Chunks in the classroom

An experiment on the use of Cognitive Linguistic principles in academic vocabulary instruction among Norwegian upper secondary ESL-students

Bjørnar Meling

Master Thesis in English Linguistics

Department of Foreign Languages

University of Bergen

May 2019

Abstract in Norwegian

Ordsekvensar, også kalla 'chunks' på engelsk, er ein viktig del av det engelske vokabularet. Denne masteroppgåva undersøkjer mogelege effektar av å bruka læringsstrategiar inspirert av kognitiv lingvistikk i undervisning av akademiske ordsekvensar.

Forsking frå kognitiv lingvistikk viser at dei semantiske og strukturelle eigenskapane ved mange ordsekvensar ikkje er tilfeldige. Dei kan forklarast med utgangspunkt i prinsipp om likskap eller nærleik som dannar grunnlag for til dømes metaforar, rim og allitterasjonar. Lingvistiske studiar viser også at undervisningsmetodar som fører til auka forståing av desse mekanismane, og auka evne til å kjenna igjen ordsekvensar i språket, er pedagogisk effektive.

For å undersøkja desse effektane nærmare, vart det gjennomført eit eksperiment med to eksperimentgrupper og ei kontrollgruppe. Deltakarane gjennomførte fire ulike oppgåver. Masteroppgåva måler bruk, evne til gjenkjenning og forståing av ordsekvensar. Den same testen vart nytta som pre-test og post-test før og etter eit undervisningsopplegg. Dette undervisningsopplegget nytta didaktisk metode basert på teori om 'chunks'. Testresultata blei analysert med ein ANCOVA-test, som presenterte blant anna sannsynsverdiar (p-verdiar) og verdiar for effektstorleik, samt statistisk deskriptive mål i form av gjennomsnittsverdi og standardavvik.

Studien fann at læringsopplegget hadde ein liten, positiv effekt på gjenkjenning av ordsekvensar, men at opplegget ikkje hadde signifikant effekt på eksperimentdeltakarane sin bruk og forståing av ordsekvensar. Analysen indikerte også at akademiske idiom er særleg eigna i undervisning inspirert av kognitiv lingvistikk. Funn i oppgåva tyder også på at høgt-presterande og homogene klassar får større utbyte av undervisningsopplegg basert på kognitiv lingvistisk teori.

Acknowledgements

First, I would like to thank my supervisor Dagmar Haumann for helpful advice and valuable guidance throughout my writing process.

Furthermore, I would like to express my gratitude to Michel Cabot for his invaluable proofreading. Also, a big thank you to all the students who made the experiment come true!

Finally, I am grateful to my family; my father, brother, sister, and to my kind and supportive girlfriend, who encouraged me and helped with everything from making statistical charts to babysitting.

Table of contents

List of tables

List of figures

List of abbreviations

ANCOVA	analysis of covariance
AVL	Academic Vocabulary List
AWL	Academic Word List
BICS	Basic interactive communication skills
CALP	Cognitive/academic language proficiency
CL	Cognitive Linguistics
CM	conceptual metaphor
EAP	English for academic purposes
ESL	English as a second language
FLT	foreign language teaching
GSL	General Service List
L1	first language
L2	second language
LME(s)	linguistic metaphorical expression(s)
SLA	second language acquisition
VKS	Vocabulary Knowledge Scale
VLT	Vocabulary Levels Test

1. Introduction

1.1 Outline and motivation

Vocabulary is an important part of language learning, as it is essential for all kinds of language proficiency. It has been said that people often think of language as knowledge of words, and this fact is made clear by the sayings and expressions seen in everyday speech all around the world (Singleton 1999: 8). This thesis focuses specifically on strategies for learning and teaching academic vocabulary in the form of *chunks*, which is defined as 'sequences of words which native speakers feel is the natural and preferred way of expressing a particular idea or purpose' (Lindstromberg and Boers 2008a: 7). The concept of chunks relies on the idea that certain words have 'an especially strong relationship with each other in creating their meaning' (Wray 2008: 30).

Numerous introductory volumes to the field of English vocabulary research have pointed to the significance and widespread use of chunks in language (Carter 2012; Nation 2013; Schmitt 2015). Chunks include various kinds of multi-word units, such as idioms ('kick the bucket', 'all hands on deck'), phrasal verbs ('bring about', 'keep up'), noun phrases ('great expectations', 'stomach pain'), binomials ('thick and thin', 'bread and butter'), and compositional collocations ('commit suicide', 'as a matter of fact'). There are many terms strongly related to chunks, such as formulaic sequences, prefabricated word strings (also known as prefabs), phrases, as well as fixed expressions and idioms (FEIs). This thesis uses several of these terms when referring to other studies; my usage of the terms mirrors the way in which the terms are used in these studies. However, in the reporting of the results of the tests conducted in this thesis, I consistently use the term *chunks*.

The term *chunks* was adopted in this thesis because it is a relatively frequent word in the English vocabulary; to date, more than 100,000 entries in the newly released iWeb Corpus (Davies 2018) have been counted for the term. Moreover, to most people, the term *chunks* has several familiar meanings that relate to everyday life (e.g., 'a chunk of bread' or 'a chunk of the profit'). Because of these features, this term is easier to explain, and easier to remember, than some of the alternatives mentioned above. Consequently, the term is suitable in pedagogical contexts, such as the ones described in this thesis.

English vocabulary, specifically English *academic* vocabulary, can be a valuable tool for several reasons. First, the students may need this kind of vocabulary in higher education or in their professional careers at a later stage. Secondly, chunks are essential for the mastery of several *competence aims* for the upper secondary school level in the Norwegian national curriculum. Some of the more relevant competence aims entail enabling the student to perform the following:

(1) express oneself fluently and coherently in a detailed and precise manner suited to the purpose and situation
(2) introduce, maintain and terminate conversations and discussions about general and academic topics related to one's education programme
(3) understand and use an extensive general vocabulary and an academic vocabulary related to one's education programme
(4) understand the main content and details in texts of varying length about different topics

(Utdanningsdirektoratet 2006)

These competence aims are normally incorporated in the written exams at the end of each English course as well; in these exams, students typically receive tasks that require them to read non-fiction texts of current interest, and to discuss the 'pros and cons' of a subject.

Academic vocabulary can be defined as 'words common in different kinds of academic texts' (Nation 2001: 12). In this definition, however, academic vocabulary is limited to academic *words*. This thesis focuses on *word strings, and more specifically— chunks* as part of academic vocabulary. Along this line, the thesis discusses how and whether academic vocabulary differs from other kinds of vocabulary, by examining older and more recent corpus-based vocabulary lists.

Cognitive Linguistic (CL) research may provide important knowledge about methods for teaching and learning vocabulary, including academic chunks, because this field of research offers theories on how to understand and categorize vocabulary in meaningful ways. Studies show that categorisation of words plays a key role in retrieving and memorising vocabulary, as pointed out by Sophia Skoufaki (2008): '[…] in unprompted free recall tasks people tend to retrieve words in category clusters [and] when words are presented in categories they are better remembered' (Schmitt 1997 in Skoufaki 2008: 102). Recent studies in Cognitive Linguistics have provided tools for introducing vocabulary by means of categories; specifically, this is done by explaining the semantic and structural patterns that *motivate* the use of language (Boers et al. 2004; Boers and

Lindstromberg 2008c; Bobrova and Lantolf 2012; Falck and Gibbs 2013; Noroozi and Salehi 2013). In Cognitive Linguistics, linguistic motivation is viewed as both primary and pervasive in language (Boers and Lindstromberg 2008b). In other words, language use is not arbitrary, It is contingent on human experiences that are processed by means of conceptual categories and phonological properties. The concept of *linguistic motivation* is a central premise for the teaching principles that are presented in this thesis as part of CL-inspired teaching.

The methods of CL-inspired teaching are derived from specific *cognitive processing theories*, namely (a) levels of processing theory, (b) dual coding theory, and (c) trace theory. These theories explain the need for complex tasks which promote the *deep-processing* of language and the analysis of motivational patterns in words and expressions—also known as *elaboration* in applied CL theory. The theories also explain the need to learn language through the use of different senses, creating awareness of imagery in content, and visualising imagery in presentation of chunks. Cognitive processing theories may also explain the way CL-inspired teaching is often organised. This kind of teaching can involve introducing target vocabulary through a variety of noticing activities, elaborating on different aspects of the vocabulary, and repeating the target vocabulary after a certain amount of time.

The goal of the experiment that is described in this thesis is to examine the effects of CL-inspired teaching on L2 speakers' knowledge of academic chunks. This knowledge includes *awareness* of chunks, ability to comprehend the *meaning* of chunks, and ability to *recognize* and *reproduce* chunks. To examine the effects, the thesis uses a quasi-experimental research design—including a pre-test and a post-test—to measure the knowledge of a limited number of target items.

