constructed text.

[20]   **We Were Here First - "Native Americans - The Original Inhabitants"**
          Link the words that mean the same – or nearly the same.
          perish   important
          innumerable   newcomer
          notable   stupid
          tribe   nation
          foolhardy   many
          diverse   die
          settler   varied
          (Sørhus et al. 2006b)

Noticeable from exercise [20] is that although the words in the two columns are synonyms (e.g. *foolhardy* and *stupid*, *perish* and *die*, *notable* and *important*) their use might be distinct

in terms of collocations or register. One way to find out these differences is to check their concordance lines in the corpus or the most frequent collocates of each of them.

Table 4.38 lists the words with their collocations found in the text. There is a combination with the word *foolhardy* in the text, *a foolhardy attack*, which was not found in the corpus or the dictionary. Another combination not found in the dictionary was *Indian tribe*, although it appeared in the corpus with an MI of 8.73 and frequency of 658. The fact that *European settlers* is a collocation but not *African settlers* for example might be interesting for discussion in terms of historical facts.

Table 4.38 Collocations from exercise [20] in *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
diverse	diverse group - diverse groups - diverse group	Adj+N	(589) 219 370	 5.53 5.42
notable	notable exception - notable exceptions - notable exception	Adj+N	(437) 186 251	 11.06 10.17
settler	European settlers	Adj+N	145	7.75
tribe	nomadic tribes	Adj+N	40	10.10

Table 4.39 shows the ten most frequent bigram noun collocates of the adjective *notable* in the corpus (Adj+N). What stands out in this table is the repetition of words, indicating that *notable* has fewer frequent collocates than the ones seen so far so that the ten most common collocates are variations of only six words (*exception*, *example*, *feature*, *performance*, *difference*, and *success*).

Table 4.39 Collocates of the adjective *notable* in COCA

Collocates	Raw frequency	MI
exception	266	11.19
exceptions	190	12.02
example	79	6.00
feature	49	8.15
performances	46	9.12
examples	44	7.94
differences	40	6.59
successes	37	9.81
difference	35	6.08
features	30	6.57

Exercise [21] below is from the same part of *Passage* as the previous one, and the words are taken from a text about the aborigines in Australia, a constructed text. The exercise is to match synonyms.

[21]   **We Were Here First - "Aboriginal Australians"**
       The verbs on the left are all taken from the text. Link them with their synonyms on the right.
       adapt   hunt
       emerge   come out
       take place   change
       christen   name
       provide   give
       stalk   happen
       (Sørhus et al. 2006b)

Only two of the words in the exercise above are part of collocations in the text, as table 4.40 shows. The combination *provide (someone) with* was found in the text and in the dictionary but its MI was below 3.0.

Table 4.40 Collocations from exercise [21] in *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
adapt	adapt to	V+P	3590	3.68
emerge	emerge from	V+P	8252	6.10

Table 4.41 presents the ten most frequent 4-gram noun collocates of the verb *provide* in the corpus (V+N). As in the case of *obtain* versus *get* (see comment about table 4.13), it would be interesting to compare the uses of *provide* and *give*. The comparison might reveal not only differences in collocations but also in register. In contrast with previous tables, the frequencies in this table are higher while the MI scores are lower.

Table 4.41 Collocates of the verb *provide* in COCA

Collocates	Raw frequency	MI
information	4515	4.98
services	2785	5.28
support	2606	4.36
care	1817	3.77
evidence	1730	4.56
opportunities	1621	6.03
opportunity	1446	4.99
access	1208	4.76
service	1083	3.50
assistance	1071	5.77

Exercise [22] is from the same part as the two previous ones and it is also a matching synonyms exercise. This is the first exercise whose words are taken from a song, 'Took the Children Away' by Archie Roach, an authentic text. The song lyrics are displayed as a text in the textbook.

[22]   **We Were Here First - "Took the Children Away"**
Connect the words on the left with the synonyms on the right.

origins	make someone feel stupid
snatch	subject, topic
adopt	take away
apologize	say sorry
orphanage	roots
humiliate	foster
issue	children's home

(Sørhus et al. 2006b)

Table 4.42 exhibits the collocations found in the text with the words from exercise [22]. It should be noted that unlike most of the grammatical collocations that appeared before in this chapter, *snatch_from* was entered in the corpus as a 4-gram and not a bigram, since there is usually an object between the verb and the preposition. Moreover, although the collocation found in the text is in the past tense, the lemma SNATCH was entered in the corpus. The combination *adopted by* was found in the text and in the corpus but not in the dictionary.

Table 4.42 Collocations from exercise [22] in *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
apologize	apologize for	V+P	2282	4.23
snatch	snatched_from	V+P (4-gram)	752	3.91

Table 4.43 below lists the ten most frequent 4-gram noun collocates of the verb *adopt* in the corpus (V+N). The table reveals that the word *adopt* has stronger collocates when used in its more abstract sense of 'to accept or start something new' (*CALD* 2005:17), as in *adopt a resolution* (MI 6.07), *adopt policies* (MI 5.88), and *adopt an approach* (MI 5.07). When the sense is 'to take a child', the MI scores are slightly lower. This word might be useful to talk about politics, for example in phrases like *the government has adopted (a) new policy/approach/strategy/measures*.

Table 4.43 Collocates of the verb *adopt* in COCA

Collocates	Raw frequency	MI
children	395	3.38
policy	306	4.48
approach	294	5.07
child	285	3.89
policies	278	5.88
strategy	188	5.23
resolution	182	6.07
standards	164	4.85
baby	156	4.06
measures	139	4.94

Still from the fourth part of the *Passage* material, exercise [23] is the only one based on a listening activity. The text students listen to is 'Dracula' by Abraham 'Bram' Stoker and the teaching material indicates that the text was 'slightly adapted' (Sørhus et al. 2006b). The transcripts are provided only in the teacher's resources on the website.

[23]   **We Were Here First - "Dracula"**
The words on the left are taken from the text. Link them up with words that mean the same - or nearly the same.

aid	story
fancy	faint
swoon	help
narrative	memory
remembrance	confused
cleanse	imagining
bewildered	drink
draught	wash

(Sørhus et al. 2006b)

Since this exercise is based on a literary text with many infrequent words, the combinations found in the transcript are either rare or inexistent in the corpus and none of them is listed in the dictionary. Some of these are: *sleeping draught* (appeared twice in the text and presented an MI of 9.82 in the corpus), *horrible fancies* and *awful narrative* (found in neither), and *in a half swoon* (found in the corpus with an MI of 4.52). Therefore, there is no collocation table for exercise [23].

Table 4.44 presents the ten most frequent 4-gram noun collocates of the verb *cleanse* in the corpus (V+N). As the table shows, the collocates are not so frequent compared to previous tables and their MI scores are mostly close to 3.0, except for *palate*. It might be interesting to contrast the uses of *cleanse* with the synonym *wash* given by the exercise.

Table 4.44 Collocates of the verb *cleanse* in COCA

Collocates	Raw frequency	MI
body	29	4.82
skin	16	5.39
face	16	3.53
blood	15	4.76
area	14	3.76
system	14	3.24
palate	13	10.37
mind	12	3.81
land	11	4.04
air	10	3.27

Exercise [24] is taken from the fifth part of *Passage*, called 'The Road to Freedom', and students are supposed to match words with their definitions. The text where the words are

taken from is a factual text with excerpts from autobiographies, so that it is partly constructed and partly authentic.

[24]   **The Road to Freedom - "The Road to Freedom - Three Key Figures"**
        Which noun fits the definition?

Part of a train where the passengers sit:	obligation
The feeling of bitterness at some injustice:	resentment
Something you have a duty to do:	cowardice
Reaching agreement by talking together:	negotiation
Attitude not based on rational thinking:	prejudice
Showing a lack of courage:	compartment

(Sørhus et al. 2006b)

Table 4.45 displays the collocations found in the text. It stands out that the collocation *color prejudice*, even though appearing twice in the text, has a raw frequency of 8 tokens and an MI score of 3.04, barely fitting the definition of collocation provided in chapter 1.

Table 4.45 Collocations from exercise [24] in *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
negotiation	negotiations with	N+P	1717	4.04
obligation	fulfill my obligation - fulfill_obligation - fulfills_obligation	V+N	(202) 198 4	 9.34 6.70
prejudice	color prejudice (2x)	N+N	8	3.04
resentment	resentment towards	N+P	27	7.33

Table 4.46 lists the ten most frequent bigram noun collocates with the noun *prejudice* in the corpus (N+N). It is noticeable that *prejudice* has a stronger bonding to *racial* (MI 10.02) and *ethnic* (6.58) than with the collocate *color*, which appears in the text. The difference between the raw frequencies is also visible, especially between *racial* (202) and *color* (8). The question here is why this infrequent and weak collocation appeared in the text instead of the more common, stronger, and similar in meaning, *racial prejudice*.

Table 4.46 Collocates of the noun *prejudice* in COCA

Collocates	Raw frequency	MI
racial	202	10.02
social	40	4.60
old	34	3.90
ethnic	27	6.58
cultural	24	5.40
religious	23	5.32
personal	22	4.70
blatant	19	10.24
subtle	17	7.41
sexual	17	5.03

Exercise [25] contains words from the same text as the previous one. In this exercise students have to match synonyms.

[25]   **The Road to Freedom - "The Road to Freedom - Three Key Figures"**
Join up the words that mean the same - or nearly the same.

murmur	mutter
resentment	answer
persist	choose
select	continue
assassinate	keep separate
retort	anger
segregate	murder

(Sørhus et al. 2006b)

Table 4.47 exhibits the results for the words in exercise [25]. No combinations were found with the verb *segregate* in the text, only with other forms, such as *segregated society* (37 tokens, MI 6.71), *racial segregation* (213 tokens, MI 11.02), and *patterns of segregation* (9 tokens, MI 5.13). Among these, *racial segregation* is listed in the *OCD*, but since it belongs to another part of speech than the word in the exercise, it was not included in the table.

Table 4.47 Collocations from exercise [25] in *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
murmur	murmured politely	V+Adv	5	7.26
persist	persist in	V+P	1200	3.45
resentment	resentment towards	N+P	27	7.33

Table 4.48 presents the ten most frequent 4-gram noun collocates of the verb *assassinate* in the corpus (V+N).

Table 4.48 Collocates of the verb *assassinate* in COCA

Collocates	Raw frequency	MI
president	129	5.29
leaders	32	5.58
king	22	5.02
minister	15	5.27
leader	15	4.83
officials	12	3.72
Pope	9	6.20
character	9	4.29
commander	6	5.32
militant	5	7.02

The collocates of *assassinate* listed in table 4.48 above confirm its definition, which is 'kill someone famous or important' (*CALD* 2005:66), as seen in collocates like *president*, *king*, and *Pope*. In fact, all of the collocates could be said to be 'important people' except from *assassinate one's character,* which is used when someone degrades somebody's image or

reputation. The collocates of the verb *assassinate* could be contrasted with the ones from *murder*, which is given as the synonym in the exercise.

Exercise [26] takes its words from the sixth part of *Passage*, called 'The Power and the Glory' from a constructed text about the English monarchs. In this exercise students have to match words with their Norwegian translations.

[26]   **The Power and the Glory - "English Monarchs"**
Match the English words with their Norwegian translations:

resistance	avstamning
infidelity	stamme
tribe	erobre
execute	nymotens
allegiance	knust
descent	henrette
conquer	blodbad
source	utroskap
new-fangled	motstand
devastated	kilde
slaughter	troskap

(Sørhus et al. 2006b)

Most of the words above are not part of any collocations in the text, as table 4.49 below shows. The combination *promise allegiance* was found in the text, but it was not found in the dictionary and, in the corpus, it is infrequent (3 tokens) and has an MI below 3.0. The combination *Germanic tribes*, on the other hand, was found in the corpus with an MI of 10.45 but it is not listed in the dictionary. Thus, these were omitted from table 4.49.

The adjective *new-fangled* should be given some extra attention. Firstly, it is unusual that *new-fangled* is presented as a single word and not a collocation in the exercise, especially when it collocates strongly with *ideas*, as it was found in the text. Furthermore, when looking for the most used nouns with *new-fangled*, *ideas* and *idea* are the ones on top of the list, followed by other combinations with frequencies lower than 3. In the case of the word *resistance*, the collocation found in the text was *put up fierce resistance* (V+Adj+N), listed both in the corpus and the dictionary and another exception with regard to the syntactical form (like *get good grades* and *in the majority*). If we look for *put up_resistance*, allowing a word between *up* and *resistance*, and looking for the lemma PUT (entered as *[put] up * resistance*), COCA shows 35 tokens of the type *put up no resistance*, *put up little resistance*, *put up stiff resistance*, and *put up tremendous resistance*. This exemplifies other ways in which collocations with more lexemes can be found in the corpus and shows how much variation of results can be encountered. This is the reason why I have devised simpler criteria than the whole spectrum available in the corpus.

Table 4.49 Collocations from exercise [26] in *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
descent	direct descent	Adj+N	8	4.63
new-fangled	new-fangled ideas	Adj+N	(10)	
	- new-fangled ideas		3	9.30
	- new-fangled idea		3	7.97
	- newfangled ideas		3	7.32
	- newfangled idea		1	4.40
resistance	put up fierce resistance	V+Adj+N	3	7.24

Table 4.50 below lists the ten most frequent 4-gram verb collocates of the noun *allegiance* in the corpus (V+N). According to the table, the verb *pledge* is the most frequent and strongest collocate of *allegiance* in the corpus and it appears in the dictionary. It did not, however, appear in the text while the free combination *pledge allegiance* was included. Thus, the question can be raised again: Why would the textbook writers include the weak and rare combination *promise allegiance* in the constructed text instead of the more common and stronger collocation *pledge allegiance*? This matter will be discussed further in chapter 5.

Table 4.50 Collocates of the noun *allegiance* in COCA

Collocates	Raw frequency	MI
pledge	82	10.04
switched	35	8.56
swear	34	8.68
pledged	33	9.21
owe	30	8.35
pledging	25	11.14
swore	21	9.18
declare	17	7.96
owed	16	8.07
sworn	15	8.22

Exercise [27] is from the same section as [26], and the text the words are taken from is an excerpt from the novel 'Animal Farm' by George Orwell, thus, an authentic text. It is a matching definitions exercise.