Several earlier studies (Olsen 1999; Skoglund 2006; Lervåg and Aukrust 2010) suggest that Norwegian L2 learners of English lack the sufficient knowledge of vocabulary that they need in order to succeed in academic writing and academic text comprehension in further education. One of these studies (Skoglund 2006) concludes that Norwegian learners have substantial shortcomings in the area of vocabulary size and knowledge:

> Norwegian learners of English have a relatively small vocabulary and a lack of
> vocabulary knowledge. This deficiency could hinder Norwegians in the future,

but with the help of further research, improved teaching, and interest from all parties concerned, vocabulary skills could improve.

(Skoglund 2006)

These findings are of particular concern to educators in upper secondary school, since this is the highest level of English required in compulsory education. Upper secondary school is also the final preparatory stage before further English studies at university or college. The importance of vocabulary knowledge is reflected in the curriculum for English instruction in upper secondary education. Vocabulary is a fundamental requirement for many of the four listed competence aims listed above (Utdanningsdirektoratet 2006).

In addition to targeting schools and educators, this study is aimed at learners of *English as a second language* (ESL). Several studies on Norwegian language learners of English have noted that the boundary between English as a first language, ESL, and English as a foreign language (EFL), is far from clear-cut (Hellekjær 2005; Rindal 2014). While English is not an official second language in Norway, Rindal observes that the country 'has seen an increase in English language access and domain use' in recent years, and that English is frequently characterized as a second language in Norway (Rindal 2014: 8). Since many of the works of literature that are referred to in this thesis apply the term *second language* (L2) rather than *foreign language*, this thesis opts for the term ESL rather than EFL when discussing the language acquisition and use of English among Norwegian speakers. Clearly, however, some of the literature on first language acquisition is not relevant to Norwegian students of English. Nevertheless, although this study is primarily aimed at Norwegian ESL learners, the study is also aimed at vocabulary acquisition more generally. This is why some parts of the literature on vocabulary knowledge and acquisition (or language learning in general) does not address L2 speakers in particular.

1.2 The research question

The overarching research question of this study is as follows:

How does teaching inspired by Cognitive Linguistics compare to traditional teaching, in terms of recognition, understanding and application of academic chunks among Norwegian ESL-students in upper secondary school?

The question is formulated in accordance with Andrews's (2003: 23–50) assertion in that the *main research questions* should mirror the core aim of the study. It is possible to argue that there are research questions which contribute to or *derive from* this question; such questions would be known as *subsidiary research questions* (cf. Andrews 2003). For instance, it is necessary to know what is entailed in the teaching of academic chunks that is inspired by Cognitive Linguistics; this subtopic ultimately relates to the theoretical framework of the study, which is dealt with in Chapter 2, and thus the question does not need to be explicitly stated here. Another thing that would be useful to know is the level of vocabulary knowledge among Norwegian L2 learners of English, and the relevant forms of vocabulary knowledge. However, this would require an in-depth examination that falls outside the scope of the thesis. Finally, one might ask *which* of the learning activities that are inspired by Cognitive Linguistics, that have been most effective with regards to recognition, knowledge, and use of academic chunks. However, an examination of this question would require a comparison between different types of CL-inspired teaching that would call for a more extensive, long-term experimental design.

1.3 Thesis structure

Chapter 2 introduces theory on the topics of L2 vocabulary learning, chunks, and academic vocabulary; the chapter also describes the Cognitive Linguistic framework and how this can be applied to vocabulary teaching and learning. Chapter 3 explains the methods used in the thesis. In Chapter 3 I discuss the selection of target items, sampling, the teaching sessions, the testing procedures, the analysis of the test results, questions related to reliability and validity, practical limitations, and potential ethical issues. Chapter 4 presents and analyses the results from each of the test components in the experiment, and provides a summary of these test results. Chapter 5 discusses the test results in light of the theory and relevant findings from other studies. Lastly, Chapter 6 summarizes the findings, provides an answer to the research question, discusses pedagogical implications of the findings, and points to potential implications for further research.

2. Theoretical framework

This chapter presents the theoretical framework for the present study. Section 2.1 examines vocabulary knowledge and acquisition, both in general, and also more specifically among L2 learners of English.

Section 2.2 addresses the principle of idiomaticity in language, and introduces the term *chunks* as well as related terms. The section also presents different types of chunks. Section 2.3 discusses the notion of academic vocabulary, and explains what kind of words and expressions are included in academic vocabulary lists. It also considers the importance of chunks in academic vocabulary, and it discusses some of the critiques of the notion of academic vocabulary. Section 2.4 introduces Cognitive Linguistics and explores the connection between research on second language acquisition (SLA), chunks, and CL-theory. This section also examines the methods of organising CL-inspired teaching of L2 vocabulary.

2.1 Vocabulary and second language learners of English

This section outlines some central principles of vocabulary knowledge and learning, as well as how knowledge of vocabulary can be measured. Here, the study draws upon the distinction between incidental and deliberate/intentional learning, and presents basic elements of vocabulary learning programs. The section also presents research on vocabulary acquisition theories and on the learning burden that is associated with different kinds of vocabulary.

2.1.1 Vocabulary as knowledge of words

One understanding of *vocabulary* is '[…] the body of words used in a particular language' (Merriam Webster 2018). The term variably refers to a number of different concepts, including how many words a person knows (i.e., breadth knowledge) or how well the person knows certain words (i.e., depth knowledge), what is the extent to which a person can recognise form or meaning of words, or how many 'content words' a person knows. The understanding of vocabulary thus depends on the understanding of words.

Singleton (1999: 10–11) examines three possible routes to understanding the concept of *word*. The first relates to the way we count words. In the phrase 'going, going, gone'—which is often used to close bidding in auctions (cf. Merriam Webster 2018)—there are three units or *tokens*, whether identical or not. However, there are two *types* of

units, units that can be distinguished from one another, namely the units *going* and *gone*. We can also consider the constituents of the phrase instances of the paradigm of *go*, in which case we have one *word expression* or *lexeme* (Singleton 1999: 10). In corpus research, there is a need to use word categories when counting words, since learners might know different inflections or derivations of a word if they know the main word, which in this case is *go*. The two most common categories are *lemmas* and *word families* (Nation 2013: 10–11). A lemma is the main form of a word, a *headword* and its inflections (e.g., *proceed, proceeds, proceeding*), while a word family consists of a headword and its closely related *derivations* (e.g., *process, proceed, procedure, proceedings, procedural*). Lemmas may include many forms within a certain word class, whereas a word family may also include words from different word classes.

The second way of understanding *words* is by way of linguistic levels (Singleton 1999: 12). This refers to the principle that a word can be approached as different kinds of entities, for example as an orthographic entity with a series of letters, a phonetic entity with particular acoustic properties, a morphosyntactic entity that requires a certain position in sentences, or a semantic entity by virtue of its meaning and its associations to other words (Singleton 1999: 11).

Singleton's third dimension in approaching the concept of *words* concerns 'the extent to which semantic content is being treated as criterial' (Singleton 1999: 10–11). This relates to the distinction between *content words* and *grammatical words*. Content words such as *boat*, *tree*, or *house* have substantial meaning even out of context. Grammatical words such as *if*, *of*, or *the* have little or no independent meaning; they have predominantly a grammatical function. The distinction that is made here is not always straightforward, though, because many words may be considered part of both categories. Examples include the preposition *within* and the conjunction *while*.

Nation (2013) has a slightly different approach when he separates between different kinds of word knowledge in terms of *form*, *meaning*, and *use*. *Form* includes knowledge of pronunciation, sound, and written form. Meaning includes knowledge of what words mean and how they relate to other words semantically. Lastly, use includes knowledge of where, when, and how to use the word.

2.1.2 Vocabulary size and L2 learners

Vocabulary knowledge is an essential part of mastering a language. According to a range of studies, there is a close relationship between vocabulary size and reading (Laufer 1992), vocabulary size and class grades (Laufer and Goldstein 2004), L2 reading ability, and lexical size (Albrechtsen et al. 2008). According to a study by Alderson (2006), vocabulary size constitutes 37–62% of the variance in the proficiency scores of writing, listening and reading.

While these numbers show that a certain amount of vocabulary is needed for various language proficiencies, they do not show *how much* is needed for specific language tasks, let alone how much vocabulary is needed for L2 learners to be able to use vocabulary the way native speakers do. According to Nation (2013: 12), a native speaker accumulates around 1000 word families each year until the age of 25, where they will have accumulated a vocabulary of approximately 25 000 word families. As mentioned earlier, word families consist of both words and their derived forms, which means that the number of single words (i.e., tokens) in a native speaker's repertoire may well be a six-digit number. As for the number of words needed for an L2 learner of English, Schmitt (2015: 6–8) suggests that a knowledge of high-frequency vocabulary from the thousand most common word families is essential, but in order to read a variety of texts without unknown vocabulary being a problem, learners may need up to nine thousand word families. Nation (2013: 14) used texts based on the thousand most frequent word families from the British National Corpus. Nation argues that a 98 percent coverage should be a common goal for learners, since 2% unknown words can be regarded as 'manageable'. In order to understand 98 percent of the words in these texts, he found that readers need to know between six thousand and nine thousand word families.

2.1.3 Receptive and productive vocabulary

According to Schmitt (2015: 80), the aspects of word knowledge are so numerous and multifaceted that it would be difficult and extremely time-consuming to include all aspects in a test-battery. Hence, there is a need for a pedagogically useful and feasible way of categorising vocabulary knowledge for measurement. Many aspects of vocabulary knowledge can be understood in terms of the *form-meaning link* (Schmitt 2015: 49–50). This is a premise for many vocabulary tests, where participants must either explain or translate a certain target vocabulary or provide the target vocabulary in question after they

have been given meaning-clues. Another important aspect of vocabulary knowledge relates to the distinction between active and passive knowledge (Laufer and Goldstein 2004) or productive and receptive knowledge (cf. Nation 2013; Schmitt 2015). Passive or receptive knowledge is needed to 'receive language input from others through listening or reading and [...] comprehend it' while active or productive skills are needed to 'produce language forms by speaking or writing to convey messages to others' (Nation 2013: 46–47). These categories are not always easy to distinguish. For example, one might argue that listening and reading include productive and active features, such as producing meaning. Nevertheless, Laufer and Goldstein (2004) consider two types of passive knowledge and two types of active knowledge. These are passive recognition and passive recall, and active recognition and active recall. By contrast, Schmitt (2015) uses the terms *meaning recognition* and *meaning recall*, and *form recognition* and *form recall* In this thesis, these latter terms are throughout the rest of the study unless in those instances where I refer to Laufer and Goldstein's (2004) research. In the vocabulary tests measuring the two categories of receptive knowledge, the form is given, and the learner must either choose the correct L1 translation or explanation from several alternatives (i.e., meaning recognition) or produce the meaning by means of an L1 translation or explanation (i.e., meaning recall). Conversely, when productive knowledge is tested, the meaning is given in the form of a context, an explanation, or as an L1 translation, and the form must be recognised from several alternatives (i.e., form recognition) or it must be produced (i.e., form recall). Table 2.1 shows these four test formats, which Laufer and Goldstein (2004) refer to as degrees of vocabulary knowledge.

Table 2. 1: Degrees of vocabulary knowledge.

	Recall	**Recognition**
Active (retrieval of form)	Supply the L2 word	Select the L2 word
Passive (retrieval of meaning)	Supply the L1 word	Select the L1 word

Note: Adapted from Laufer and Goldstein (2004: 407)

Apart from the fact that these different test formats measure different kinds of vocabulary knowledge, and therefore should be applied accordingly, they are also different with regards to difficulty, relevance to everyday interpersonal communication, and practical

considerations. In accordance with earlier research, e.g. Stoddard (1929; in Nation 2013: 54), Laufer and Goldstein (2004) found evidence that these four test formats form a hierarchy of difficulty. They tested 435 L2 learners in high schools and universities using all four test formats, and they found that active recall proved to be the most difficult, thus representing 'the highest degree of form-meaning knowledge strength' (Schmitt 2015: 85). Passive recognition proved to be the easiest test format in the same study. Based on Laufer and Goldstein's (2004) research, it can be stated that as a general rule, active/productive tasks are more difficult than passive/receptive tasks, and recall tasks are more difficult than recognition tasks. Schmitt also pointed out that the recognition tasks are less relevant for everyday life use of vocabulary, since this kind of knowledge is typically only used when choosing entries in a dictionary after having looked up a word. In the majority of interpersonal communication involving everyday life situations, people are expected to have the form-meaning link already established at the recall knowledge level (Schmitt 2015: 88). However, recognition tasks can be more efficiently administered through multiple-choice tests that automatically register the correctness of the answers, whereas recall tasks depend on researchers' own judgement of the correctness of each answer (i.e., produced form or meaning), which would be determined subjectively.

In conclusion, recognition tasks are more suitable for larger samples where the administration of recall tests would be too time-consuming, while recall tasks may provide more accurate information about an individual's overall vocabulary knowledge.

2.1.4 Breadth and depth of vocabulary

Schmitt (2015: 187) highlighted the importance of being able to distinguish between the *breadth* and the *depth* of knowledge of vocabulary. While some tests focus on the quantitative aspect of vocabulary knowledge, by counting the vocabulary *size* (i.e., breadth), other tests set out to measure the *level of knowledge of certain words* (i.e., depth).

According to Schmitt (2015: 197), one of the best known vocabulary size tests aimed at ESL-learners, is the Vocabulary Levels Test (VLT). This test measures vocabulary knowledge at four frequency levels of word families, namely 2000, 3000, 5000, and 10 000. Each of these levels is considered necessary in achieving its corresponding key goals: (a) 2000 word families are sufficient for engaging in daily

conversation; (b) 3000 word families are needed for the initial access to authentic reading; (c) 5000 word families are necessary for independent reading; and (d) 10,000 word families would enable learners to have an advanced usage in most settings. The VLT uses a *form recognition* test format where the participant chooses the correct form corresponding to a short 2–5 word-explanation. As discussed in section 2.1.3, the recognition test format has its advantages and disadvantages. This format can be administered efficiently on a large scale, but it measures knowledge that is not in itself relevant to everyday life situations.

While vocabulary tests that measure breadth operate with easily quantifiable variables such as lemmas and word families, vocabulary tests that measure depth require well considered definitions of *levels of knowledge of vocabulary items*. The level of knowledge of a word should not be regarded as a continuous variable in such measurements. The idea of a scale ranging from *knowing nothing* to *knowing everything* about a word makes little sense, particularly when applied to SLA, as explained by Schmitt (2015: 217): 'If a person knows the spelling, pronunciation, and morphological rules of language, then they will already know something about almost any new lexical item they meet'. Nevertheless, there are some standardized tests for measuring depth of vocabulary; these tests use categories that reflect the hierarchy of the task difficulty which is related to receptive and productive vocabulary, as discussed earlier. One such test is the Vocabulary Knowledge Scale (VKS), the most widely used vocabulary depth test according to Schmitt (2015: 218). This test uses a five-level scoring system to determine the knowledge of depth of each target item. Each level is measured by self-reporting or a combination of self-reporting and a demonstration of knowledge by either translating the item, supplying synonyms, using the target items in sentences, or a combination of the above (as shown in Table 2.2).

Table 2. 2: The five-level scoring scale in the Vocabulary Knowledge Scale (VKS).

I.	I don't remember having seen this word before.
II.	I have seen this word before, but I don't know what it means.
III.	I have seen this word before, and I think it means _____. (synonym or translation)
IV.	I know this word. It means _____. (synonym and translation)
V.	I can use this word in a sentence: _____. (Write a sentence.) (If you do this section, please also do Section IV.

Note: Adapted from Schmitt (2015: 218–219)

Schmitt (2015) listed several problematic issues concerning the Vocabulary Knowledge Scale (VKS). One problematic issue is that the two first tasks are only self-reporting tasks, while the three others also require a demonstration of vocabulary knowledge. Another concern is that the ordinal nature of the scoring system in the test makes the test inappropriate for parametric statistics, which requires a continuous dependent variable. The third challenge concerns the fact that the intervals between the knowledge levels are inconsistent—that is, the first four levels measure the form-meaning link, whereas the fifth category jumps to 'strong enough mastery to use the word in a semantically appropriate way in a sentence' (Schmitt 2015: 220). It is important to recognise these issues when interpreting results from tests using the VKS.

2.1.5 L2 vocabulary learning and the role of the language teacher

There are many avenues to learning vocabulary, and the process of learning vocabulary takes place in a variety of arenas of life. However, while native speakers learn most of their language through interpersonal encounters in everyday life situations, non-native speakers acquire language more extensively in a classroom setting. It is therefore important that vocabulary learning programs for L2 learners are organised in a pedagogically balanced way. The programs should include different kinds of student activities, such as meaning-focused and language-focused activities. The programs should also include both teacher-instructed and independent work. Nation (2013: 2) provides a useful overview for teachers and researchers of vocabulary, in which he presents four 'strands'—that is, four basic elements that should be part of a language

learning course, namely meaning-focused input, language-focused learning, meaning-focused output, and fluency development. Table 2.3 shows the four strands and the teaching and learning activities associated with each strand.

Table 2. 3: The four strands of a well-balanced language learning program

The four strands	Activities and techniques
Meaning-focused input	• Reading graded readers • Listening to stories • Engaging in communication activities
Language-focused learning	• Direct teaching of vocabulary • Direct learning • Intensive reading • Training in vocabulary strategies
Meaning-focused output	• Communication activities with written input • Prepared writing • Linked skills
Fluency development	• Reading easy graded readers • Repeated reading • Speed reading • Listening to easy input • 4/3/2 • Rehearsed tasks • Ten-minute writing • Linked skills

Note: Adapted from Nation and Webb (2011: 2).

A well-balanced learning program can implement the four strands with the following steps and principles in mind. First, learners should encounter new language from reading and listening to the language in a way that they can focus on the content presented through language. This happens best if the learners are familiar with more than 98% of the words they encounter. Secondly, learners should engage in a 'usefully-focused' *deliberate*

learning of language items, since 'the cumulative process of learning new words can be given a strong boost by the direct study of certain features of the word' (Nation 2013: 2). Thirdly, learners should be subject to speaking and writing activities because these activities require a different focus than listening and reading activities. Speaking and writing activities also encourage the learners to 'listen like speakers' and 'read like writers'. Finally, learners should practice fluency by using the vocabulary they already know. Whereas vocabulary use requires vocabulary knowledge, vocabulary knowledge also requires vocabulary use. Thus, without fluency training, the three other strands become less effective (Nation 2013: 2).