[27]   **The Power and the Glory - "Animal Farm"**
         Match the words on the left with the correct definition in the right hand column.

oppression	the practice or extending a state's rule over other territories
totalitarian	the theory that a country's wealth should belong to the people as a whole
philosophy	an open or armed fight against a government
socialism	characteristic of a one-party state
rebellion	the search for knowledge and truth, especially about the nature of man and his behavior and beliefs
imperialism	to be governed cruelly

(Sørhus et al. 2006b)

Only two words from the exercise above are part of collocations in the text, as displayed by table 4.51. *British imperialism*, in the same way as *European settlers* in exercise [20], is a collocation which shows a historical aspect that can be discussed in class.

Table 4.51 Collocations from exercise [27] in *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
imperialism	British imperialism	Adj+N	22	7.55
totalitarian	totalitarian regime - totalitarian regimes - totalitarian regime	Adj+N	(210) 109 101	 12.97 10.99

Table 4.52 presents the ten most frequent bigram noun collocates of the adjective *totalitarian* in the corpus (Adj+N). Since this section of *Passage* deals with power and politics, this is one of the words whose collocates could be explored in class. Some of the collocates show a stronger bonding to *totalitarian*, especially *totalitarian regime* and *totalitarian dictatorship*.

Table 4.52 Collocates of the adjective *totalitarian* in COCA

Collocates	Raw frequency	MI
regimes	109	12.97
regime	101	10.99
state	71	5.61
system	56	6.97
government	33	5.99
states	29	5.66
society	27	6.96
rule	21	7.89
dictatorship	19	11.61
governments	18	8.45

Exercise [28] contains words taken from the last section of *Passage*, called 'Global Village'. It is a matching synonyms exercise and the words are taken from a (constructed) text about football.

[28]   **Global Village - "Want to buy a football club?"**
         Link up words that mean the same - or nearly the same.
         outlet   dismiss
         loathe   promising
         revenue   hate
         auspicious   clever
         hectic   income
         canny   payment
         sack   busy
         fee   shop
         (Sørhus et al. 2006b)

Some of the words from the exercise above are quite unusual. Thus, teaching students their collocates can help them understand their meanings, as in the case of *auspicious start*. *Start* is

the second most common noun collocate of *auspicious* in the corpus (17 tokens), preceded only by its synonym *beginning* (27 tokens). The combinations *canny businessman*, *media revenues*, and *sales outlets* were found in the text and corpus but not in the dictionary.

Table 4.53 Collocations from exercise [28] from *Passage*

Words	Collocations found in the text	Type of collocation	Raw frequency	MI
auspicious	an auspicious start	Adj+N	17	6.77
fee	transfer fees - transfer fee - transfer fees	N+N	(28) 16 12	 4.73 4.03
hectic	hectic schedule - hectic schedules - hectic schedule	Adj+N	(104) 23 81	 10.51 10.14

Table 4.54 presents the ten most frequent bigram adjective collocates of the noun *outlet* in the corpus (Adj+N). Collocations such as *retail* and *fast food outlet*; *electric* and *electrical outlet*; *creative*, *social*, and *emotional outlet*; and *GFCI outlet* (a boat appliance) can show the different meanings of this noun.

Table 4.54 Collocates of the noun *outlet* in COCA

Collocates	Raw frequency	MI
retail	256	11.01
electrical	202	10.97
other	89	3.38
fast-food	59	11.49
creative	59	7.94
electric	23	6.73
social	16	3.14
emotional	15	5.59
major	15	3.51
GFCI	14	13.44

This was the last exercise from *Passage*. The total number of vocabulary items in the fourteen exercises selected from the website is 102; among these only one is a collocation (0.98%): *take place* from exercise [21]. The summary below will give an overview of the results.

4.5 Summary

This chapter has presented the results of the search in the textbooks, corpus, and collocation dictionary according to the criteria established in chapters 1 and 3. Comments have been added between the tables for matters of practicality and many of these will be discussed further in chapter 5.

Table 4.55 below presents a summary of the results for *@cross* and *Passage*. The table shows first the total number of exercises from each material, then the number of authentic,

constructed, and mixed exercises. Next, the total number of vocabulary items is listed; 156 from @*cross* and 102 from *Passage*, which comprise 258 items. In @*cross*, 93.6% of all the items constitute single words, while in *Passage* the percentage is 99.1%; the remaining items are word combinations. @*cross* presents 18 words in the exercises that are not in the texts, which comprise 12.3% of all the single words. This means that only 128 words from @*cross* (87.7% of all single words) were found in the texts. In *Passage* all the words have been found in the texts.

The table also displays the total number of collocations found, which is 55 in @*cross* and 36 in *Passage*. Since this number includes cases of one word with more than one collocation, each word that was part of at least one collocation was counted, giving a total of 45 in @*cross* and 35 in *Passage*. When calculating the percentage of these words among *all* single items (146 and 101), the result is 30.8% in @*cross* and 34.6% in *Passage*. However, in the case of @*cross*, when only the words found in the texts are considered (128), the result is 35.1%. Since the first research question is to find out the percentage of words that are part of collocations *in the texts*, only the second percentage (35.1%) will be taken into account.

Table 4.55 Summary of results from the two textbooks

		@*cross*		*Passage*		Both	
Type of text	total	14		14		28	
	authentic	-		7		7	
	constructed	14		4		18	
	mixed	-		3		3	
Items of vocabulary from the exercises	total	156		102		258	
	single words	146	93.6%	101	99.1%	247	95.7%
	combinations	10	6.4%	1	0.9%	11	4.3%
	single words not found in the texts	18	12.3%	-	-	18	7.3%
	single words found in the texts	128	87.7%	101	100%	229	92.7%
Collocations	total number of collocations	55		36		91	
	words that are part of collocations in the texts among all single words	45	30.8%	35	34.6%	80	32.4%
	words that are part of collocations among the ones *found* in the texts	45	35.1%	35	34.6%	80	34.9%

As often commented between the tables in the previous sections, many of the combinations found in the texts are included in COCA but not in the collocation dictionary (*OCD*), while others are included in the *OCD* but not in COCA. As explained above, these are not considered collocations in the present study. Figure 4.1 below shows that a total of 124 combinations from the texts were found in COCA (with frequencies above 3 and MI above 3.0); among these, only 91 were included in the *OCD*. Thus, 33 combinations have been excluded by using the *OCD*, 17 in @*cross* and 16 in *Passage*. On the other hand, 95

combinations from the texts were found in the dictionary, from which 91 were listed in COCA (with frequencies above 3 and MI above 3.0), which means that only 4 combinations from the dictionary have been excluded by using the corpus. Thus, it seems like the corpus has a wider range of collocations than the dictionary but that the dictionary focuses on the most common words, probably the ones which are most pedagogically relevant.
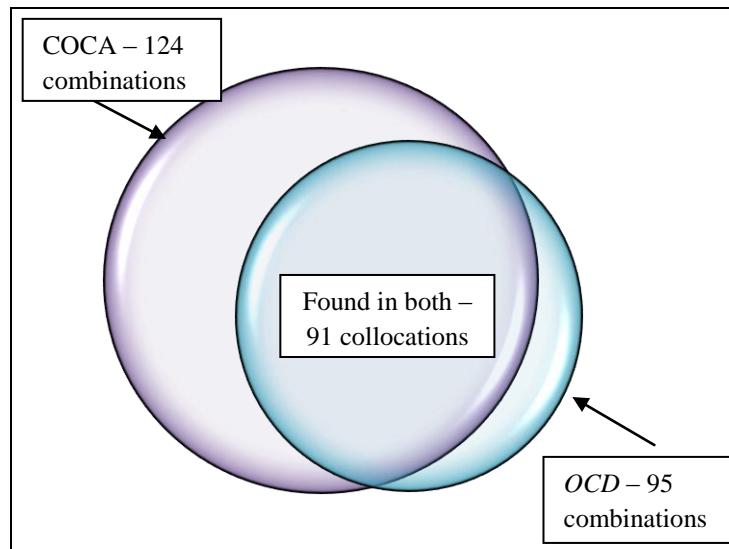

COCA – 124 combinations

Found in both – 91 collocations

OCD – 95 combinations

Figure 4.1 Number of combinations in COCA and the *OCD*

Chapter 5 will give an interpretation of the results, suggest ways in which the exercises can be improved and draw on research on collocations, corpora, and foreign language vocabulary acquisition to propose ideas for vocabulary teaching.

5. DISCUSSION AND SUGGESTIONS FOR TEACHING

5.1 Introduction

The aim of this chapter is to analyze the results reported in chapter 4, interpret them in light of the background in chapter 2, and discuss how these findings achieve the aims outlined in chapter 1. I will offer suggestions on how the exercises can be improved and on other ways in which collocations can be integrated into language teaching. These suggestions are not meant to be prescriptive, but to broaden teachers and textbook designers' awareness of collocations.

5.2 First aim: Research questions and hypothesis

In this section the main findings of the investigation will be reported and the first research question and the hypothesis will be addressed. Then, other interesting findings from the two materials will be discussed. At the same time, these findings will be considered in relation to the competence aims of the L06 in order to address the second research question.

The summary table in chapter 4 (4.55) shows that most of the items of vocabulary in both books are single words, which was expected since these exercises have been selected *because* they present lists of single words. The few items presented as combinations, 6.4% in *@cross* and 0.9% in *Passage*, were not part of the search in the texts and corpus. What is most interesting for the present study, however, is the percentage of words that are part of collocations in the texts. In *@cross,* 35.1% (45 out of 128) of the single words in the exercises appeared as collocations in the texts they are taken from, whereas the result for *Passage* is 34.6% (35 out of 101). These results enable us to achieve the first aim, which was to investigate if words that are usually taught as separate items in textbook exercises are part of collocations in the texts they are taken from. Addressing the first research question, these collocations comprise more than one third of all the words taught as single items in *@cross workbook* and the *Passage* website. Considering that the words from the exercises were not selected with collocations in mind, one third can be regarded as a high proportion. The results have thus corroborated the hypothesis that a large proportion of collocations would be found.

As explained in chapter 1, the corpus and dictionary have been used in this study as methods to exclude free combinations and combinations less pedagogically relevant. Figure 4.1 shows that COCA included more collocations than the *OCD*, which is probably due to the fact that the corpus contains a larger number of words than the dictionary. The latter only includes entries for 9.000 words and it prioritizes nouns (McIntosh et al. 2009:v-vi).

The next subsections will address the second research question – which of the collocations would be useful for Norwegian students in the first year of upper secondary school.

5.2.1 Discussion of *@cross* exercises

Some particular features stand out in the *@cross* material, two of which are that a number of words from different exercises are in fact not in the texts and that many collocations found were closely related to chapter themes. These and other findings will be discussed below.

Words not present in the texts

One of the first peculiarities of the *@cross* exercises is the fact that many of them practice words which are not present in the texts indicated by the headings. For example, in exercise [2] the heading is: 'If you need help, look at pg [*sic*] 41-44 in @cross Textbook. There you will find the words used' (Rodgers et al. 2006b:34), and in exercise [1], 'Having read the text on Berwick, find the English words corresponding to the following Norwegian words by drawing lines' (Rodgers et al. 2006b:28). However, not all the words appear in the text, which means that learners are expected to practice a completely decontextualized item. According to Gass (1988; in Nation 2001:159), activities matching first or second language synonyms, definitions or pictures (ten of the fourteen exercises in *@cross* are of this type), are exercises to test if students have comprehended the input received, in this case, the text they have read. If the word is not in the text, there is no input from which to learn.

The intention in some of the cases might have been to present to students other words related to the same topic as those that appeared in the text. An example would be exercise [5] which practices words related to media and government and adds the word *anarchy* that does not appear in the text. The same happens in exercise [9] about aid work, where the words *responsibility*, *health*, and *dedication* are added in the exercise without appearing in the text. Exercise [13], which deals with education, adds *electives* and *classes*. Nonetheless, if the intention of adding words in these exercises was to supply students with other items under the same theme, this might in practice have a negative effect. Hoey talks about his experience learning other languages using materials that worked with vocabulary in themed lists: 'I can only speak for myself as a very average learner, but I am conscious of quickly forgetting themed lists. Because the words are learned without reference to any context in which they might be used, they tend to get confused with each other' (Hoey 2000:227). Thus, although most of the words in the exercises are in a way already decontextualized for being presented

as single words, it is even less productive when they are not part of a text where students can check their meaning.

Repeated words

Another feature noticeable in @*cross* is that many words are repeated in different exercises, which, according to research on language acquisition, is essential for the retrieval of vocabulary (Schmitt & McCarthy 1997:276). Repetition is among the learning strategies suggested by Nation (2001:218), as displayed in table 2.1, and he also argues that meeting the word in new contexts aids in the process of generating language (2001:222). There are eight words that have been repeated in exactly the same form in @*cross* in contrast with only two in *Passage*. The reason why @*cross* seems to give more importance to recycling vocabulary might be that the exercises in the workbook are meant as the main exercises, while in *Passage* the website might be used as extra practice of what students have seen in the textbook.

Morgan Lewis attests that 'learning is holistic, cyclical and evolves over time' (2000:12), which means that recycling is essential to internalize the meaning of a word. He then suggests that teaching a word's collocates is a way of accelerating this process since it reactivates and expands the meaning of a word by showing how it combines with new ones. An illustration is the word *election*, found in the text for exercise [4] in the collocation *general election* and then in the text related to exercise [14] as *hold an election* and as *free election*. If the exercises had presented *election* in its collocations, students could have widened their knowledge of this word at the same time as recycling it. Thus, teaching the same word more than once and always providing some collocates can be a way of achieving the competence aim of L06 which says that students should 'master a wide vocabulary' (2006b).

Collocations with familiar words

Nation affirms that '[collocation exercises] can help learners expand their knowledge of words that are already familiar to them' and that training collocations can encourage students to widen the vocabulary they already know (1990:101). @*cross* contains some collocations that fit this category, such as ***medieval*** *art* [1], ***coastal*** *road* [2], ***incredibly*** *busy* [3], ***trendy*** *restaurants* [3], *drinking* ***water*** [9], *pure* ***water*** [9], and ***consider*** *necessary* [14]. The words in bold are the ones presented by the exercise (supposedly new), and the others are the collocates found in the text (words students probably know already). Teaching these two types of words in a collocation might help students remember the meaning of the new ones as they learn

more about the collocational behavior of the ones they already know. This is confirmed by Hill (2000:67) who argues that the main idea of vocabulary work should be to teach students the collocations of the words they are already partly familiar with (see section 2.5.3).

Collocations related to chapter themes

In contrast with *Passage*, each chapter of *@cross* is based on a theme, which is displayed in the table of contents of *@cross textbook* (Rodgers et al. 2006a:7-13). These themes can often be recognized by looking at the collocations that have been found in the texts. For example, exercise [5] is taken from chapter 8 where one of the themes is 'British parliament and politics' and some of the collocations found are: *general election*, *constitutional monarchy*, *Houses of Parliament*, *Members of Parliament*, and *elect representatives*. Another case is exercise [6] whose chapter has 'education in the UK' as a theme and some of the collocations found are: *comprehensive school*, *national curriculum*, *pursue an education*, and *vocational education*. In exercise [7] the theme of the chapter is 'higher education' and collocations like *study abroad*, *graduate student*, and *receive a scholarship* have been found. Another example is exercise [14], part of chapter 6, whose theme is 'The American political system'. The collocations found are: *pass/sign a bill*, *hold an election*, *free election*, *executive branch*, *judicial branch*, and *legislative branch.*

Some of the competence aims of L06 establish that students are expected to 'choose an interdisciplinary topic for in-depth studies within his or her own programme area and present this' and 'express himself/herself in writing and orally with subtleness, proper register, fluency, precision and coherence' (2006b). To make a presentation about interdisciplinary topics, students should first be able to select the appropriate vocabulary within these subjects. Learning collocations related for example to politics, government, and school systems in the US and the UK, as shown above, might help them complete their task with fluency and accuracy. The *OCD* states that 'a student who chooses the best collocation will express himself or herself much more clearly and be able to convey not just a general meaning, but something more precise' (McIntosh et al. 2009:v). Hence, learning collocations within subject topics can prove convenient for students.

5.2.2 Discussion of *Passage* exercises

Two contrasting features that can be observed between *Passage* and *@cross* are that in *Passage* all the words listed in the exercises are in fact in the texts and the units in *Passage* do not contain any specific themes like in *@cross*. Other features which will be discussed further

in this subsection are: infrequent and formal words, the issue of constructed versus authentic texts, the great number of grammatical collocations found, and the case of near synonyms.

Infrequent words

A common feature among many of the *Passage* exercises is the occurrence of infrequent and more formal words (e.g. in exercise [25] *retort*, in [23] *swoon* and *cleanse*, in [18] *skid* and *shin*, in [17] *enunciate*, in [15] *commence*, etc.) [10]. This might be due to the fact that many of the *Passage* exercises are based on literary texts, which does not happen in *@cross*. In any case, this can pose advantages and disadvantages for the students. One disadvantage is to be spending time learning words that they might not encounter again or have no need for. Morgan Lewis (2000:14) stresses that there would be a great improvement in students' vocabulary if less time was spent on learning rare words and instead they would concentrate in learning how to use the ones they already know in their collocations.

On the other hand, learning formal and rare words can be beneficial. According to L06, students should be able to 'read formal and informal texts in various genres and with different purposes' and 'write formal and informal texts with good structure and coherence on personal, interdisciplinary and social topics' (2006b). Thus, in order to understand and write formal texts, it is essential to have the necessary vocabulary. Nation notes that some advantages of teaching infrequent words are: that they might contain a prefix, root, or suffix that will help students when meeting other words; it may be a word that occurs a number of times in a specialized text which is of interest to the students; or still 'it may be a collocation of a much more useful word' (1990:137). We can see some examples of the last case in the *Passage* findings, such as *cause **consternation*** [16], ***clasped** my hands* [17], ***enunciate** every word* [17], *grow **weary*** [18], *took a **sip*** [19], ***sipping** tea* [19], ***new-fangled** ideas* [26], and *an **auspicious** start* [28]. The words in bold are the infrequent words brought up by the exercises and the others are the collocates found in the texts. The former could be argued to be, following Nation, collocations of other 'more useful' words, in this case, *cause*, *hands*, *word*, *grow*, *take*, *tea*, *ideas*, and *start*. In *CALD* all of these are marked as 'essential' words, which they define as 'the most common and useful words in English' (2005:i).

Thus, one way or another, if infrequent or more formal words ought to be learned, it is probably more effective when they are presented with one or more collocates.

---

[10] In order to determine if the words mentioned in this subsection are frequent or infrequent I have looked them up in *CALD* which contains a 'Frequency Information system' (see *CALD* 2005:vii).

Authentic versus constructed texts

The issue of authentic versus constructed texts is particularly relevant in relation to some examples found in *Passage*. As it can be noted from table 4.55, all the exercises selected in *@cross* present words based only on constructed texts while *Passage* includes seven authentic, four constructed, and three mixed. Whether this has had an influence on the results concerning number of collocations is unclear since the proportion of collocates was approximately the same in both materials. On the other hand, the combinations that appear in constructed and mixed texts prompt some questions in relation to the type of input the books provide for the learners.

A case in point is *promise allegiance* commented on in chapter 4, table 4.50. This combination was found in a constructed text about English monarchs, but it is not listed in the collocation dictionary and, in the corpus, it presents a raw frequency of 3 and an MI of 2.40 score (which is why it is not listed in the collocation table). If the corpus is searched to find more about the uses of *allegiance*, *pledge* appears as the most frequent verb collocate, and the collocation *pledge allegiance* shows a frequency of 529 and an MI of 12.52. Thus, the question is: why would the textbook writers include such an infrequent and weak combination (*promise allegiance*) instead of the stronger and more common collocation (*pledge allegiance*)? The reason might be that *allegiance* is a new word for students and the writers attempted to combine it with a word that students probably already knew (*promise*). Be that as it may, *pledge allegiance* is what students will most likely encounter, as in the American oath of loyalty to the republic: 'I pledge allegiance to the flag of the United States of America [...]', which is used in schools, government meetings, and even in sports events. These findings could indicate that constructed texts tend to include weaker combinations than authentic ones.

However, another example should be considered before suggesting any tendencies. In the text related to exercise [24], the infrequent and weak combination *color prejudice* was used instead of the stronger and more common *racial prejudice*. The former was included in the dictionary and in the corpus, but its raw frequency was 8 and MI score 3.04; *racial prejudice*, on the other hand, presented a frequency of 202 and an MI of 10.02 (see table 4.46). Then the question arises again: Why was *color prejudice* used instead of *racial prejudice*? Unlike the case of exercise [26], this exercise is based on a mixed text so that parts of autobiographies are combined with the constructed text and *color prejudice* was used by Ghandi when talking about his experiences in America, which is the authentic part of the text. It could still be argued that Ghandi was not a native speaker of English. Nevertheless, the

main point here is that there may be no assurance that authentic texts have stronger and more common collocates than constructed ones. These, however, are very preliminary results and should be investigated further to draw any definite conclusions.

Grammatical collocations

Another particular feature of the *Passage* results is the large number of grammatical collocations found in the texts. Calculating the proportion of this type of collocation in both materials, I found that 5.6% in @*cross* are grammatical collocations in contrast to 34.3% in *Passage*. The ones from *Passage* are (exercise number indicated in brackets): *detached from* [15], *yell at* [17], *glanced at* [19], *pushed through* [19], *adapt to* [20], *emerge from* [20], *apologize for* [22], *snatched from* [22], *negotiations with* [24], *resentment towards* (which appeared twice) [24] and [25], and *persist in* [25]. At first glance, it might not be clear why *Passage* contains a larger proportion of grammatical collocations. However, it is observable from the examples that most of these collocations are of the type V+P; thus, having a suspicion that the presence of verbs in the exercises might have influenced the results, I counted all the verbs appearing in the exercises from both materials and found that 10.5% of the words in @*cross* are verbs while in *Passage* the proportion is 44.5%. Hence, it seems like verbs tend to be in grammatical collocations more often than nouns and adjectives.

Grammatical collocations might not be so relevant to teach when they are congruent to Norwegian, as for example *adapt to* (*tilpasse til*) and *snatched from* (*tatt fra*). On the other hand, those which show a different grammatical structure should probably be taught, such as *apologize (to someone) for (something/doing something)*, (*beklage seg for noe*, literally 'apologize oneself for something') or *persist in* (*fortsette med*, 'continue with'). Emphasizing collocations which are not obvious to students and highlighting the ones that might be structurally different from the L1 can, as a matter of fact, be recommended regarding any collocation and not only grammatical ones.

A note of caution is pertinent here relating to the percentages of verbs provided above. The number of verbs given is only approximate since the distinction between verbs and nouns are not always straightforward in the exercises, even when checking the synonyms given. An example is exercise [23], where *aid* is the keyword and the synonym given is *help*; another is [26] where the English word *descent* should be linked to the Norwegian translation *stamme*, which can be a verb or a noun. In fact, some of the exercises, especially the two mentioned, seem confusing due to the lack of context or clues. In exercise [23], the words are infrequent and the synonyms do not give any indication of the part of speech they belong to. Moreover,

in one case the keyword and its synonym are written in different forms: *fancy* (supposedly a verb) is to be matched with *imagining*. One of the competence aims of L06 is that students should be able to 'use relevant and precise terminology to describe the forms and structures of English' (2006b). Therefore, it would probably be advantageous to include at least the part of speech of the words.

A final comment on grammatical collocations is that their frequency tends to be higher than the lexical type (see collocation tables from *Passage*), which is owing to the fact that prepositions are a generally frequent class in the corpus.

Collocations of near synonyms

The great majority of the exercises selected in the *Passage* website are of the matching synonyms type. While this kind of exercise can help students relate the new words with known ones, the way they have been designed might pose some challenges to students. The reason is that, different from @*cross*, which gives definitions, explanations, or collocations as matching options, *Passage* provides mostly single words. This type of exercise might lead to the same misconception as the translation ones (these are discussed in subsection 5.2.3), i.e. that students may think that there is a one-to-one correspondence between synonyms. What is clear from the exercises, however, is that the synonymous words provided have very distinct uses and that corpus information can reveal their different meanings, as previous studies on near-synonyms have found (see Kennedy 1991 and Partington 1998).

An example is the verb *assassinate* [25] that should be matched with *murder*. When looking at table 4.48, which lists the ten most common collocates of *assassinate*, it becomes clear that this word is used with 'important people', as shown by the collocates *president*, *king*, *minister*, and *Pope*. Indeed, as mentioned in chapter 4, the definition of *assassinate* given by *CALD* is 'kill someone important or famous' (2005:66). To attest the differences between these two verbs, the ten most frequent noun collocates of *murder* were checked in the corpus and the following nouns have been found (in order of frequency): *wife*, *husband*, *father*, *daughter*, *parents*, *son*, *thousands*, *brother*, *civilians*, and *sister*. These results contrast with the ones from table 4.48 in that only 'common people' are listed, many of which are family members. Hence, an important distinction between these two verbs is hidden by the simple 'match the synonyms' association required by the exercise. Although *assassinate the president* or *murder his wife* are not considered collocations by this study, showing students the collocates of the synonyms is certainly beneficial.

Another example is the word *notable,* which is to be matched with the synonym *important* in exercise [20]. The collocate table of *notable*, 4.39, can be compared to the table of collocates of *important* (5.1) shown below:

Table 5.1 Collocates of the adjective *important* in COCA

Collocates	Raw frequency	MI
thing	4219	6.01
role	2563	6.58
part	2431	5.11
things	1268	4.01
factor	1105	6.88
point	1073	4.28
issue	1031	5.04
issues	960	5.16
question	776	4.10
aspect	682	7.29

Comparing both tables it becomes evident that none of the ten most used collocates of these words is the same. In addition, the frequencies with the adjective *important* are much higher while the MI scores are lower, probably for being a more frequent word than *notable* in the corpus. The MI scores and frequencies displayed by table 4.39 indicate that the collocation *notable exception* found in the text is highly relevant (most frequent and strongest in the corpus) and could have been taught as a chunk in the exercise. It should be noted that in case the words were to be taught in collocations, the model of the exercise would probably have to be altered, which will be discussed further in section 5.3 below. The difference between these two words could also be debated in terms of register or genre. *A notable example*, for instance, could be taught as a more formal or more academic variation of *an important point.* This could be brought up in class in relation to spoken and written language. One of the aims specified in L06 is that students should 'select appropriate listening, speaking, reading and writing strategies adapted to the purpose, situation and genre' (2006b).

Other examples related to the issue of genre and register are *provide* and *give* in exercise [21], where *give* is presented as the synonym of *provide*. Table 4.41 shows that the collocates of *provide* seem neutral, such as *information, support*, *care*, and *assistance*. On the other hand, when looking at the most common noun collocates of *get* and checking their concordance lines in the corpus, we find informal expressions like *get a chance*, *get in trouble*, *get the impression*, *get a grip*, and *get your ass out of here*.

Frank Palmer (1976) proposes five types of synonymous relations: register relation (e.g. *nasty smell* vs. *obnoxious effluvium*), collocation relation (e.g. *rancid butter* vs. *addled eggs*), connotation relation (e.g. *thrifty* vs. *stingy*), dialectal relation (e.g. *autumn* vs. *fall*), and

the category of partial synonyms (e.g. *promise* vs. *pledge*). The ones that are present in the *Passage* exercises seem to fit in the categories of register (*commence* vs. *begin* [15], *answer* vs. *retort* [25], *loathe* vs. *hate* [28]), collocation (*cast* vs. *throw* [18], *push* vs. *shove* [19], *notable* vs. *important* [20]), and partial synonyms relations (*sip* vs. *drink* [19], *tribe* vs. *nation* [20], *assassinate* vs. *murder* [25]). Pointing out these relations to students or asking them to discuss what kinds of relations the synonyms in the exercise bear can help them more fully understand the meanings of the words.

My contention here is that teaching the collocates of a word can help students differentiate the meanings of near synonyms. As Carter affirms, 'we should […] note that synonymic relations between words can be usefully distinguished with reference to the different collocational ranges of the synonyms involved' (1987:53-54).

5.2.3 Discussion of common features

Although @*cross* and *Passage* presented individual peculiarities, there are central aspects which are common to both materials. The issues discussed in this subsection are connected to translation exercises, unexpectedly uncollocated words, collocations which convey sociocultural aspects, and collocations which evoke communicative situations.

Translation exercises

Among the twenty-eight exercises selected, one fourth is translation exercises: six in @*cross* and one in *Passage*. Although making use of the L1 has its advantages when learning another language (when for example drawing parallels or for true cognates), it might also be problematic, as discussed in chapter 2 (section 2.3). Translation exercises of the type 'match the words with the correct translation' used in these two materials might give students the false idea that there is a one-to-one correspondence between words in the two languages, especially when only one translation is given for the English words, as in exercises [1], [6], and [7] in @*cross* and [26] in *Passage*.