It is worth pointing out that the direct teaching of vocabulary is only part of the second strand, which is language-focused learning. Each of the strands is also assigned a similar amount of time in Nation's language program. This means that the three meaning-focused strands (i.e., 1, 3, 4) should be allotted 75 percent of the time, and the language strand should be given 25 percent. However, the meaning-focused strands require only small amounts of effort from the teacher in the form of guidance or time-taking. On the other hand, the language-focused strand often calls for more effort as a result of planning and facilitation. For this reason, there is a tendency that too much time is spent on instructed, language-focused activities in the classroom, according to Nation (2013: 95).

Historically, several of the principles underlying this proposed language program have frequently been subject to debate by language researchers, one of which is Stephen Krashen. Krashen's input hypothesis states that language—or vocabulary in particular—is essentially acquired by sufficient comprehensible input, and this is done through free and voluntary reading or listening. This hypothesis questions the pedagogical benefits of learning words one by one deliberately (i.e., the skill-building hypothesis). It also opposes the output hypothesis, which emphasizes the importance of language production in L2 learning (Swain 2005: 471). Krashen pointed to a range of studies showing that better writers read more outside of school; he also mentioned studies showing how children who are subjects to comprehensible input outside of school have greater competence in vocabulary and spelling, and how children who grow up in print-rich environments have better vocabularies (Miller 1977; Greany 1982; Rice 1986; Anderson, Wilson and Fielding 1988; in Krashen 1989: 441). According to Krashen, the human mind holds a 'language acquisition device', through which language is subconsciously acquired—

hence the process of acquisition. Krashen distinguishes between *acquisition* and *learning*, the latter of which he regards as a conscious (i.e., intentional, deliberate) process. Language learning goes on when mental faculties outside the language faculty are used, in which case 'only a limited amount of "language-like" competence can be developed' (Krashen 1989: 454). In a review of more recent research on academic L2 learning by Nagy and Townsend (2012), Krashen challenged the entire notion that anyone can consciously learn 'more than very modest amounts of academic language' (Krashen 2012: 234).

Krashen's learning vs. acquisition is described as a no-interface position, as in an absolute separation between the explicit and implicit knowledge in the learner's mind. This position has been heavily criticised by opponents adhering to the weak-interface position (e.g., Ellis 2008) or the strong-interface position (e.g. DeKeyser 1995; Pawlak 2014). Elgort (2011) believes that explicit learning may contribute more or less to implicit knowledge. In a recent study by Suzuki and DeKeyser (2017) on Japanese L2 learners of English, the authors found explicit strategies to be the more dominant learning route (as opposed to implicit strategies), 'which [helped] learners proceduralize and automatize linguistic knowledge, ultimately impacting the acquisition of implicit knowledge' (Suzuki and DeKeyser 2017: 778). Also, Elgort (2011: 399) provided findings that led her to conclude that deliberate learning is an effective way to learn L2 vocabulary and that the learning/acquisition-dichotomy suggested by Krashen is not justified.

There are several important differences between L1 and L2 vocabulary acquisition, and Nation (2013: 92) pointed out that it is important to keep these fields apart when addressing the teaching and learning methods. L1 language researchers question the value of deliberate vocabulary teaching for several reasons. First, by the time native speakers start an English course, they have already acquired so many words that vocabulary teaching would only marginally increase their vocabulary. Secondly, there is too much to learn about each word, so teaching would not cover all aspects regardless. Thirdly, learning a word requires a lot of time; according to some estimates, learners need to spend a minimum of 15 minutes per word. Fourth, incidental learning requires less effort and time from the teacher and is therefore a better alternative. Lastly, a number of studies have revealed uncertainty about the effect of vocabulary teaching.

However, there are several reasons why deliberate vocabulary teaching may prove particularly effective for L2 vocabulary learning, according to Nation (2013: 94). L2 learners are typically presented with far smaller samples of language than native speakers are, and the contexts in which the language is learned are often less helpful (Lightbown and Spada 2013: 68). Therefore, they often lack knowledge of everyday life vocabulary, which native speakers have accumulated through interactions with family and peers since childhood. Since this kind of vocabulary consists of a relatively small number of words, vocabulary teaching is often considered suitable. Another reason is that L2 learners need to bridge the gap between their proficiency level and the proficiency level needed to understand words in an 'unsimplified' input that they meet outside the L2 classroom (Nation 2013: 94).

2.1.6 Vocabulary acquisition strategies

Research on vocabulary teaching techniques and vocabulary learning activities have been lacking for many years, as pointed out by Nation (2013: 101). However, in recent years, there has been considerable research on this field, and according to Nation and Webb (2011: 3), the involvement load hypothesis (Laufer and Hulstijn 2001) is the best known and best researched way of analysing vocabulary teaching techniques.

According to Nation (2013: 100), the involvement load hypothesis builds upon the levels of processing theory, which is presented in section 2.4.2. Common to both theories is the notion that learning requires cognitive effort—also regarded as 'involvement'. The involvement load hypothesis specifies three important factors that are required for involvement—namely *need*, *search*, and *evaluation*—each of which may be either absent (-), moderately present (+), or strongly present (++) (Nation and Webb 2011: 3). These factors are shown in Table 2.4.

Table 2. 4: Contributing factors for involvement according to the Involvement Load Hypothesis

	Absent (-)	Moderately present (+)	Strongly present (++)
Need	Target vocabulary is not regarded as necessary to complete task	Target vocabulary is regarded as necessary to complete task	In addition to the vocabulary being regarded as necessary to complete a task, the learner feels the need to know the target vocabulary
Search	Form and meaning of target vocabulary are supplied as part of task	Learner must search for the meaning of the target item (involves 'receptive' skills)	Learner must search for the form of the target item (involves 'productive' skills)
Evaluation	Learner does not have to decide whether word choice is appropriate (e.g., the teacher evaluates word choice)	Learner decides whether word choice is appropriate when context is provided	Learner decides whether word choice is appropriate and provides a suitable context for the word

Note: Adapted from Nation (2013: 98)

Need is a motivational factor, which is present if the learner regards the target vocabulary as necessary either to complete a task or to expand his or her own vocabulary. The second factor, *search*, is present if the learner has to look up the meaning and/or the form of the target vocabulary. The third factor, *evaluation*, is present if the learner has to evaluate the appropriateness of the word choice either with or without a given context.

Some studies have indicated that the learner's effort or involvement is not always the most important variable. In a study by Hummel (2010), using a pre-test and a post-test, French ESL-learners were given a recall task before (pre-test) and after (post-test)

two kinds of treatment, namely active translation and rote-copying.[1] Hummel assumed *a priori* that the participants who engaged in active translation would perform better than those who engaged in rote-copying due to the more extensive, elaborated processing involved in the former task. However, the results showed that while all participants performed better in the post-test than they did in the pre-test, those who engaged in rote-copying performed significantly better than the active translation group. According to Hummel, one possible explanation for this result might be that the translation task preoccupied the 'processing resources' needed for the memorisation of vocabulary.

Hummel's (2010) study indicates that the 'less engaging', simpler vocabulary exercises may in some cases prove more beneficial for vocabulary retention than the vocabulary exercises that were more engaging and complex. At least, this applies when the time span is short. With reference to these findings, we can speculate as to whether the involvement load hypothesis primarily applies to long-term learning rather than short-term learning In that case, we might have to adjust for potential learning restraints from cognitive processing 'overload' when designing learning activities in accordance with the involvement load hypothesis.

2.1.7 Vocabulary and learning burden

The acquisition of vocabulary depends on a number of factors, including the lexical properties of words, e.g., imageability and idiomaticity, and the similarity in form and meaning.

One study has shown that if a concept is easy to *imagine* and to *experience* by the senses, learners will likely also find it easy to learn the *words* for the concept (Schmitt 2015: 53). These two factors often correlate, but not always. As a case in point, a study by De Groot (2006: 473; in Schmitt 2015: 53) suggested that abstract concepts such as *anxiety* and *jealousy* have proved easier to imagine than some concrete but infrequent words, such as *armadillo* and *encephalon*. This might be because experiencing feelings of anxiety and jealousy is more common than experiencing armadillos or observing an encephalon.

---

[1] Hummel (2010: 65) explained 'rote-copying' in this study as 'exposure and copy', and pointed out that 'participants were not required to engage in an active translation process, although they were exposed to sentences and their translation equivalents.'

Some words or expressions have literal meanings while others have idiomatic meaning, as can be seen in the two verb phrases 'carry out garbage' and 'carry out a task'. Phrases with literal meanings, such as the former, are easier to understand since they often deal with more concrete matters (Schmitt 2015: 53). However, according to Schmitt (2015: 53), research has indicated that phrases with idiomatic meanings are in fact more frequently used (Conklin and Schmitt 2008; in Schmitt 2015: 53). This shows that even though idiomatic phrases are less straightforward, their frequent use might suggest that they are easier to learn.