Using translation in class should not be strictly banned as it was in the Direct Method, but a more effective way to do it would be to translate words as chunks into the students' L1, particularly when a term is not congruent in the languages. Nesselhauf (2005:260) concludes in her study of collocations used by German learners that more emphasis should be given to non-congruent than to congruent collocations in teaching. This seems to be the case of the collocations: *general election* [5], *pursue an education* [6], *comprehensive school* [6], *graduate student* [7], *different opinions* [7], *consider something necessary* [14], and *pass a*

law or *a bill* [14]. These would probably be translated respectively as *stortingsvalg* ('parliament election'), *ta en utdanning* ('take an education'), *grunnskole* ('base school'), *akademiker/student* ('scholar', 'student'), *ulike meninger* ('different opinions'), *se noe som nødvendig* ('see something as necessary'), and *opprette en ny lov* ('establish a new law'). These collocations might seem simple in English but their translations are not literal and many of them illustrate Swan's (1997) list of aspects that might interfere when translating from one language into another (see section 2.3). For instance, the case of *graduate student* would be an example of the different notions of a 'word', since university or college students are called simply *student* in Norwegian (the English word *student* is *elev* in Norwegian*)*. So in English the concept is realized in two words while in Norwegian it is realized in only one. Another aspect that might generate confusion is false cognates, which is the case of *different opinions*. This collocation seems fairly obvious in English and some might even question if it is indeed a collocation. However, the translation to Norwegian is *mening*, which is a false cognate, and students tend to utter sentences like *my meaning is that...* ('My opinion is that...'). Swan (1997:159) also notes that words might have different grammatical contexts, which is the case of *consider necessary*. The Norwegian version of this collocation needs the preposition *som* ('as'), whereas in English the collocation has no preposition. Morgan Lewis says that only after an item is shown to students in its bigger context is it safe to translate it into the learners' L1, and 'not word-for-word but whole phrase to whole phrase, bearing in mind that the structure of the expression may be very different in one language from the equivalent expression in the other' (2000:16). In the case of the collocations *put up fierce resistance* in exercise [26], the literal translation in Norwegian would be 'make hard resistance' (*gjøre hard motstand*), which is not a collocation in English.

However, there are examples which are probably obvious for Norwegian students for being congruent and for having one of the collocates as a true cognate, such as *hold an election* (*holde valg*) [14], *direct descent* (*direkte nedstamming*) [26], and *mark the boundary* (*markere grensen*) [1]. In such cases the L1 can be helpful. On the other hand, Hill et al. (2000:94-97) report about a study on teaching collocations to business English students that concludes that learners profited from positive feedback on their right use of collocations, which enabled them to fully internalize what was only half-known. Thus, transparent collocations can probably be omitted from exercises so that more unpredictable ones can be given priority. However, when students use them correctly in their writing, a positive feedback can help them internalize what might have been just a successful guess.

Unexpectedly uncollocated words

As seen in the summary table in chapter 4 (table 4.55), more than one third of the single words presented in the exercises were part of collocations in the texts. Although my claim is that many of these single items should have been presented in collocations instead, I do not intend to argue that absolutely all need to be shown with a collocate. Cases like *broadsheet* [5], *anarchy* [5], *happy* [10], *evidently* [15], *forehead* [17], *orphanage* [22], *humiliate* [22], and *bewildered* [23], among others, are high information words (see Woolard 2000:33) and not usually part of strong and frequent collocations[11]. It would be awkward for instance to present *happy* with *birthday*, which is shown as the strongest collocate of *happy* in the corpus; or present *Russian orphanage,* where *Russian* is the strongest and most frequent adjective collocate of the noun *orphanage*.

On the other hand, there are words which occur repeatedly as part of strong and frequent collocations but that for some reason are presented in the exercises as single items. This is the case of *environmentally* in exercise [8], which quite unexpectedly appears without any collocates. By looking at the collocate table (4.15), it is apparent that the weakest combination, *environmentally aware*, is still strong (MI 7.86) compared to collocates in other tables. In fact, after the word *allegiance*, this is the word with the strongest collocates among all the twenty-eight investigated (see collocate tables). The word *environmentally* appears 3052 times in the corpus; in 545 of the cases it appears with *friendly*, 316 with *responsible*, and 286 with *sound*, which means that more than one third of its occurrences are with only these three collocates. Thus, I argue that there is sufficient evidence that this word should have been taught as a collocation and not as a single item.

Another example that should be given attention is the word *new-fangled* in exercise [26] of *Passage*. This word only appears 54 times in the corpus and in all of them followed by a noun. Among these, it appears 6 times with the collocate *idea(s)*, 2 with the word *glass* and only once with all the other nouns. Taking into consideration that in the text where this word is taken from the collocation *new-fangled idea* was found, it is surprising that the adjective is presented as an individual item in the exercise. Michael Lewis maintains that 'collocations are not words which we, in some sense, "put together", they co-occur naturally, and the first task of the language teacher is to ensure that they are not unnecessarily taken apart in the classroom' (2000:132).

---

[11] As mentioned above, all of the words from the exercises were entered in the corpus to find their ten most common collocates, so the assertion here that they are not part of strong and frequent collocations is based on corpus information.

A final interesting case is of *auspicious start* in exercise [28]. As commented in chapter 4, *auspicious start* and *auspicious beginning* are the most frequent Adj+N collocations in the corpus and the second and third strongest. Taking into consideration that the meaning of these noun collocates is very close, that this adjective is uncommon, and the fact that *auspicious* appears together with *start* in the text, my claim is that they should have been presented as a collocation in the exercise.

Collocations evoking communicative situations

Hill et al. (2000:101) and Morgan Lewis (2000:15) emphasize that the communicative potential of many collocations can be explored in the classroom. As Morgan Lewis affirms:

> Many collocations have immediate pragmatic force or are situationally evocative. For example, it is hard to think in which situation someone might say: *This is a corner.* But if I say to you: *This is a dangerous corner*, it immediately suggests two people in a car as they approach a corner where lots of accidents have happened. (Morgan Lewis 2000:15)

Morgan Lewis gives other examples of this kind, like *routine check-up*, *disperse the crowd*, and *widely available* (see section 2.5.3), which can suggest, respectively, a visit to the doctor, a police action after an incident, and talking about a new product. He argues that these real-world situations must be brought to students' attention in class (Morgan Lewis 2000: ibid.). Similarly, students could be asked in what kind of situation they would use or see the following collocations found in the texts: *incredibly busy* [4], *feel frustrated* [10], *delicious meal* [12], *have a reputation* [6], and *hectic schedule* [28]. Some ideas could be to use *incredibly busy*, *feel frustrated*, and *hectic schedule* to talk about a busy lifestyle; a *delicious meal* when complimenting someone on the food or thanking for the food; and *have a reputation* when commenting on someone's behavior (e.g. *He has a reputation for being rude*). These are only suggestions and students could use their creativity to imagine other real life situations. They could also be asked to write a dialogue using these collocations and then practice it with a partner. This kind of activity could be an aid to accomplish L06's communicative aim of 'take initiatives to start, finish and keep a conversation going' (2006b). The central point is that, as research in vocabulary acquisition confirms (see section 2.3), visualizing examples of the word and 'creating contexts, collocations and sentences containing the word' (Nation 2001:222) can foster retention and production of new vocabulary. More ideas on how collocations can be integrated in teaching will be proposed in section 5.4 below.

Collocations to talk about news and current events

L06 determines that students should be able to 'present and discuss international news and current events' (2006b) and some of the collocations found in the texts could be helpful to accomplish this aim. When preparing for a news presentation, students could use *high unemployment* [4] to talk about the job market; *cause consternation* [16] to talk about some polemic event or issue; or *put up fierce resistance* [26] to talk about a rebellion against the government, for example. Certain collocations are very restricted to the language of news and advertizing and they should also be given some attention in class, such as *hit the headlines* and *innovative features* respectively (McCarthy & O'Dell 2002:124-126).

Collocations revealing various meanings of a word

Some of the collocate tables with words from the exercises reveal that different collocates can show a word's various meanings. For instance, the collocates of *bright* (*starts*, *red*, *future*, etc.) listed in table 4.5 show that this word has the meanings 'full of light', 'strong in color', and 'full of hope for success or happiness', besides the 'intelligent' meaning like in *bright student* (*CALD* 2005:152). Thus, teaching students some of the collocational fields of *bright* ('things that shine', 'colors', 'people', etc.) can help them use this word.

Another example is seen in table 4.29, where the verb collocates of *slack* show the different meanings of this noun. The concordance lines of the collocates show that this word is used in expressions such as *pick/take up the slack*, *cut him/her some slack*, *his/her face/jaw/fingers…went slack*, and *dressed in black/dark/pressed…slacks*. Another example is the collocates of *snap* from exercise [19] shown in table 4.37: *snap your fingers*, *snap pictures/photos*, *snap peas* and *beans* (which are compounds), and *snap to attention*. All these examples are parallel to Hunston's list of collocates of the word *shed* (2002:12; see section 2.5.4) and to her argument that collocational information can bring forward the different meanings of a word (2002:76). Hence, teaching the collocates of a word and their collocational fields can aid students in using the words they are taught.

Collocations conveying sociocultural aspects

Maybe one of the most interesting findings in this study has been how collocations can reflect sociocultural aspects of the American society. In this study I use the term 'culture' in a broad sense to mean behaviors, ways of living and thinking, traditions, historical facts, and current events in a society. Investigating how these aspects are reflected in the language is a vast field

of research. However, as discussed in chapter 2, not so much has been done connected to how collocations in English can reflect the culture of its native speakers.

A first example is the word *suburbs* from @cross. Some of its collocates listed in table 4.4 are especially interesting: *white*, *affluent*, *wealthy*, and *middle-class*. These reveal that people living in the suburbs are probably white, rich, and middle-class. The collocations conveying economic status are the strongest ones – *affluent* with an MI of 10.40, *middle-class* with 8.72, and *wealthy* with 8.23. If we look for the collocates of *neighborhood* for example, we find among the ten most common ones *black*, *working-class*, *poor*, and *tough*. This kind of information provided by the corpus can foment a discussion about the class system in the US and how the different classes are spread around a city. Another example is *coverage* shown in table 4.9 in exercise [5]. The collocates of this noun indicate that this word is commonly used to talk about health insurance, like *full coverage* and *medical coverage*. These collocations can incite a discussion about the health system in the US.

Another thought-provoking example is the collocates of the word *prom*, displayed in table 4.19 of exercise [10]. Firstly, the prom tradition might be interesting to discuss since Norwegian students have nothing of the type, so collocates like *senior*, *junior*, and *high school* could be useful in this context. Secondly, what might be striking to Norwegian students are the collocates *white*, *integrated*, and *separate*, which show that in many places in the USA blacks and whites still have separate proms. Looking at the concordance lines of these three words, the following sentences are found: 'Saltzman's fine documentary "Prom Night in Mississippi" tells about the first integrated prom in the history of Charleston high school - in 2008'; 'White students plan the white prom, and black students plan their own'; and '[…] a black boy tried to get into the white prom to see his friend, and they told him he had to leave, and they had him escorted out' (Davies 2008). These would certainly be relevant to bring to students' attention in the context of school and racial issues in the US.

Another example worth mentioning is the collocates of the adjective *attractive* in table 4.23, exercise [12]. What these collocates indicate is that women are referred to as *attractive* much more than men: *woman* has a raw frequency of 260, *women* 85, and *girl* 49, totaling 349 hits, against 89 applying to *men*. This aspect is probably not only related to the American society but to the general belief that women's appearance is more important than men's. This could be compared for example with the collocates of *smart*, which show *guy* as the second most frequent collocate while the first word designating a female person (*girl*) only appears in the fourteenth place.

A final example is the high frequency of the collocate *center* with the node *trade* in table 4.17, exercise [9]. This is an indication of how events, like the attack on the World Trade Center, can influence the language. As commented in chapter 2, it would be interesting to investigate in the corpus the concepts that were often used after the attack, such as *war on terror*, *anti-terror campaign*, *terrorist attacks*, *terrorist networks*, etc. (see section 2.5.5). Due to space constraints I have looked only at the most frequent noun collocates of the word *terrorist* and found the following collocates (raw frequency and MI score indicated respectively in parenthesis): *attacks* (2207, 11.59), *groups* (755, 8.18), *organizations* (402, 8.85), and *threat* (267, 8.29). Also interesting is that the raw frequency of the word *terrorist* from the year 2000 to 2002 has risen from 243 to 2001 tokens.

In sum, it seems like the corpus is a good asset for teachers when discussing sociocultural differences between English speaking countries and Norway. However, since this corpus is of American English, most of the information reveals features related to the US. Using the British National Corpus (BNC) would probably be a better option when discussing characteristics of the British culture. For teachers wanting to use the BNC, it is available online with the same interface as COCA, so all the corpus tools explained in the present study can be employed to search the BNC. It should be noted that knowledge of culture and society is central in L06, as some of the competence aims reveal: 'the pupil shall be able to discuss social conditions and values in various cultures in a number of English-speaking countries [and] present and discuss international news and current events'. Also under the 'main subject areas' it is explained that 'working with various types of texts and other cultural expressions is important for developing linguistic skills and understanding how others live, and their cultures and views on life' (L06 2006b). Calling attention to culturally-loaded collocations in these texts would be a way of achieving some of the objectives of the Norwegian curriculum.

5.2.4 A summary: What kinds of collocations are useful for students?

This section summarizes what types of collocation might be useful to teach Norwegian students in the first year of upper secondary school. The suggestions are based on the findings described above, research on collocations and foreign language vocabulary acquisition, and on the competence aims of L06. Thus, answering the second research question, the types of collocations found in the books which would be useful for Norwegian students are displayed in table 5.2 below:

Table 5.2 Useful collocations for Norwegian students among the ones found in the texts

Types of collocation	Examples
with words that appear repeatedly in the exercises	*general election*, *free election*, *hold an election*
with words that are already familiar to students but which expand the knowledge of the known word	*incredibly busy*, *trendy restaurants*
related to chapter themes	*constitutional monarchy* (government in the UK), *vocational education* (education)
when teaching infrequent words, those containing the infrequent word and a known word	*grow weary* and *auspicious start*
frequent and strong in corpora	*pledge allegiance*, *environmentally friendly*
collocations of near synonyms to avoid using them interchangeably independent of context	*notable example* and *important point*
non-congruent with the Norwegian equivalent	*put up fierce resistance*, *graduate student*
evoking communicative situations	*hectic schedule*, *feel frustrated*
useful to talk about news and current events	*high unemployment*, *cause consternation*
revealing various meanings of a word	*bright future* and *bright light*, *cast a vote* and *cast doubt*
conveying sociocultural aspects (These do not need to be directly taught, but brought to students' attention in order to foment discussions about cultural issues.)	*affluent suburbs*, *white prom*

Thus, it seems like most of the collocations found in the textbooks would be useful to teach Norwegian students and, most importantly, that there is an even greater variety yet to be explored in class. My claim, however, is not that traditional vocabulary exercises should be merely turned into collocation exercises. Although they can be improved by collocational information, as will be discussed in the next section, I argue for a change in the way we *view* language. As Hill observes, language is 'a predominantly lexical phenomenon' (2000:47), and this view of language should be reflected in foreign language teaching and learning. This does not mean that we should teach students all the collocations of English, but that awareness of collocations should be encouraged as a learning strategy. Woolard maintains that it is 'essential […] that the teacher equips the students with search skills which will enable them to discover significant collocations for themselves, in both the language they meet in the classroom and, more importantly, in the language they meet outside the classroom' (2000:34). This statement is particularly relevant for the context of Norwegian upper secondary students who probably meet English words more often outside than inside the classroom.

5.3 Second aim: How the exercises can be improved

It is paramount that research on corpus linguistics and particularly on collocations finds its way into the classroom. The second aim of the present study is to suggest how the vocabulary exercises can be improved and complemented by previous research and ideas from existing collocation exercises. As mentioned in section 2.3, the exercises selected from @*cross* and

Passage are the type used to consolidate word form and meaning in memory, which is the fourth step in the vocabulary acquisition process (Hatch & Brown 1995). This section will give suggestions on how the vocabulary exercises discussed above can be improved using corpus information so that they will still work at consolidating form and meaning at the same time as they will take students a step further in the learning process: how to *use* the words. The suggestions are also based on previous research and teaching materials on collocations.

Woolard affirms that 'students with limited time available for study will not learn high priority lexis if it is not deliberately selected and incorporated into learning materials. Collocations, then, must become part of that planned language input' (2000:32). One way to turn collocations into part of the input is including more common and stronger collocations in the texts constructed for the textbooks. As discussed above, the collocation *pledge allegiance* could have been used in the text instead of *promise allegiance*.

Relevant to how the exercises can be improved, it is generally accepted that students learn best when words are practiced in context (Paribakht & Wesche 1993); what I would like to add is that teaching the collocates of a word is also a form of contextualization and should not be regarded as less important. On these grounds, and inspired by the example of McCarthy et al. (2006a:48; see appendix 2) exercise [6] in @*cross* could be redesigned as shown by figure 5.1:

[**6**] Find these words and expressions in the text. Match them with the correct translation.	
1. pursue an education	7. academically giftet
2. vocational education	8. comprehensive school
3. national curriculum	9. compile (v)
4. enhance (v)	10. Beacon school
5. evolve (v)	11. coaching (n)
6. build up a reputation	

Figure 5.1 Suggestion of how to improve exercise [6]

Using translation in this exercise is not a problem as long as it is a translation of the whole expression or collocation (see Morgan Lewis 2000:16, section 2.3), but another option would be to list definitions in English. What has been changed here is that the collocations found in the text were added, the combinations that were there from before were kept, and the single words which did not have collocates were given parts of speech.

Another suggestion is to turn a simple translation exercise like [14] into a collocation exercise where the collocates are to be matched (if in doubt students can check them in the text) and then translated to Norwegian. This kind of exercise practices collocations, uses the

context of the text to help students guess meanings, and makes them think about the concepts as a unit in Norwegian. The suggestion is presented in figure 5.2 below:

[14] Match the words on the left with their collocates on the right (some of the collocates can be used with more than one word). Then write them down as in the text and then translate them into Norwegian.
1. pass   a. election
2. consider   b. branch
3. hold   c. law
4. executive   d. necessary
5. legislative   e. bill
6. free
7. judicial
8. sign

Figure 5.2 Suggestion on how to improve exercise [14]

In the exercise in figure 5.2, only the collocations found in the text were used; however, this exercise would be more challenging if other words and collocations related to government were also added.

In exercises like [19] where some of the key words have different meanings determined by their collocates (e.g. *snap*, *push*, and *tug*), these collocates can be presented and students are supposed to find the words in the text, as shown in figure 5.3 below. As a follow up, students can discuss in pairs the different meanings conveyed by the collocates. This exercise is inspired by one in Michael Lewis (2000:112; example 2 in appendix 2).

[19] Try to find in the text the verbs that complete the following collocations. The same verb completes all three examples. Thereafter, discuss with a partner the different meanings of each word.

1. …………… pictures   2. …………..… the button   3. ……………… of war

  …………….. to attention   ……………. up princes   ……………… at somebody's heart

  …………..…… your fingers   ……………. the limits   an emotional ………….

Figure 5.3 Suggestion on how to improve exercise [19]

Another suggestion is one related to exercise [28] in *Passage*. Three words from this exercise were part of collocations in the text (*hectic schedule*, *auspicious start*, and *transfer fees*) and the word *outlet* was checked in the corpus for its most frequent collocates (see table 4.54). One way to adapt this exercise to include collocations could be to group their collocates and ask students to match the key word with the group of collocates, as in an exercise made by Michael Lewis (2000:115; see example 3 in appendix 2). In case students need to check the meaning of the words first, they can look at the glossary provided next to the text in *Passage* (Sørhus et al. 2006a:311-314). Checking the meaning would be part of the third step in the vocabulary acquisition process according to Hatch & Brown (1995, section 2.3) and the

collocation exercise suggested in figure 5.4 would aid in achieving the fourth step, which is 'consolidating form and meaning to memory'. In order to reach the fifth step, 'using the words', students could be asked to write sentences or short dialogues including the collocations. The suggestion is shown in figure 5.4 below:

[28]   Match the following words with their group of collocates. One of the words does not match any groups.
auspicious   canny   sack   fee   hectic   revenue   loathe

  a. entry, registration, membership, admission, school
  b. beginning, start, opportunity, event, day
  c. tax, oil, advertising, government, export, sales
  d. schedule, day, pace, life, work, lifestyle
  e. businessman, politician, marketing, move, eyes, smile

Figure 5.4 Suggestion on how to improve exercise [28]

The collocates listed in figure 5.4 above were all taken from the corpus. They were not checked in the dictionary since many of the node words are not included in *OCD*.

Some of the vocabulary exercises have a potential to help students internalize new collocations and discuss cultural differences. Nesselhauf says that 'the topic of an essay could serve as a trigger for teaching a number of collocations […] that occur frequently in connection with this topic' (2005:265). She also recommends that the collocations are introduced before students write the text. Hence, exercise [10], in which students have to write a text about American teenagers, could be supplemented so that certain collocations are discussed first in relation to the American teenage culture and then students can use them to write a text about the topic, as illustrated in figure 5.5.

[10] The following collocations have been found in an American corpus. What do they tell you about teenagers in the US and the American culture?

  a. sports hero   d. shopping lists
  b. senior prom, white prom   e. fashion trends, latest trends
  c. feel frustrated   f. bullying prevention

  Now write a paragraph describing American teenagers and their situation using the collocations above and the following words:
   weird, wired, committed, happy, cliques
  Add other words from the texts which you think fit in.

Figure 5.5 Suggestion on how to improve exercise [10]

In the text related to exercise [10], the following combinations were also found: *wired teenagers*, *member of a clique*, and *clique identity*. They could also be discussed since they convey the close relationship of the new generation with technology and the importance of

belonging to a clique in American high schools. Something similar to the suggestion in figure 5.5 could be done in exercise [9] so that students would have to use collocations like *medical aid*, *cancel de debt*, *drinking water*, etc. (see table 4.16) to write the text about aid work.

Exercise [25] asks students to match synonyms. The heading is 'Join up the words that mean the same – or nearly the same' (Sørhus et al. 2006b) and some of the synonyms are: *murmur* and *mutter*, *persist* and *continue*, *select* and *choose*, *assassinate* and *murder*, *retort* and *answer*. This exercise could be complemented so that after students match the synonyms, they could try to find out the differences between them. Nesselhauf recommends that in more advanced levels the simple matching synonyms activities should 'be replaced by exercises actually focusing on the differences in meaning and usage of near-synonyms' (2005:266). Whether or not students at the first year of upper secondary school in Norway are advanced is disputed. What is certain is that changing the focus of the exercises would add some challenge to the activity and give students a deeper understanding of the uses of the words they are learning. Since this exercise is on the website, one way to redesign it would be to hyperlink the words so that students could click on them and see some information taken from corpora to be analyzed. They could discuss in pairs and draw conclusions about the various uses of the two words. Hatch & Brown show that this type of activity is possible using programs like Hypercard or Toolbook. They explain that by clicking or touching a word, grammatical or cultural information, a translation, or the meaning of the word is provided (Hatch & Brown 1995:408). It should also be possible to include corpus data so that information as shown in figures 5.6, 5.7, and 5.8, could pop up when students clicked on the words. The heading of the exercise could be: 'Look at the corpus information and discuss the differences between the following synonyms'. Figure 5.6 below show the most common collocates of the words *assassinate* and *murder*.

assassinate				murder		
1	☐	PRESIDENT	129	1	☐	WIFE	214
2	☐	LEADERS	32	2	☐	HUSBAND	82
3	☐	KING	22	3	☐	FATHER	71
4	☐	MINISTER	15	4	☐	DAUGHTER	57
5	☐	LEADER	15	5	☐	PARENTS	51
6	☐	OFFICIALS	12	6	☐	SON	47
7	☐	POPE	9	7	☐	THOUSANDS	36
8	☐	CHARACTER	9	8	☐	BROTHER	34
9	☐	COMMANDER	6	9	☐	CIVILIANS	21
10	☐	MILITANT	5	10	☐	SISTER	21

Figure 5.6 Lists of collocates of the synonyms *assassinate* and *murder* in COCA

Based on this information students will probably conclude that the difference between *assassinate* and *murder* is mainly the type of people killed and the motivations for killing.
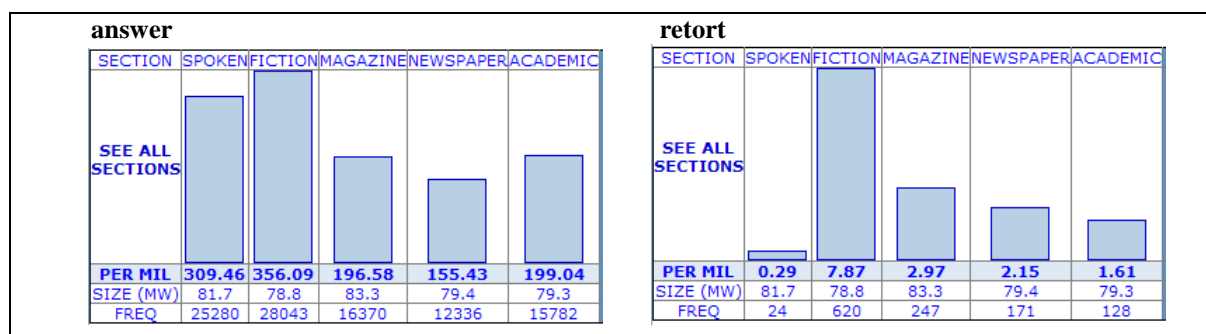

Figure 5.7 Use of the synonyms *answer* and *retort* in different sections in COCA

Looking at the charts in figure 5.7, students might conclude that *answer* is more used in spoken English while *retort* is mostly used in fiction, and that *answer* is much more frequent.
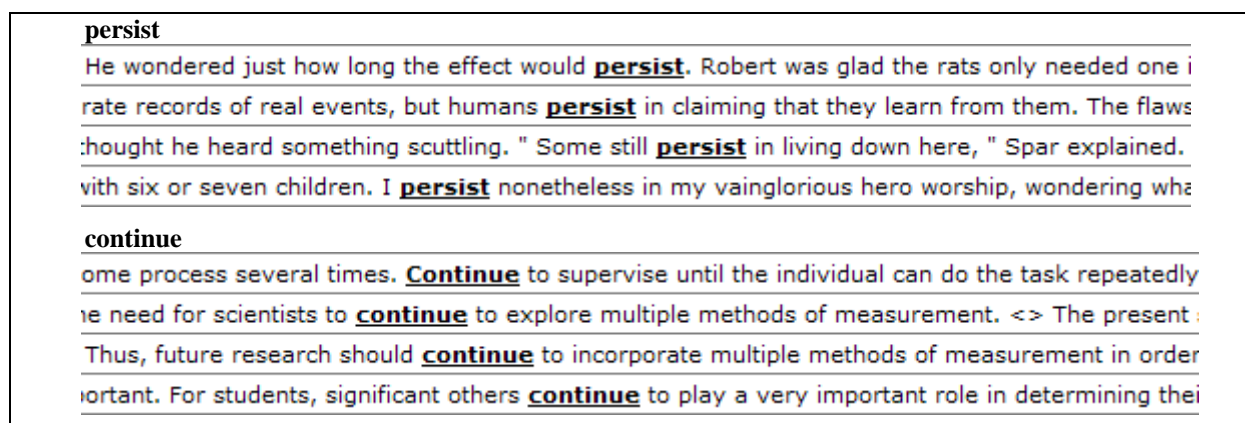

Figure 5.8 Concordance lines of the synonyms *persist* and *continue* in COCA

With reference to *persist* and *continue* in figure 5.8, students will probably see that *persist* is more commonly followed by the preposition *in* and *continue* by *to*. These are only some examples of how corpus can help students discover language patterns. L06 states that pupils must 'be allowed to choose tasks which will challenge them and give them the opportunity to explore, both alone and together with others' (2006b). Obviously, the textbooks do not need to present the raw data as displayed here, the information could be summarized and provided in other ways. The difference is that when observing raw data to draw conclusions about how language works, students learn inductively (discovery learning) and when the data is interpreted for them the learning is deductive (see section 2.6). In case students find working with raw data overwhelming, the exercise can be constructed simply *informed* by corpus. The main idea is that corpus information is used.

Turning now to grammatical collocations, McCarthy et al. (2006b:60) include an exercise in which students have to choose the correct preposition to complete a statement and then write if the statement is true of false (see example 4 in appendix 2). Thus, it works with collocations and relates them to familiar experiences. This type of exercise could be a follow up of a matching definitions exercise containing verbs that are part of grammatical collocations. Since this was not the case of any of the exercises investigated, I have used two verbs from exercise [21] to exemplify how this could be done, as shown in figure 5.9 below:

---

[21] Circle the correct prepositions to complete each sentence. Then write T (true) or F (false) depending on whether the sentence is true or false for you. Correct the false statements.