Similarity in form or meaning is another aspect of language that affects learning burden. Some words and expressions are synonymous. This means that they are different in *form* but similar in *meaning*. For example, *expire*, *pass out*, *bite the dust*, and *kick the bucket* all mean *die* in a given context (Schmitt 2015: 49–50). Conversely, polysemy refers to cases where *multiple* meanings might be expressed by *one* form. For instance, the word *carry* means two different things in the two example verb phrases in the previous paragraph. Schmitt (2015: 50–51) points out that the vocabulary used in the English language—as opposed to that of other languages—has a large number of inconsistent form-meaning relationships; this might be due to the language's historical development. English was originally a Germanic language, but it retained a great number of words from French, Latin, and Old Norse over the course of many centuries. This lack of formal similarity between semantically-related words makes the English vocabulary more difficult to learn compared to vocabulary from languages with more transparent formal relationships (Schmitt 2015: 51).

The learning burden also depends on whether words have content or not. *Content words*, such as 'house', 'throw', or 'warm', are considered to have substantial meaning even out of context, while *function words*, such as 'the', 'in', or 'of' have largely a grammatical role and little or no independent meaning (Singleton 1999: 11).

2.2 Chunks of language

This section covers the identification and the use of 'chunks of language', often referred to simply as chunks. The study examines the term by comparing it to related terms and discusses the distinction between different kinds of *word strings*— some of which are chunks and some of which are not. Furthermore, this section presents different ways to classify chunks based on function, formal features, and transparency. The section also

investigates the spread and the use of chunks by different groups of society. Finally, the section considers the relevance of chunks for L2 learners and how the acquisition of word sequences differs from acquisition of words.

2.2.1 Definition of chunks and related terms

Chunks is another term for *formulaic sequences*, which is defined as 'multiple-word strings that behave as single units, e.g., realizing a single meaning or function' (Alali and Schmitt 2012: 153). The term chunks is used by many researchers in recent works (Lindstromberg and Boers 2008a; 2008b; Boers et al. 2010a; Davis and Kryszewska 2012), but there is a range of closely related terms, including formulaic sequences (Jones and Haywood 2004; Read and Nation 2004; Schmitt et al. 2004a; Schmitt and Underwood 2004), fixed expressions (Moon 1998), lexical phrases (Schmitt and Carter 2000), prefabs (Erman and Warren 2000), multi-word units or collocations (Schmitt 2015: 119). Research on chunks focuses on different aspects. In many cases, the emphasised aspects determine which term is used. For instance, in research which focuses on the relationship between two-word pairs, the term *collocations* is most frequently used. By contrast, in research on holistic storage of forms (Schmitt 2015: 119), it is more common to use terms such as *chunks, formulaic sequences* and *prefabricated expressions*

2.2.2 The ubiquity of chunks in language use

Many definitions of vocabulary do not specify the basic unit of vocabulary beyond its *word* concept. However, research has suggested that a huge part of the English language is made up of chunks. Erman and Warren (2000: 37), found that 58.6 percent of the words in spoken language and 52.3 percent of the words in written language are chunks or prefabs.. A study by Foster (2001; in Schmitt and Carter 2004: 1) provided more modest estimates, suggesting that 32 percent of all unplanned native speech is made up of chunks. These numbers show that chunks are not a peripheral feature. On the contrary, they are 'ubiquitous and a chore characteristic of language', as pointed out by Schmitt (2015: 117).

2.2.3 Identification of chunks

There are two main reasons why chunks, or prefabs can be difficult to identify. First, due to their emergence through a gradual process of conventionalisation,[2] some members of a language community might identify a chunk that is not recognised as a chunk by other member of the same language community. Secondly, chunks can be easily overlooked, because some of them appear to be transparent word combinations but may turn out to be idiomatic or non-compositional on closer scrutiny. According to Erman and Warren (2000: 33), chunks 'are probabilistic, some more than others'. In other words, there will always be some uncertainty related to the identification.

Computer software has contributed to a change in the way the term *chunks* is defined. Schmitt and Carter (2004: 2) and Schmitt (2015: 117) point out that idioms, proverbs, and sayings have long been recognized as chunks because of their 'non-compositional' nature. In other words, 'their meaning [can] not be derived from the sum of meanings of the component words' (Schmitt and Carter 2004: 2). However, with the recent technological advances in computerised methods in corpus research, collocations have come to be defined as chunks. A number of language analysis tools, such as Wordsmith, MonoConc Pro, or WMatrix (Schmitt 2015: 335–345), now enable the detection of collocative patterns of words. For example, by analysing concordance lines that include the words *stomach*, *bow*, and *gap*, we find the compound noun *stomach pain*, the binomial pair *bow and arrow*, and the function phrase *mind the gap*.

The constituent words of chunks co-occur fairly frequently in native speakers' language use. One way of identifying chunks is therefore to study large text collections of written or spoken language discourse, and to detect *recurrence*. Schmitt, Grandage, et al. (2004b: 128) use the term *recurrent clusters* to refer to word strings that occur frequently together in a sufficiently large corpus. . Recurrent clusters are often chunks but not always. For example, we can see from a collocates search in the British Nation Corpus that *commit suicide* is a frequently used word string due to the word *commit* occurring most frequently in conjunction with *suicide*. In this case, the word string is both a recurrent cluster and a chunk, since according to the Oxford Collocations dictionary,

---

[2] Conventionalisation is the process towards conventionality of a linguistic unit, which relates to 'the idea that linguistic expressions become part of the grammar of a language by virtue of being shared among members of a speech community' Evans (2007: 21).

commit suicide is regarded a well-established verb-noun collocation, regardless of its frequency (McIntosh et al. 2009).

However, according to Schmitt, Grandage, et al. (2004b: 128) we cannot with certainty identify chunks on the basis of corpus research. This is because many recurrent clusters, such as *it is the*, *if you,* or *of it*, would probably not be called chunks by the human mind. Wray (2002) elaborated on these aspects of chunks (i.e., 'formulaic sequences') in the following definition of the term, calling it

> […] a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.
>
> (Wray 2002: 9)

This definition explains how we recognise chunks and how we use them in language. Schmitt et al. (2004b) shared this understanding, defining the term *chunks* as units of words which are 'stored holistically' (Schmitt et al. 2004b: 128) in the mind. In other words, while recurrent clusters of words are simply word strings which co-occur frequently in a corpus, chunks are word strings that are stored as entire units and therefore processed and used in a different way. Another definition by Wray (2008) in a more recent work, adds to the understanding of chunks (also here in terms of 'formulaic sequences') and describes the chunk as a 'morpheme equivalent unit' that is seen as

> […] a word or word string, whether incomplete or including gaps for inserted variable items, that is processed like a morpheme, that is, without recourse to any form-meaning matching of any subparts it may have.
>
> (Wray 2008: 12).

Here, chunks are defined in terms of their functional equivalence to other basic components of vocabulary.

2.2.4 Classification of chunks

Boers and Lindstromberg (2009: 9–12) examine different aspects of chunks, and they suggest that we classify chunks on the basis of their function, formal features, and degree of transparency.

When it comes to function, chunks may be used as social routine fillers, conversational fillers, interactional sentences heads, situation evaluators or discourse organizers, and they sometimes have referential or message-oriented function. Some of these functions are more useful to L2 learners than others. For example, social routine fillers can be useful to 'fit in' with native speakers, and discourse organisers can be useful in academic text composition.

We can also define chunks by way of their *formal* features. Examples by Lindstromberg and Boers (2008a: 8) are sentence heads (e.g., '*Could you… ?*' or '*Would you… ?*'), phrasal verbs (e.g., *break down* or *wipe out*), compounds (e.g., *credit card* or *weather forecast*), strong collocations (e.g., *tell a story* or *stark naked*), and grammatical frames (e.g., '*as … as …*' and '*the -er the -er*'). Another example of formal classification is provided by Gibbs's (1994, in Gibbs 2007: 698–699), who presents a 'rough list of [...] different forms of idioms and formulaic language':

(1) Sayings: a. take the bull by the horns b. let the cat out of the bag
(2) Proverbs: a. A bird in the hand is worth two in the bush. b. A stitch in time saves nine.
(3) Phrasal verbs: a. to give in b. to take off
(4) Idioms: a. kick the bucket b. to crack the whip
(5) Binomials: a. spick and span b. hammer and tongs
(6) Frozen similes: a. as white as snow b. as cool as a cucumber
(7) Phrasal compounds: a. red herring b. dead-line
(8) Incorporating verb idioms: a. to babysit b. to sightsee
(9) Formulaic expressions: a. at first sight b. how do you do?

(Gibbs 2007: 698–699)

Simpson and Mendis (2003) report that several studies (Nattinger & DeCarrico, 1992; Moon, 1998; Wray, 1999, 2000, 2002; Wray & Perkins, 2000) consider idioms as 'one subcategory of the more general lexical phenomenon of formulaic language'. This is in congruence with the list above. However, in many studies of chunks, the focus is instead on the distinction between *pure idioms* and *open collocations*. This dichotomy makes sense if we see chunks in terms of the two variables transparency and compositionality, which leads us to the final classification-method outlined by Boers and Lindstromberg (2009).