1. _____ I always adapt well (**in** / **to** / **at**) new environments.
2. _____ I think that the parents should provide (**with** / **to** / **for**) their children until they can get a job, no matter their age.

Figure 5.9 Suggestion of a grammatical collocation exercise based on exercise [21]

Textbooks can also supply tips about common usage based on corpus information. In *Touchstone* (McCarthy et al. 2006a), there is usually a 'curiosity' box in the vocabulary section where some data from corpora is provided (see example 5 in appendix 2). This could be done for example in relation to exercise [1] from @cross. In this exercise, the first word to be matched with its translation is *attractive* and the exercise is about a town in the UK. A curiosity box could be displayed next to the exercise, as follows:

---

[1] In the text about Berwick, the adjective *attractive* is used to describe a town. Here are the words most used with *attractive* in a corpus in order of frequency:

1. attractive woman   5. attractive women
2. attractive alternative   6. attractive place
3. attractive option   7. attractive girl
4. attractive man   8. attractive feature

Figure 5.10 Suggestion of a 'curiosity box' using the adjective *attractive*

This approach might be effective in getting students more interested in corpora without overwhelming them with frequencies and MI scores. Since the text is about a town in the UK, information from a British corpus might be more appropriate here. Many other suggestions could be given exploring all the potential of the corpus.

5.4 Third aim: Other ways of integrating collocations in language teaching

In addition to the suggestions given above, other activities can be prepared by teachers or offered by textbooks in order to include collocations in language teaching. Some options are,

as provided by Conzett (2000; see appendix 2, examples 7 and 8), make an 'odd one out' exercise with collocations; give a few collocates of a word and ask students to add others; make thematic collocation lists; and play collocation dominoes (writing part of a collocation in each end). Hill et al. (2000) recommend brainstorming collocations on a topic before students write a text about it; bringing to class an extra text on the topic of the essay where students can find useful collocations; exercises that teach alternatives to *very* (e.g. *highly qualified* and *bitterly disappointed*); matching collocates and using them to fill in gaps; etc.

One of my suggestions is to give students a list of collocations that convey cultural aspects to be discussed when appropriate to the theme of the lesson or the chapter. Another idea is to ask students to write down confusable words like *do* and *make* and dictate collocates to be written under the correct verb. Howarth says that learners tend to avoid collocations with delexical verbs (*take*, *make*, *give*, *do*, etc.) 'due to uncertainty over appropriate collocability' (1998:181); thus, working with these verbs can be a way of addressing the problem (see example 6 in appendix 2). Students can also be given sentences from their own writings which contain collocational mistakes to be corrected. In this case they might need to consult a corpus or collocation dictionary, which is a good opportunity to teach how to use them.

Teachers can also call attention to important collocations that appear in the texts from the textbook, as suggested by Conzett (2000:72; see section 2.4). The relevance can be according to, as discussed above, lesson or chapter topic, a grammatical point, a distinction from a false cognate in Norwegian, or the need of the collocations to write a text. Another option is to suggest collocates on the margin of students' essays when they have written an odd word combination. Woolard says that 'one obvious way of finding out which words our students do not expect to find together is through the mis-collocations they make in their production of language' (2000:30). If a student has written for example *high poverty* in an essay, the teacher can write on the margin *extreme/severe/absolute poverty*. The corpus can also be used to check students' doubts about word uses or to illustrate a grammatical point. Dealing with corpus examples can encourage discovery learning, so that students can infer the rules by looking at usage-based examples. If the teacher finds the concordance lines too demanding for students due to the amount of unknown words, he or she can previously edit the concordance. In a study by Chambers, one of the students affirms: 'I discovered that achieving results from my concordance was a highly motivating and enriching experience. I've never encountered such an experience from a textbook' (2005:120). Hence, the corpus can be used as an asset by teachers to improve the classes and to motivate students.

An idea suggested by Nation (2001:302) is to encourage students to keep word cards: the word to be learned can be written on one side and the meaning on the other and drawings can be included. Although Nation suggests that the cards should not carry too much information, I argue that collocates and example sentences should also be added to lead students from the fourth step in vocabulary acquisition – consolidating form and meaning to memory – to the fifth, knowing how to use the word (see Hatch & Brown 1995). Woolard (2000:43) suggests that students keep a lexical notebook which can be divided by situations (e.g. at the bank), functions (e.g. complaining), and topics (e.g. occupations). When students encounter the words again, they can add a sentence with the new context in their notebooks as a way of recycling and expanding their vocabulary knowledge.

McCarthy & O'Dell (2005) have compiled a resourceful book called *English collocations in use* that can be used by teachers to complement vocabulary work. It includes collocations taken from the Cambridge International Corpus which are taught grouped by themes such as 'travel', 'relationships', 'sport', 'music', 'study and learning', 'work', etc. They are first presented in context and can then be practiced through exercises. Learning tips are also given based on analysis of learner errors from the Cambridge Learner Corpus (McCarthy & O'Dell 2005:3).

In conclusion, course material and teachers should help students perceive the lexical patterns existing in the English language and point out that learning these patterns can develop fluency. Conzett states that what is most important for teachers, 'more than worrying whether or not something is a collocation, is to shift their and their students' focus away from individual words to chunks of language. These chunks improve the fluency and accuracy of the English students produce' (Conzett 2000:80). Teaching students to notice collocations in the language input they receive and to have a system to record them are some of the best learning strategies that teachers can equip learners with.

5.5 Summary

This chapter has attempted to discuss and interpret the results of the investigation taking into account the background in chapter 2 and the three aims outlined in chapter 1. Concerning the first aim and first research question, the percentages from table 4.55 reveal that more than one third of the single words from the exercises which appeared in the texts are part of collocations. This confirms the hypothesis that a high proportion of collocations would be found in the texts. Some peculiarities from each material have been discussed separately while the features observed both in *@cross* and *Passage* have been interpreted together. The

findings have shown that collocations can be useful to widen students' vocabulary and also to provoke discussions about sociocultural matters. As a means to address the second research question, a list of useful collocations for Norwegian students in the first year of upper secondary school has been suggested. As for the second aim, ideas for improving the exercises have been given based on previous research and on existing collocation exercises. Finally, achieving the third aim, ideas have been given on other ways in which collocations can be integrated into language teaching.

In the present study a measurable definition of collocation was needed in order to make quantitative claims. Consequently, combinations that were potentially useful for Norwegian students have been excluded either because of the information provided by the corpus or the dictionary. As shown in figure 4.1 in the results chapter, 33 combinations have been excluded by the use of the dictionary and 4 have been left out by the use of the corpus. In practice, the theme being discussed, students' L1, and cultural relevance are probably more important than whether a combination is or is not a collocation.

Many of the ideas discussed in the present study, and especially in this chapter, are connected to Beck et al.'s concept of 'rich instruction' for vocabulary teaching. They emphasize the importance of relating words to one another and to personal experiences, and say that discussing words is crucial for the learning process. This is expressed by their quotation below:

> Although the instruction would need to include associating words with definitions, it would need to go well beyond that. […] Students should be required to manipulate words in varied and rich ways, for example by *describing how they relate to other words and to their own familiar experiences*. To promote and reinforce deep processing, activities should include much *discussion of the words* and require students to *create justifications for the relationships and associations that they discover*. This feature we labeled rich instruction. (Beck et al. 1987:149, added emphasis)

All in all, this study has attempted to give an idea of the range of collocations that appear in textbooks and show their relevance for vocabulary teaching. Furthermore, it has sought to stress the potential of corpora for teaching and to show how it can be used to complement vocabulary exercises, which is especially relevant in the Norwegian context where computers and internet are readily available. It has also tried to emphasize the importance of integrating collocations not only in vocabulary practice but also in English teaching in general.

6. CONCLUSION

This chapter will summarize the findings of the present study and draw conclusions based on what has been discussed above. Suggestions for further research will be offered and the chapter will conclude with the main findings of the investigation.

6.1 Summary and conclusion

This thesis has attempted to show how pervasive collocations are in the English language and their importance for the teaching of English as a foreign language. In order to raise awareness of the phenomenon, the first aim was to investigate two textbooks for Norwegian upper secondary students to find out if the words presented in isolation in the vocabulary exercises were part of collocations in the texts they were taken from. Recapitulating, the first research question was:

(a) What proportion of these words appears in collocations in the texts?

Based on the results presented in chapter 4 and discussed in chapter 5, it has been found that 34.9% of all the isolated words in the selected exercises from @*cross* and *Passage* were part of collocations in the texts they were taken from. The individual result for each textbook was 35.1% in @*cross* and 34.6% in *Passage*. This shows that more than one third of all single words from the exercises appeared in collocations in the texts, which confirms the hypothesis that a high proportion of collocations would be found. Although what is a high or low proportion is difficult to define, we can conclude that above one third is more than one would normally expect given that the choice of words for the exercises did not take collocations into consideration. The next research question was:

(b) Which of these collocations are useful for Norwegian students in the first year of upper secondary school?

The collocations found in the texts were considered in relation to previous research combined with the competence aims of the Norwegian curriculum (L06). A list was provided in chapter 5 indicating which kinds of collocations would be useful for teaching. Some of these are: collocations related to the themes of the textbook, collocations of near synonyms, collocations to talk about news and current events, and collocations conveying sociocultural aspects. The

findings related to the first aim can be taken as evidence to claim that at least some of them should have been listed as chunks in the exercises.

Fulfilling the second aim, findings on corpora, collocations, and foreign language vocabulary acquisition have been used together with existing material on collocations to improve and complement the exercises selected from the books. The third aim has been achieved by suggesting other ways in teachers and textbook writers can integrate collocations into language teaching, particularly in the area of vocabulary.

Although this study deals with material designed for Norwegian students in the first year of upper secondary school, I argue that some of the implications can be transferred to students of other nationalities and at other levels. For example, many of the types of collocations that are recommended for teaching can be extended to any learner of English: collocations which are frequent and strong in the corpus, collocations revealing various meanings of a word, collocations that convey sociocultural aspects of language use, collocations of near synonyms, collocations of words that are already familiar to students but which expand the knowledge of the known word, and the ones related to themes being discussed in class (e.g. politics, environment, education, etc.). In addition, the conclusion that vocabulary exercises can be complemented by corpus information can also be applied to any English teaching context.

In sum, this thesis has detected a large amount of unexplored collocations in texts from textbooks that have the potential to improve textbook vocabulary exercises and to expand student's knowledge of single words. The present study has also attempted to show that corpora can be used as a resource to complement language teaching in general and that collocational knowledge can help students communicate more successfully and fluently.

6.2 Further research

Some interesting findings have been made in the course of this investigation which could be further explored. A future study could be to use the corpus to find the most common collocates of all near synonyms provided by the exercises and check the concordance lines for each of them. This would certainly lead to new findings about the use of these words. Other words that could be researched are synonyms that are problematic for students, such as *wound* and *injury* as exemplified by Morgan Lewis (2000:13; see subsection 2.5.4).

A study that would be valuable for language teaching would be to investigate collocations common to different topics or themes usually discussed in the classroom. Lists of

salient collocations could be compiled (see example 7 in the appendix) and recommendations could be given regarding how these collocations can be used in class.

The collocations given most attention in the present study were the ones found in the texts of the textbooks and it has not been discussed extensively what other specific collocations would have been more relevant to include in the exercises. The reason why the ones found in the texts were given prominence is the fact that students have encountered them already in the input, and including them in the exercises would help the input become intake. However, there are probably other useful collocations to be taught, as the ones suggested by Nesselhauf (2005:237-273) in her study of deviant collocations produced by German students. How these collocations can be integrated into the syllabus can be a topic for further research.

Another finding that could be studied in more depth is the sociocultural aspects conveyed by certain collocations. This could be carried out in connection with the general studies program for students of the second and third year of upper secondary school. Their main subject areas in English are 'international English', 'social English', and 'English literature and culture'. The textbooks for this course and/or various corpora could be used to find examples of culturally-loaded collocations that can serve as starting points for discussions in the classroom about other English speaking countries. It would also be interesting to conduct a study similar to the one by Teliya et al. (1998) discussed in chapter 2. Studying stereotypes, cultural concepts, connotations, and conceptual metaphors in English collocations seems to be a large and underexplored topic of research.

Yet another area that could be pursued in further studies is the teaching of English for the specialized courses in Norway – the vocational education programs. Although the competence aims established for these courses are the same as for the general studies, teaching specific collocations for the areas of restaurant and cooking, health and social studies, media and communication, among others, would be of great benefit to these students. Small corpora could be created by scanning texts within these fields and concordance programs available online could be used to look at concordance lines, collocations, frequencies, etc. (see Gavioli 2005). There is also a program online called 'Sketch Engine' (http://www. sketchengine.co.uk/; accessed on November 16, 2009) which can be used to build your own corpora. Using certain genres and subgenres of COCA is another option.

An investigation could also be conducted applying the ideas proposed by the present study. If the exercises suggested are further developed, they could be used in class and students' vocabulary knowledge could be tested. Corpus consultation can also be implemented. Testing students taught in more traditional ways and then comparing the results

can reveal if collocations indeed contribute to vocabulary learning. In addition, classroom procedure could be recorded and teachers and students could be interviewed regarding the use of corpora. In line with Yoon, 'we need an empirical report from actual teaching that uses easily accessible general corpora to encourage teachers and students to use the new corpus approach' (2008:32-33).

Finally, the treatment of vocabulary in the textbooks analyzed should not be judged only by the findings of the present study since the material selected comprises a small amount of all the vocabulary work provided by the book and website. A topic for further research would be to include other types of exercises in the investigation. Furthermore, the new editions of @*cross* and *Passage* could be investigated to find out if the vocabulary exercises have been changed and if collocations have been given a more central place.

6.3 Final remarks

The interest in collocations has increased in the last decades and considerable research has been done to find out about the collocational behavior of words. However, there seems to be a gap between what has been discovered and the application of the findings in language teaching. According to Meunier & Gouverneur, 'studies on the actual treatment of phraseology in ELT material are rare. Information on the selection of learning and teaching-prone formulaic sequences is nowhere to be found, and precise guidelines on how to teach formulaic sequences is just as scarce' (2007:121). This thesis attempts to fill this gap by providing new insight into the treatment of collocations in language material, by recommending types of collocations that can be taught, and by showing how these can be integrated in vocabulary exercises and teaching.