The third way to classify chunks is by locating them on a continuum from the opaque to the transparent. Ebeling and Ebeling (2013: 2) explain that chunks range from 'opaque, non-compositional idioms [to] fully transparent, compositional, sequences of words'. Thus, chunks can be classified within the range from opaque to transparent and

from non-compositional to compositional. According to Schmitt (2015), idioms are typically opaque and non-compositional, and they can be defined as 'semantically opaque in the sense that their meaning is figurative and not predictable from the literal meanings of its constituents' (Schmitt 2015: 120). Collocations, on the other hand, can be defined as 'any sequence of words that is frequently found in the language in a relatively fixed form and [that] merits the learners' attention because of its semantic unpredictability' (Malec 2010: 129). With regards to the two variables mentioned above, collocations are transparent and compositional, while idioms are opaque and non-compositional. However, the difference between idioms and collocations is far from clear-cut, because many collocations are non-transparent and thus similar to idioms. For example, 'heavy smoker' and 'criminal lawyer' are two collocations that cannot be understood in terms of their *literal* meaning, which would define them as 'overweight nicotine-user' and 'law-breaking attorney' respectively. They are characterised by, what Singleton calls 'peculiar semantics' (2000; in Malec 2010: 128).

Despite these issues in differentiating collocations and idioms, chunks can be categorised in a scale ranging from *pure idioms* to *open collocations*, as in the continuum of collocability (cf. Table 2.5).

Table 2. 5: The continuum of collocability

Pure idioms	Figurative idioms	Restricted collocations	Open collocations
'kick the bucket'	'kick your heels'	'kick the habit'	'kick the ball'

Note: Adapted from Malec (2010: 129)

Sinclair (1987; in Moon 1998: 128) argues that language use has two underpinning principles: The *open choice principle* and the *idiom principle*. The open choice principle, which is the traditional view, understands virtually *all* language as formed by word-by-word construction according to language rules. By contrast, the idiom principle states that language is also formed by the use of preconstructed multi-word units as 'prefabricated routines'.

2.2.5 The importance of chunks to L2 vocabulary learners

Numerous studies have concluded that the knowledge of chunks is equally important or even more important for language use than the knowledge of *single* words

(Pawley and Syder 1983, Jackendoff 1995 and Mel'čuk 1995; in Schmitt and Carter 2004: 1). According to Wray (2004: 249), the prefabrication of word strings might be experienced as particularly empowering by L2 learners, because they ease the expression of otherwise complex information. L2 learners are regularly faced with the challenge of balancing speed, fluency and accuracy. This balance might be perceived as trivial to native speakers. For L2 learners, on the other hand, conversations around the dinner table or asking for directions in unfamiliar places might represent significant challenges. Chunks have the potential to significantly ease communication in such situations.

The functions of chunks become evident when we examine more closely why it is that certain professions or native-speaking groups in society use chunks more often than others. According to Kuiper and Haggo (1984; in Schmitt and Carter 2004: 5), 'smooth-talkers' such as sportscasters and auctioneers are examples of such groups. Individuals in these professions depend on the ability to convey information in a short amount of time, and the use of chunks makes this process more efficient. This is because memorised strings of words can be retrieved more quickly and easily than the same strings of words made up creatively. In other words, chunks are cognitively more efficient than creatively-generated language. Schmitt and Carter (2004: 5) found evidence of this in functional language usage in corpus: 'For example, when shifting a topic, we commonly use a formulaic sequence like *by the way*, but create novel phrases like *It's time for a topic change* much more rarely' [authors' emphasis]. The use of chunks also tends to increase *accuracy* in communication. Hence, as Wray (2004) points out, chunks might reduce the risk of 'instilling [the] message with inappropriate pragmatic overtones' (Wray 2004: 249).n addition to the productive advantages for using chunks, there are also receptive advantages. A study on eye movement patterns by Underwood, Schmitt and Galpin (2004) suggested that both native and non-native speakers of English process words in chunks faster than words in nonformulaic contexts.

2.3 Academic vocabulary

The term *academic vocabulary* can be defined as words and expressions that are 'common in the English academic register' (Hyland and Tse 2007: 235). This section presents some definitions of the term; it also discusses the contexts in which the term is used and examines some critical views to the notion of academic vocabulary. The section accounts for the study of corpora of academic written and spoken language and the construction of

academic vocabulary lists that is based on corpus research. The section also discusses chunks as part of academic vocabulary, and the genre-specific approach as a pathway to learning chunks.

2.3.1 Definition of academic vocabulary

Academic vocabulary is also known as 'subtechnical vocabulary', 'semitechnical vocabulary', or 'specialized non-technical lexis', according to Hyland and Tse (2007: 235). Additionally, Baumann and Graves (2010: 4) list a range of related terms which might have alternatively been used, such as 'academic background', 'content vocabulary' and 'academic language skills', and 'academic literacy', the last of which was defined by Moore (2008; in Baumann and Graves 2010: 5) as 'the reading and writing used in school contexts'. Baumann and Graves (2010: 5) also list several definitions for academic vocabulary that relate the term to the school context.

Unlike general vocabulary, academic vocabulary is restricted to those words and expressions that are useful for *academic* undertakings. Notably, however, academic vocabulary, does not include technical words and expressions. Thus, many low-frequency words and expressions in engineering and other technical areas of expertise are not categorised as units of academic vocabulary.

Academic vocabulary includes words and expressions that appear across a range of academic disciplines. The pedagogical value of academic vocabulary lists is demonstrated by efficiency in vocabulary acquisition. Instead of having to interact with vocabulary in each academic discipline, students can interact with one kind of vocabulary as preparation for studies within *several* academic disciplines. Moreover, academic vocabulary is a 'key element of essayist literacy' (Lillis 2001; in Hyland and Tse 2007: 235), and academic vocabulary is considered more advanced than most of the words and expressions students generally encounter in higher education (Hyland and Tse 2007: 235).

2.3.2 Academic vocabulary and L2 learners' proficiency

Academic vocabulary serves important supportive functions in academic text comprehension and is thus necessary for students in higher education to acquire it. The acquisition of academic vocabulary is regarded challenging for new students. Yet, academic vocabulary is often omitted from glossaries. It is also frequently used without explanation by lecturers (Hyland and Tse 2007: 236). Therefore, the development of new

techniques and strategies for teaching and learning academic vocabulary in higher education, and preparatory stages such as upper secondary school is therefore crucial.

Language proficiency in academic vocabulary and general vocabulary are developed differently in language learners. This is suggested by research on basic interpersonal communicative skills (BICS) and cognitive/academic language proficiency (CALP), which are two terms that were first introduced by Cummins (1980). BICS refers to conversational fluency in a language, while CALP refers to students' abilities to 'understand and express […] concepts and ideas that are relevant to success in school' (Cummins 2008: 491). CALP is considered the result of both 'language proficiency' and 'cognitive and memory skills' and is considered the major determinant of educational progress, as shown in Figure 2.1 (cf. Cummins 1980).

Figure 2.1: Relationship of CALP to language proficiency, cognitive and memory skill and educational progress
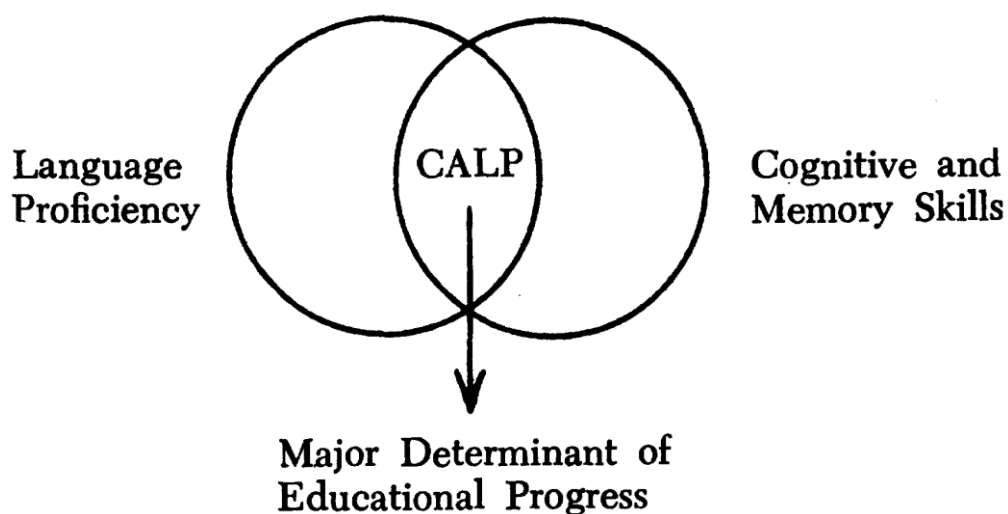
Note: Adapted from Cummins (1980: 178)

Cummins (1980; 2008) claims that the two skill types of BICS and CALP can be distinguished empirically. This distinction is imperative for teachers to understand in order to avoid misjudgement of language abilities, particularly among L2 learners. While conversational fluency (BICS) is commonly developed by L2 students after one to three years of living in an English-speaking environment, the development of CALP may take five to seven years. Moreover, BICS is developed faster in *younger* L2 students, while CALP is developed faster in *older* L2 students. Vincent (1996) discusses some of the

hurdles in teachers' and administrators' judgement of L2-learners language abilities. In a study of second-generation Salvadorian students in Washington DC, she found that teachers were often deceived by the native-appearing abilities of L2 students who had lived in the country for only two to three years. These L2 students had attained *conversational* fluency. They were able to speak without an accent, and to converse with their peers about day-to-day activities and about frequently discussed topics. However, their academic language was lacking. This caused many teachers to overlook the ESL-background of the students. The students were instead believed to be 'slow learners' because of their poor performances in school assignments, and they were frequently enrolled into special education classes. In many of these cases, the students were simply following a normal progression of development, but their academic language proficiency did not match their level of conversational fluency (BICS).