The findings of the present study can be summarized in four main points: (1) single words taught in the textbooks have a high proportion of collocates in the texts they are selected from, (2) these collocations convey relevant sociocultural information encoded in language use that can foment discussions in the classroom, (3) corpora can be used to improve vocabulary exercises, and (4) the integration of collocations in vocabulary teaching is feasible and it is beneficial for learners. Thus, I believe that this study has contributed to the field of applied linguistics, and more specifically, to the fields of corpus linguistics and English language education by trying to fill the gap between research findings on collocations and the applications of these findings in language teaching.

REFERENCES

Bartsch, Sabine. 2004. *Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence.* Tübingen: Gunter Narr Verlag.

Beck, Isabel L., Margaret G. McKeown, & Richard C. Omanson. 1987. The effects and use of diverse vocabulary instruction techniques. In *The nature of vocabulary acquisition*, eds. Margaret G. McKewon & Mary E. Curtis, 147-163. Hillsdale, N.J.: Lawrence Erlbaum.

Benson, Morton, Evelyn Benson, & Robert F. Ilson. 1997. *The BBI dictionary of English word combinations*. 3rd ed. Amsterdam: John Benjamins.

Bernardini, Silvia. 2000. *Competence, capacity, corpora: A study in corpus-aided language learning*. Bologna: CLUEB.

Bernardini, Silvia. 2002. Exploring new directions for discovery learning. In *Teaching and learning by doing corpus linguistics*, ed. Bernhard Kettemann & Georg Marko, 165-182. Atlanta, Georgia: Rodopi.

Bernardini, Silvia. 2004. Corpora in the classroom: An overview and some reflections on future developments. In *How to use corpora in language teaching*, ed. John M. Sinclair, 15-36. Amsterdam: John Benjamins.

Berns, Margie. 1990. *Contexts of competence: Social and cultural considerations in communicative language teaching.* New York: Plenum Press.

Berntzen, Jorunn & Alf Bårtvedt, eds. 2005. *Targets.* Oslo: Aschehoug.

Biber, Douglas, Susan Conrad, & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

British National Corpus. http://www.natcorp.ox.ac.uk/ or http://corpus.byu.edu/bnc/ (same interface as COCA; accessed November 16, 2009).

Brown, Dorothy F. 1974. Advanced vocabulary teaching: The problem of collocation. *RELC Journal* 5(2): 1-11.

Cambridge advanced learner's dictionary (*CALD*). 2nd ed. 2005. Cambridge: Cambridge University Press.

Carter, Ronald. 1987. *Vocabulary: Applied linguistics perspectives*. London: Allen & Unwin.

Carter, Ronald & Michael McCarthy, eds. 1988. *Vocabulary and language teaching*. London: Longman.

Carter, Ronald & Michael McCarthy. 1999. The English get passive in spoken discourse: Description and implications for an interpersonal grammar. *English Language and Linguistics* 3: 41-58.

Chambers, Angela. 2005. Integrating corpus consultation in language studies. *Language Learning & Technology* 9(2): 111-125. http://llt.msu.edu/vol9num2/chambers/ (accessed November 6, 2009).

Chambers, Angela & Íde O'Sullivan. 2004. Corpus consultation and advanced learners' writing skills in French. *ReCALL* 16(1): 158-172.

Chomsky, Noam. 1959. Review of *Verbal behaviour* by B.F. Skinner. *Language* 35: 26-58.

Church, Kenneth & Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16(1): 22-29.

Clear, Jeremy. 1993. From Firth principles – Computational tools for the study of collocation. In *Text and technology: In honour of John Sinclair*, eds. Mona Baker, Gill Francis, & Elena Tognini-Bonelli, 271-292. Amsterdam: John Benjamins.

Clear, Jeremy. 2000. Do you believe in grammar? In *Rethinking language pedagogy from a corpus perspective*: *Papers from the third international conference on teaching and language corpora*, eds. Lou Burnard & Tony McEnery, 19-30. Frankfurt: Peter Lang.

Coady, James. 1993. Research on ESL/EFL vocabulary acquisition: Putting it all in context. In *Second language reading and vocabulary*, eds. Thomas Huckin, Margot Haynes, & James Coady, 46-64. Norwood, N.J.: Ablex.

Coady, James. 1997. L2 vocabulary acquisition through extensive reading. In *Second language vocabulary acquisition*, eds. James Coady & Thomas Huckin, 225-237. Cambridge: Cambridge University Press.

Collins, Heloisa. 2000. Materials design and language corpora: A report in the context of distance education. In *Rethinking language pedagogy from a corpus perspective*, eds. Lou Burnard & Tony McEnery, 51-63. Frankfurt: Peter Lang.

Conzett, Jane. 2000. Integrating collocation into a reading & writing course. In *Teaching collocation*, ed. Michael Lewis, 70-87. London: Heinle, Cengage Learning.

Council of Europe. 2001. The user/learner's competences. In *Common European framework of reference for languages: Learning, teaching, assessment*, 101-130. Strasbourg: Cambridge University Press.

Cowie, Anthony P. 1981. The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics* 2: 223-235.

Cowie, Anthony P. 1994. Phraseology. In *The encyclopedia of language and linguistics*, eds. R.E. Asher & J. Simpson, 3168-71. Oxford: Pergamon Press.

Cowie, Anthony P., ed. 1998a. *Phraseology. Theory, analysis, and applications*. Oxford: Clarendon Press.

Daasvand, Siri. 2001. Some aspects of teaching and learning vocabulary in English as a foreign language. Unpublished thesis. Department of British and American Studies and Department of Teacher Education and School Development. University of Oslo.

Davies, Mark. 2008. Corpus of Contemporary American English (COCA). http://www.americancorpus.org (accessed October 23, 2009).

Ellis, Nick C. 1997. Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In *Vocabulary: Description, acquisition and pedagogy*, eds. Norbert Schmitt & Michael McCarthy, 122-139. Cambridge: Cambridge University Press.

Ellis, Nick C. 2001. Memory for language. In *Cognition and second language instruction*, ed. P. Robinson, 33-68. Cambridge: Cambridge University Press.

Fairclough, Norman. 1990. What might we mean by 'enterprise discourse'? In *Enterprise culture*, eds. R. Keat & N. Abercrombie, 38-57. London: Routledge.

Fano, Robert M. 1961. *Transmission of information: A statistical theory of communications*. Cambridge, MA: MIT Press.

Firth, John R. 1968. A synopsis of linguistic theory, 1930-1955. In *Studies in linguistic analysis*, eds. J.R. Firth et al., 1-32. Special volume of the Philological Society. 3rd ed. Oxford: Blackwell.

Fontenelle, Thierry. 1998. Lexical functions in dictionary entries. In *Phraseology*, ed. Anthony P. Cowie, 189-207. Oxford: Clarendon Press.

Gabrovšek, Dušan. 1998. Toward an interlingual collocational dictionary of English and Slovene. In *Language policy and language education in emerging nations: Focus on Slovenia and Croatia*, eds. Robert de Beaugrande, Meta Grosman, & Barbara Seidlhofer, 127-142. Stamford, CT: Ablex.

Gass, Susan M. 1988. Integrating research areas: A framework for second language studies. *Applied Linguistics* 9(2): 198-217.

Gavioli, Laura. 2005. *Exploring corpora for ESP learning*. Amsterdam: John Benjamins.

George, Horace V. 1983. Classification, communication, teaching and learning. English Language Institute, Victoria University of Wellington.

Halliday, Michael A.K. 1966. Lexis as a linguistic level. In *In memory of J.R. Firth*, eds. C.E. Bazell, J.C. Catford, M.A.K. Halliday, & R.H. Robins. 148-162. London: Longmans.

Hatch, Evelyn & Cheryl Brown. 1995. *Vocabulary, semantics, and language education*. Cambridge: Cambridge University Press.

Hausmann, Franz J. 1989. Le dictionnaire de collocations. In *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*, eds. Franz J. Hausmann, Herbert E. Wiegand, & Ladislav Zgusta, 1010-1019. Berlin: Walter de Gruyter.

Heian, Bente, Gro Lokøy, Brynjulf Ankerheim, & Ion Drew. 2006. *Experience.* Oslo: Gyldendal.

Hidalgo, Encarnación, Luis Quereda, & Juan Santana, eds. 2007. *Corpora in the foreign language classroom.* New York: Rodopi.

Hill, Jimmie. 2000. Revising priorities: From grammatical failure to collocational success. In *Teaching collocation*, ed. Michael Lewis, 47-69. London: Heinle, Cengage Learning.

Hill, Jimmie & Michael Lewis, eds. 1997. *The LTP dictionary of selected collocations.* Hove: Language Teaching Publications.

Hill, Jimmie, Morgan Lewis, & Michael Lewis. 2000. Classroom strategies, activities and exercises. In *Teaching collocation*, ed. Michael Lewis, 88-117. London: Heinle, Cengage Learning.

Hoey, Michael. 2000. A world beyond collocation: New perspectives on vocabulary teaching. In *Teaching collocation*, ed. Michael Lewis, 224-243. London: Heinle, Cengage Learning.

Hornby, Albert S., Edward V. Gatenby, & Harold Wakefield. 1942. *Idiomatic and syntactic English dictionary*. Tokyo: Kaitakusha.

Howarth, Peter. 1998. The phraseology of learners' academic writing. In *Phraseology*, ed. Anthony P. Cowie, 161-186. Oxford: Clarendon Press.

Howatt, Anthony P.R. (with H.D. Widdowson). 2004. *A history of English language teaching*. 2nd ed. Oxford: Oxford University Press.

Hulstijn, Jan H. 1993. When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *Modern Language Journal* 77: 139-147.

Hunston, Susan. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hymes, Dell. 1972. On communicative competence. In *Sociolinguistics*, eds. J.B. Pride & J. Holmes, 269-293. New York: Penguin.

Hymes, Dell. 1979. On communicative competence (extracts). In *The communicative approach to language teaching*, eds. C.J. Brumfit & K. Johnson, 5-26. Oxford: Oxford University Press.

Johansson, Stig. 2007. Using corpora: From learning to research. In *Corpora in the foreign language classroom*, eds. Encarnación Hidalgo, Luis Quereda, & Juan Santana, 17-28. New York: Rodopi.

Johns, Tim. 1994. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In *Perspectives on pedagogical grammar*, ed. T. Odlin, 293-313. Cambridge: Cambridge University Press.

Kennedy, Graeme. 1991. *Between* and *through*: The company they keep and the functions they serve. In *English corpus linguistics: Studies in honour of Jan Svartvik*, eds. Karin Aijmer & Bengt Alternberg, 95-110. London: Longman.

Kennedy, Claire & Tiziana Miceli. 2001. An evaluation of intermediate students' approaches to corpus investigation. *Language Learning and Technology* 5(3): 77-90.

Kjellmer, Göran. 1984. Some problems relating to the study of collocations in the Brown corpus. In *Computer corpora in English language research*, ed. Stig Johansson, 25-33. Bergen: Norwegian Computing Center for the Humanities.

Knowledge Promotion (L06). 2006a. Knowledge promotion – Kunnskapsløftet. http://udir.no/Artikler/_Lareplaner/_english/Knowledge-promotion---Kunnskapsloftet/ (accessed October 7, 2009).

Knowledge Promotion (L06). 2006b. Common core subjects in primary and secondary education. http://www.utdanningsdirektoratet.no/Artikler/_Lareplaner/_english/ Common-core-subjects-in-primary-and-secondary-education/ (assessed November 3, 2009).

Knowledge Promotion (L06). 2006c. Programme for specialization in general studies. http://www.utdanningsdirektoratet.no/Artikler/_Lareplaner/_english/Programmes-for-general-studies/Languages-Social-Sciences-and-Economics-Studies-/ (accessed October 7, 2009).

Krashen, Stephen. 1989. We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal* 73(4): 440-464

Krashen, Steven & Tracy Terrell. 1983. *The natural approach: Language acquisition in the classroom*. Oxford: Pergamon Press.

Larsen-Freeman, Diane. 1986. *Techniques and principles in language teaching*. Oxford: Oxford University Press.

Larsen-Freeman, Diane. 1997. Chaos, complexity, science and second language acquisition. In *Applied Linguistics* 18(2): 141-165.

Lewis, Michael. 1993. *The lexical approach: The state of ELT and a way forward*. Hove: Language Teaching Publications.

Lewis, Michael. 1997. *Implementing the lexical approach: Putting theory into practice*. Hove: Language Teaching Publications.

Lewis, Michael, ed. 2000. *Teaching collocation.* London: Heinle, Cengage Learning.

Lewis, Morgan. 2000. There is nothing as practical as a good theory. In *Teaching collocation*, ed. Michael Lewis, 10-27. London: Heinle, Cengage Learning.

McCarthy, Michael & Felicity O'Dell. 2002. *English vocabulary in use*. 2nd ed. Cambridge: Cambridge University Press.

McCarthy, Michael & Felicity O'Dell. 2005. *English collocations in use*. Cambridge: Cambridge University Press.

McCarthy, Michael, Jeanne McCarten, & Helen Sandiford. 2006a. *Touchstone 3 student's book.* New York: Cambridge University Press.

McCarthy, Michael, Jeanne McCarten, & Helen Sandiford. 2006b. *Touchstone 3 workbook.* New York: Cambridge University Press.

McEnery, Tony, Richard Xiao, & Yukio Tono. 2006. *Corpus-based language studies: An advanced resource book*. New York: Routledge.

McIntosh, Colin, Ben Francis, & Richard Poole, eds. 2nd ed. 2009. *Oxford collocations dictionary for students of English* (*OCD).* Oxford: Oxford University Press.

Mel'čuk, Igor. 1998. Collocations and lexical functions. In *Phraseology*, ed. Anthony P. Cowie, 23-53. Oxford: Clarendon Press.

Meunier, Fanny & Céline Gouverneur. 2007. The treatment of phraseology in ELT textbooks. In *Corpora in the foreign language classroom,* eds. Encarnación Hidalgo, Luis Quereda, & Juan Santana, 119-139. New York: Rodopi.

Moon, Rosamund. 1994. The analysis of fixed expressions in text. In *Advances in written text analysis*, ed. Malcolm Coulthard, 117-135. London: Routledge.

Moon, Rosamund. 1998. Frequencies and forms in phrasal lexemes in English. In *Phraseology*, ed. Anthony P. Cowie, 79-100. Oxford: Clarendon Press.

Moore, Johanna C. & John R. Surber. 1992. Effects of context and keyword methods on second language vocabulary acquisition. *Contemporary Education Psychology* 17: 286-292.

Nagy, William E., Patricia A. Herman, & Richard C. Anderson. 1985. Learning words from context. *Reading Research Quarterly* 20: 233-253.