These observations are also relevant to Norwegian students. Norwegian ESL-students seem to perform well in tests of general English vocabulary. Norway ranks 4th worldwide on Education First English Proficiency Index[3] (Education First 2018). However, recent research suggests that Norwegian students struggle with academic language proficiency. In a study by Hellekjær (2005), two-thirds of the students in a sample of 178 upper secondary level respondents achieved IELTS scored below the Band 6 level, which is the required minimum for admission to British and Australian universities. Moreover, one third of the students in a university level sample in the same study struggled with the English literature on their reading lists. These findings suggest that a greater focus on English for academic purposes (EAP) and on CALP in English language teaching is needed.

2.3.3 Corpus-based lists of academic words and phrases

Corpus research is of great value to the study of lexis and vocabulary. Linguistic research that is based solely on intuition might cause inadequate descriptions and conclusions, due to imprecisions and inaccuracies in judgement (Timmis 2015: 22). One weakness with linguistic research based on intuition is that unusual words and structures are frequently noticed, while ordinary ones tend to be overlooked (Krishnamurthy 2000; in Timmis 2015: 22).

---

[3] According to the official website of Education First (2018), the report surveys 88 countries where English is not the populations' first language.

Corpus research has also played a major role in 'discovering' the role of 'units beyond the word' in the lexicon, since the ubiquity and the importance of chunks has become visible in the analyses of large corpora during recent decades. These new discoveries have induced a 'reappraisal of lexis' as well as a renewed belief in the importance of vocabulary in language acquisition, according to Timmis (2015: 22–23). While grammar was previously considered the generative aspect of language that should be prioritised in language learning, some researchers now consider grammar as 'playing only a subsidiary or supportive role in communication' (Timmis 2015: 23)[4].

Research on corpus-based vocabulary lists has been conducted for over a hundred years within the field of linguistics, according to Martinez and Schmitt (2015: 440). One important contribution to the field is the General Service List (GSL) by Michael West (1953), which consists of 2000 words, and it has played a central part in vocabulary research up until recently (Martinez and Schmitt 2015: 442).

In 2000, Avril Coxhead published the Academic Word List (AWL) (Coxhead 2000). The development of this list was motivated by the need to pedagogically identify useful words for students who use English in higher education. Coxhead was the first to use a corpus collected in electronic form, and this corpus contains as many as 3.5 million words. More recently, the new General Service List (new GSL) (Brezina and Gablasova 2015) and the New Academic Vocabulary List (AVL) (Gardner and Davies 2014) were introduced. The newer lists (i.e., AVL, new GSL, and AWL) are based on much larger corpora than the old ones. They also ensure a higher replicability than the old GSL, which 'excluded important lexical items solely based on "a priory" assumptions' (Martinez and Schmitt 2015).

Unlike the AWL, both the New GSL and the AVL count lemmas instead of word families (i.e., words and their inflections instead of words and their derivations). They also account for the items' dispersion in the corpora—that is, the diversity of corpus files in which the words appear. Finally, the core high-frequency 3000 words are extracted from a far bigger corpus, namely from the Lancaster–Oslo/Bergen Corpus (1 million words), the British National Corpus (100 million words), the BE06 Corpus of British

---

[4] It is important to mention here that several recent studies have demonstrated the importance of grammar in the ESL-classroom as well; see e.g. Myhill, Jones et al. (2013: 90) on the pedagogical importance of teachers' grammatical knowledge, Batstone and Ellis' (2009) principled grammar teaching, and Pawlak's (2014) work on error correction in the EFL classroom.

English (1 million words), and the EnTenTen (12 billion web-gathered words). Gardner and Davies argued that the use of lemmas makes the list more manageable; moreover, it avoids the assumption that knowledge of a baseword can necessarily be extended to derived forms. As an example, in the AWL the word 'proceeds' is included in the word family of 'proceed'. Thus, it is not possible to tell whether the verb (where proceeds means *continues*) or the plural noun (where proceeds means *profits*) is counted. Moreover, when counting lemmas, the noun *proceedings* would rightly be counted on its own, and the words *procedure* and *procedures* would rightly be counted in one group (Gardner and Davies 2014).

However, the decision to count lemmas also has its shortcomings (Martinez and Schmitt 2015). The reason for this is that the method used to count lemmas (i.e., an 'automated part-of-speech tagging') has significant error rates and sometimes fails to assign correct word classes. More importantly, many chunks, such as 'as well as', 'as a result', or 'in result' are often falsely registered as instances of single words. Since these chunks make up huge parts of the English vocabulary, this problem causes significant inaccuracies. As an answer to these shortcomings, Martinez and Schmitt (2015: 448–450) proposed the idea of counting lexemes instead of lemmas, and they offered an important contribution with the PHRASE List (Martinez and Schmitt 2015).

In addition to the general and academic vocabulary lists mentioned above, a widely used category of vocabulary is technical vocabulary. Nation and Kyongho (1995: 35–37) differentiated between general service vocabulary, special purposes vocabulary, and technical and low-frequency vocabulary. Technical vocabulary occurs with a significant level of frequency within a few texts or maybe just within one text. The authors mentioned *isocost*, *utility*, and *duopoly* as examples of technical vocabulary that typically occurs only in economic texts. Hyland and Tse (2007: 236) estimated that technical vocabulary 'differs by subject and covers up to 5% of texts' as opposed to high-frequency words (80% of most text) and academic vocabulary (8–10% of running words in academic texts).

2.3.4 Idioms and chunks in academic discourse

As shown in the previous section, lists of academic vocabulary comprise not only words but also word strings of more narrow categories (e.g., collocations or idioms) and wider categories (e.g., formulas or chunks). Chunks are an important part of academic

vocabulary. Recent studies have shown that the use of chunks in academic speech includes not only transparent and compositional phrases, such as 'Studies show that' or 'In conclusion', but also idioms, which are often non-compositional and opaque (cf. section 2.2.4). A study by Simpson and Mendis (2003) examined a number of variables related to the frequency of idioms in academic speech. These include the frequency of idioms occurring in monologic vs. interactive speech and those in various subregisters within academic spoken language (from different academic fields); they also looked at what functions idioms perform in academic speech. The findings suggested that idioms are not at all rare in academic speech. Moreover, the use of idioms is evenly distributed in monologic and interactive speech and in the different subregisters. These findings confuted the researchers' own assumptions before the study; they had assumed that the use of idioms is more prevalent in interactive speech than in monologic speech and more prevalent in the humanities and the social sciences than in the hard sciences. The study also showed that idioms fulfil important functions of academic speech, such as emphasis (e.g., 'carrot and stick', to emphasize reward and punishment-mechanisms), paraphrase (e.g., 'dime a dozen', referring to plentiful occurrences) and metalanguage (e.g., 'cut to the chase', as in finishing the teaching session).

According to Simpson and Mendis (2003: 432), the evidence from these findings provides a rationale for including idioms and formulaic language in curricula for EAP. Furthermore, they proposed that corpora containing genre-specific vocabulary offer a valuable pathway to learning chunks. While earlier research on idioms emphasized formal properties and often applied typologies based on semantic and syntactic criteria, the current approach emphasizes the pragmatic, interactional, and discourse-specific features of idioms (Fernando, 1996; McCarthy, 1998; Moon, 1998; in Simpson and Mendis 2003: 421). Building upon this idea, the current approach holds that chunks are not used regardless of text genre and that they must be viewed as communicative devices rather than 'quirks of language'. Chunks have socio-interactional functions and relate to specific domains or institutions. In line with this view, Simpson and Mendis (2003) advocated the use of specialized corpora with specific pedagogical aims, since each formulaic sequence belongs to a certain discourse and should be taught and learned within this context. The

use of a specialized corpus, such as the one they examine (MICASE[5]), provides a rich resource for teaching materials since it offers authentic examples of the use of idioms and relieves the teachers from 'the need to create contrived contexts for idioms and teach them as disembodied items' (Simpson and Mendis 2003: 437–438).