Nation, Paul. 1990. *Teaching and learning vocabulary*. New York: Newbury House Publishers.

Nation, Paul. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nattinger, James R. 1988. Some current trends in vocabulary teaching. In *Vocabulary and language teaching*, eds. Ronald Carter & Michael McCarthy, 62-82. New York: Longman.

Nattinger, James R. & Jeanette S. DeCarrico. 1992. *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Nesselhauf, Nadja. 2005. *Collocations in a learner corpus*. Amsterdam: John Benjamins.

Osborne, John. 2000. What can students learn from a corpus?: Building bridges between data and explanation. In *Rethinking language pedagogy from a corpus perspective: Papers from the third international conference on teaching and language corpora*, eds. Lou Burnard & Tony McEnery, 165-172. Frankfurt: Peter Lang.

Palmer, Frank. R. 1976. *Semantics: A new introduction*. Cambridge: Cambridge University Press.

Palmer, Harold E. 1933. *Second interim report on English collocations.* Tokyo: Kaitakusha.

Paribakht Tahereh S. & Marjorie B. Wesche. 1993. Reading comprehension and second language development in a comprehension-based ESL programme. *TESL Canada Journal* 11: 9-27.

Partington, Alan, John Morley, & Louann Haarman, eds. 2004. *Corpora and discourse*. Bern: Peter Lang.

Partington, Alan. 1998. *Patterns and meanings*. Amsterdam: John Benjamins.

Pawley, Andrew & Frances H. Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In *Language and communication*, eds. Jack C. Richards & Richard W. Schmidt, 191-225. London: Longman.

Payne, M. E. 1988. Vocabulary learning strategies used by ESL students and their relationship to perceptual learning style preferences. Unpublished thesis. Linguistic Department, Brigham Young University, Provo, Utah.

Richards, Jack C. & Theodore S. Rodgers. 2001. *Approaches and methods in language teaching.* 2nd. ed. New York: Cambridge University Press.

Rodgers, Drew, Knut Skifjeld, Dilys Brown, & Martin Mulloy. 2006a. @*cross textbook*. Oslo: N.W. Damm & Søn.

Rodgers, Drew, Knut Skifjeld, Dilys Brown, & Martin Mulloy. 2006b. @*cross workbook*. Oslo: N.W. Damm & Søn.

Römer, Ute. 2004. Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching. In *Corpora and language learners*, eds. Guy Aston, Silvia Bernardini, & Dominic Stewart, 151-168. Amsterdam: John Benjamins.

Said, Edward. 1978. *Orientalism*. London: Routledge & Kegan Paul.

Schmid, Hans-Jörg. 2007. Entrenchment, salience, and basic levels. In *The Oxford handbook of cognitive linguistics*, eds. Dirk Geeraerts & Hubert Cuyckens, 117-138. Oxford: Oxford University Press.

Schmitt, Norbert. 1997. Vocabulary learning strategies. In *Vocabulary: Description, acquisition and pedagogy*, eds. Norbert Schmitt & Michael McCarthy, 199-227. Cambridge: Cambridge University Press.

Schmitt, Norbert. 2000. *Vocabulary in language teaching*. New York: Cambridge University Press.

Schmitt, Norbert & Michael McCarthy. 1997. *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press.

Seidlhofer, Barbara. 2000a. Operationalizing intertextuality: Using learner corpora for learning. In *Rethinking language pedagogy from a corpus perspective*, eds. Lou Burnard & Tony McEnery, 207-233. Frankfurt: Peter Lang.

Simensen, Aud Marit. 1998. *Teaching a foreign language. Principles and procedures*. Bergen: Fagbokforlaget.

Sinclair, John M., ed. 1987. *Looking up: An account of the COBUILD project*. London: HarperCollins.

Sinclair, John M. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, John M. & Antoinette Renouf. 1988. A lexical syllabus for language learning. In *Vocabulary and language teaching*, eds. Ronald Carter & Michael McCarthy, 140-160. London: Longman.

Steinberg, Danny D. & Natalia V. Sciarini. 2006. *An introduction to psycholinguistics*. 2nd ed. Harlow: Pearson Longman.

Stevick, Earl W. 1976. *Memory, meaning and method*. Cambridge, MA: Newbury House Publishers.

Stubbs, Michael. 1995. Collocations and semantic profiles: On the cause of the trouble with quantitative studies. In *Functions of Language* 2(1): 1-33.

Stubbs, Michael 1996. *Text and corpus analysis. Computer-assisted studies of language and culture.* Oxford: Blackwell.

Stubbs, Michael. 2001a. *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Stubbs, Michael. 2001b. Texts, corpora, and problems of interpretation: A response to Widdowson. *Applied Linguistics* 22(2): 149-172.

Swan, Michael. 1997. The influence of the mother tongue on second language vocabulary acquisition and use. In *Vocabulary: Description, acquisition and pedagogy*, eds. Norbert Schmitt & Michael McCarthy, 156-180. Cambridge: Cambridge University Press.

Sørhus, Theresa Bowles, Richard Burgess, & Trond Christian Anvik. 2006a. *Passage*. Oslo: Cappelens Forlag.

Sørhus, Theresa Bowles, Richard Burgess, & Trond Christian Anvik. 2006b. *Passage* website. Oslo: Cappelens Forlag. http://passage-old.cappelendamm.no/ (accessed October 28, 2009).

Sökmen, Anita J. 1997. Current trends in teaching second language vocabulary. In *Vocabulary: Description, acquisition and pedagogy*, eds. Norbert Schmitt & Michael McCarthy, 237-257. Cambridge: Cambridge University Press.

Takač, Višnja P. 2008. *Vocabulary learning strategies and foreign language acquisition*. Clevedon: Multilingual Matters LTD.

Taylor, John R. 2002. *Cognitive grammar*. Oxford: Oxford University Press.

Teliya, Veronika, Natalya Bragina, Elena Oparina, & Irina Sandomirskaya. 1998. Phraseology as a language of culture: Its role in the representation of a collective mentality. In *Phraseology*, ed. Anthony P. Cowie, 55-75. Oxford: Clarendon Press.

The BNC basic (C5) tagset. <http://www.natcorp.ox.ac.uk/docs/c5spec.html> (accessed October 27, 2009).

Utdanningsdirektoratet (Udir). 2007. Gratis læremidler i videregående opplæring. http://udir.no/Brev/_tilskott/Gratis-laremidler-i-videregaende-opplaring/ (accessed October 7, 2009).

Vestre, Sissel Kværne. 1998. Coping with words. A study of vocabulary acquisition and Norwegian learners of English. Unpublished thesis. Department of British and American Studies. University of Oslo.

West, Michael P. 1953. *A general service list of English words*. London: Longman.

Widdowson, Henry. 2000. The limitations of linguistics applied. *Applied Linguistics* 21(1): 3-25.

Williams, Raymond. 1985. Mining the meaning: Key words in the miners' strike. *New Socialist* 25: 6-9.

Willis, Dave. 1990. *The lexical syllabus*. London: Collins-COBUILD.

Woolard, George. 2000. Collocation – encouraging learner independence. In *Teaching collocation*, ed. Michael Lewis, 28-46. London: Heinle, Cengage Learning.

Yoon, Hyunsook. 2008. More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology* 12(2): 31-48. http://llt.msu.edu/vol12num2/yoon.pdf (accessed November 6, 2009).

Zimmerman, Cheryl Boyd. 1994. Self-selected reading and interactive vocabulary instruction: Knowledge and perceptions of word learning among L2 learners. Ph.D. diss., University of Southern California, Los Angeles.

Zimmerman, Cheryl Boyd. 1997. Historical trends in second language vocabulary instruction. In *Second language vocabulary acquisition*, eds. James Coady & Thomas Huckin, 5-19. Cambridge: Cambridge University Press.

Curricula

Læreplan for forsøk med 9-årig skole. 1960. Forsøksrådet for skoleverket. Oslo: Aschehoug.

Læreplanverket for den 10-årige grunnskolen (L97). 1996. Kirke-, utdannings- og forskningsdepartementet. Oslo: Nasjonalt Læremiddelsenter.

Mønsterplanen for grunnskolen. 1974. Kirke- og undervisningsdepartementet. Oslo: Aschehoug.

Mønsterplanen for grunnskolen. 1987. Kirke- og undervisningsdepartementet. Oslo: Aschehoug.

Normalplan for byfolkeskolen. 1957. Normalplankomitéen, Kirke- og undervisningsdepartementet. Oslo: Aschehoug.

This appendix lists only the competence aims for the programs for general studies (which are the same as the vocational education programs). For the objectives of the subject, main subject areas, teaching hours, and basic skills see the whole document (L06 2006b).

Competence aims after Vg1 – programmes for general studies
Competence aims after Vg2 – vocational education programmes

Language learning

The aims are that the pupil shall be able to

• exploit and assess various situations, work methods and strategies for learning English

• discuss similarities and differences between English and other foreign languages and use this knowledge when learning English

• use relevant and precise terminology to describe the forms and structures of English

• describe and assess own progress when learning English

• use a wide selection of digital and other aids independently, including monolingual dictionaries

Communication

The aims are that the pupil shall be able to

• master a wide vocabulary

• use the forms and structures of the language in spoken and written presentations

• understand extended written and oral presentations on different personal, literary, interdisciplinary and social topics

• extract essential information from spoken and written texts and discuss the author's attitudes and point of view

• express himself/herself in writing and orally with subtleness, proper register, fluency, precision and coherence

• select appropriate listening, speaking, reading and writing strategies adapted to the purpose, situation and genre

• take initiatives to start, finish and keep a conversation going

• read formal and informal texts in various genres and with different purposes

• write formal and informal texts with good structure and coherence on personal, interdisciplinary and social topics

• select and use content from different sources independently, critically and responsibly

• use technical and mathematical information in communication

• produce texts with complex content using digital media

• choose an interdisciplinary topic for in-depth studies within his or her own programme area and present this

Culture, society and literature

The aims are that the pupil shall be able to

• discuss social conditions and values in various cultures in a number of English-speaking countries

• present and discuss international news and current events

• explain the main characteristics of the development of English from an Anglo-Saxon language to an international world language

• analyse and discuss a film and a representative selection of literary texts in English from the genres poetry, short story, novel and drama

• discuss a selection of literary texts in English from various regions of the world and different periods from the 1500s up to the present

• discuss literature by and about indigenous peoples in the English-speaking world

• prepare and assess his or her own written or oral texts inspired by literature and art

This appendix lists the collocation exercises that were used as examples for the suggestions on how to improve the textbook exercises and for giving ideas on how collocations can be integrated in teaching.

Example 1

> Find these words and expressions in the article. Match them with the definitions.
>
> 1. do the trick ___g___   a. make use of
> 2. is right at your fingertips _____   b. a time when you need money
> 3. take advantage of _____   c. not having
> 4. a rainy day _____   d. a lot of money
> 5. count _____   e. is where you need it
> 6. doing without _____   f. make a difference
> 7. a bundle _____   g. be a solution

McCarthy et al. 2006b:48

Example 2

> **The missing verb**
>
> What are the missing verbs in the following collocations? The same verb completes all three examples. If in doubt, check the nouns in a collocation dictionary. Notice how important it is so learn words in phrases rather than single words.
>
> 1. ……………… a mistake   4. ……………… panic
>   a statement   a problem
>   an observation   embarrassment
>
> 2. ……………… to a complete standstill   5. ……………… danger
>   to an understanding   an accident
>   to a decision   a question
>
> 3. ……………… concern
>   embarrassment
>   fear

Michael Lewis 2000:112

Example 3

Words into groups

Match each of these nouns to one of the groups of verbs. Remember, all the verbs in the group must collocate with the noun.

 attack   battle   dispute   fight   struggle   war

Group A: avoid, get into, pick, provoke, start, win
Group B: declare, go to, lead to, prolong, wage, win
Group C: be engaged in, continue, face, give up, join, take up
Group D: be vulnerable to, carry out, launch, mount, resist, step up
Group E: fight, force, go into, lose, win
Group F: aggravate, get involved in, intervene in, put an end to, resolve

Now do the same with these:

 fine   penalty   punishment   sentence   discipline

Group A: heavy, lenient, suspended, life, long, reduced
Group B: harsh, heavy, severe, death, stiff, huge
Group C: heavy, hefty, immediate, on-the-spot, stiff, token
Group D: effective, firm, strict, slack, poor, excessive
Group E: appropriate, brutal, capital, fit, lenient, harsh

Michael Lewis 2000:115

Example 4

---

1. *If I had a problem...*

Circle the correct words to complete each sentence. Then decide if each sentence is true or

false for you. Write *T* (true) or *F* (false). Correct the false statements.

1. _____ If I had a problem, I would talk (**for / to / about**) my best friend.

    <u>If I had a problem, I would talk to my Aunt Lisa.</u>

2. _____ I always apologize (**about / to / for**) my mistakes.

    _____

3. _____ My friends never thank me (for / about / with) helping them.

    _____

4. _____ I always worry (for / about / from) taking tests!

    _____

5. _____ I usually share CDs (to / with / about) my neighbors.

    _____

6. _____ If I forgot (with / from / about) a friend's birthday, I would feel bad.

    _____

McCarthy et al. 2004b:60

Example 5

What do we make?

The most common collocations with the verb **make** are:
1. make **sure**   4. make **a decision**
2. make **sense**   5. make **a mistake**
3. make **a difference**   6. make **money**

McCarthy et al. 2006a:62

Example 6

Make up your mind

Complete the make and do expressions. Use the definitions to help you.

1. make a _____ = work to earn money
2. make a good _____ = make someone think of you positively
3. make a _____ = make a positive change
4. do your _____ = try your hardest
5. make up your _____ = decide
6. make _____ of = make jokes about and laugh at
7. make _____ = make certain
8. make a _____ = get something wrong
9. do the _____ = figure out the numbers
10. make _____ = seem logical

McCarthy et al. 2006b:44

Example 7

Thematic collocation lists

Prisons   **The workplace**

prison sentence   mental challenge
corrections office   prospective employees
prison-issue clothing   job autonomy
self-help courses   hourly wage
kill time   straight salary
re-entry into society   employee turnover
doing time   incentive schemes
alternative sentence   unskilled workers
prison capacity   external recognition

Conzett 2000:80

Example 8

Odd one out

Cross out the word which does not belong in the group:

potent car, *potent drink*, *potent drug*, *potent weapon*

Advance learners

For the following words, add one or more words with which it might be expected to occur:

potent – *potent drug*, *potent weapon*, ...
reality – *face reality*, *harsh reality*, *virtual reality*, ...

Conzett 2000:84