The findings from Simpson and Mendis' (2003) study indicated that all types of chunks—including idioms—are important in academic speech and writing, and that textbooks in the field of EAP should consequently contribute to the teaching and learning of chunks. However, studies suggest that textbooks used in EAP-courses do not cover chunks adequately. Jones and Haywood (2004: 270) addressed several issues in a review of four textbooks that were used in courses for EAP at the time of the study. First, there are too few examples in the books, and the examples given are often decontextualized. Furthermore, single words and chunks are often mixed together, and in doing so the phraseological nature of the language may not be obvious to the student. There is also no information about the frequency with which the chunks occur. Thus, students might use chunks that are rare in academic prose and end up learning vocabulary that is less useful. Moreover, there is a lack of exploratory tasks and learning strategies to help learners understand how to use the phrases. Finally, the textbooks do not teach learning strategies for the acquisition of chunks.

Jones and Haywood (2004: 277) proposed several routes to facilitate learning of academic vocabulary as an answer to these shortcomings. First, they suggested that students study academic chunks and foster deep-processing through activities such as classifying chunks into meaning-based groups, analysing and classifying them according to their structure, and comparing academic text to less formally written text. Second, Jones and Haywood (2004) argued for the use of concordance lines and corpus extracts since this allows for the study of different uses of chunks in different contexts.

2.4 Cognitive Linguistic approaches to vocabulary and implications for teaching

This section introduces some of the basic characteristics of Cognitive Linguistics and compares Cognitive Linguistics to other related approaches. It also deals with the central

---

[5] MICASE is short for the Michigan Corpus of Academic Spoken English, a specialized corpus of 1.7 million words of academic discourse (Simpson and Mendis (2003: 419).

concept of linguistic motivation in Cognitive Linguistics and presents cognitive processing theories. The next two subsections explore the semantic aspect and the structural aspect of linguistic motivation, respectively. In particular, the study introduces conceptual metaphor as a backdrop for semantic elaboration and discusses how various phonological features, such as rhyme and alliteration, give grounds for structural elaboration. Finally, the section presents theory on how to organize CL-inspired teaching.

My account of Cognitive Linguistics and conceptual metaphor theory is based on a narrow selection of relevant introductory texts from the field; this is limited to theory that is useful for the kind of classroom experiment I present in this thesis. Consequently, I focus on concepts that are important for explaining the theory behind teaching formulaic language to L2 learners of English.

2.4.1 Definition of Cognitive Linguistics

Cognitive Linguistics originated in the late 1970s and early 1980s as a result of pioneering work by a number of scholars such as George Lakoff, Ronald Langacker, and Len Talmy (Croft and Cruse 2004; Evans and Green 2006; Geeraerts 2006; Geeraerts and Cuyckens 2010). The term 'cognitive' can be defined as 'of, relating to, being, or involving conscious intellectual activity (such as thinking, reasoning, or remembering)' (Merriam Webster 2018). Geeraerts and Cuyckens explained the 'cognitive aspect' of Cognitive Linguistics in the following manner: 'Cognitive Linguistics is the study of language in its cognitive function, where *cognitive* refers to the crucial role of intermediate informational structures in our encounters with the world.'

Cognitive Linguistics is not a single unified approach, but rather it is a 'cluster of broadly compatible approaches' (Geeraerts and Cuyckens 2010: 2). However, there are several fundamental assumptions that distinguish Cognitive Linguistics from other approaches. A comparison between Cognitive Linguistics and Generative Grammar— a related linguistic approach—unveils some of the basic characteristics of Cognitive Linguistics. As a starting point, both approaches hold that there can be no knowledge without an existence of a mental representation that functions as a mediating mechanism between the epistemological subject and the object. However, while Generative Grammar holds that language is an autonomous faculty in the human mind and sees linguistic representation as *one* type of cognitive ability, Cognitive Linguistics holds that the cognitive abilities governing language use do not differ in principle from other cognitive

abilities. Cognitive Linguistics claim that 'representation of linguistic knowledge is essentially the same as the representation of other conceptual structures' (Croft and Cruse 2004: 2). Consequently, proponents of Generative Grammar are interested in knowledge *of* the language, whereas proponents of Cognitive Linguistics are interested in knowledge *through* language.

A consequence of the Cognitive Linguistic stand is that meaning is regarded as the primary linguistic phenomenon, a principle Geeraerts and Cuyckens (2010: 4) referred to as 'the primacy of semantics in linguistic analysis'. This is evident in Cognitive Linguistic vocabulary research (e.g., metaphor theory) where the emphasis is on the *concepts* in addition to the linguistic units that express these concepts. However, cognitive linguists do not only view semantic representation as basically conceptual; they consider syntactic, morphological and phonological representations to be so as well, since they must be comprehended and produced. Thus, Cognitive Linguistic research examines both the *semantic* and *structural* features of language.

Another fundamental characteristic of Cognitive Linguistics is the view that 'the categorial nature of language imposes structures onto the human perception of the world' (Geeraerts and Cuyckens 2010: 4)—that is, language is not an objective reflection of the world. Geeraerts and Cuyckens (2010: 4) referred to this aspect as the 'perspectival nature of linguistic meaning'. As a consequence of this, much Cognitive Linguistic research deals with the way language shapes our mindset by linking features from one knowledge domain to another; this is done by the use of figurative language through mechanisms as metaphor and metonymy.

2.4.2 The concept of linguistic motivation

One central concept in Cognitive Linguistics is that of linguistic motivation or non-arbitrariness. Unlike the traditional linguistic approach, which sees the formation of symbolic units such as idioms and other chunks as an arbitrary process, Cognitive Linguistics regards this as a motivated process (Kövecses and Benczes 2010: 86). Linguistic motivation can be explained or 'explicated'—a term that is frequently used in this context (see Taylor 2003; Boers and Lindstromberg 2008b; Langacker 2011)—from the perspective of either the semantic or the phonological pole of lexis, or both. On the semantic pole, the use of various kinds of conceptual metaphors can be explained by certain forms of motivation. For instance, correlation metaphors are motivated by

'correlations in experience', while resemblance metaphors are motivated by 'perceived similarity' and GENERIC-IS-SPECIFIC metaphors by 'perceived structural similarity' (Kövecses and Benczes 2010: 86). On the phonological pole, features such as alliteration, rhyme, and assonance motivate the collocational patterns of many chunks, as explained by Boers and Lindstromberg (2008c: 330): '[…] it appears that in the process of standardising word combinations, euphonious word strings are preferred over same-meaning but non-euphonious word strings. For example, alliteration and assonance help motivate the precise lexical selection in a large number of [chunks]'.

The Cognitive Linguistic view holds that linguistic motivation is present in all languages, because the cognitive mechanisms that underlie semantic and structural motivational features are determined by human universal experiences. There are however differences in how these features manifest themselves across different languages. Some of the differences in conceptual metaphors in different cultures are discussed in section 2.4.4.

Thus far, we have distinguished between the traditional view of arbitrariness and the Cognitive Linguistic view of linguistic motivation. At this point, it is important to note that the Cognitive Linguistic view only holds that language formation can be *explained* (by various semantic or phonological features). It does not hold that it can be *predicted*—that is, the ability to anticipate the formation of a certain metaphor or a certain formulaic sequence. Kövecses and Benczes (2010: 77–78) explained that the notion of predictability characterizes formal theories of language such as Generative Grammar which model themselves after 'exact science': 'In this view, which metaphors we have should be predictable, and if our theory can't predict them, the theory can be claimed to be unscientific' (Kövecses and Benczes 2010: 77). Cognitive Linguistics does not accept this view, and instead it replaces the notion of predictability with the notion of motivation. Boers et al. (2004) shed light on the nuances between arbitrariness, motivation, and prediction in the following explanation of why figurative idioms have become a more appealing target for pedagogical methods due to recognition of the notion of linguistic motivation:

> Studies in cognitive semantics (e.g. Kövecses 1990, Lakoff 1987) have revealed that many figurative expressions (including idioms) are in fact 'motivated' rather than arbitrary. While it is true that the figurative meaning of many idioms may not be fully predictable from their constituent parts, it is nonetheless often possible (in retrospect) to explain how and why that figurative meaning has arisen.

2.4.3 Conceptual metaphor theory

The initial research on vocabulary in Cognitive Linguitics took place in the late 1970s and early 1980s, and it concerned the study of conceptual metaphor theory (CMT), as outlined by George Lakoff and Mark Johnson in their work *Metaphors we live by* (1980). The traditional view of metaphor regards metaphor as a property of language, a literary device typically used by talented writers to 'ornament' a text, while the Cognitive Linguistic (CL) approach considers metaphor to be pervasive in everyday life and a necessary ingredient in human thought and communication. Deignan (2005: 4) refers to former as the 'decorative view' and the latter as 'metaphor-as-thought'.

The CL approach to metaphor examines the way abstract concepts of life are explained in terms of more concrete concepts. Metaphor is viewed as a way of understanding one conceptual domain in terms of another conceptual domain—that is, the *target domain* is being understood in terms of the *source domain.* According to convention, conceptual metaphors (CMs) are written with capital letters in the format TARGET DOMAIN IS SOURCE DOMAIN. We can distinguish between CMs and the linguistic metaphorical expressions (LMEs) in which they appear. The conceptual metaphor LIFE IS A JOURNEY is expressed by many linguistic metaphorical expressions, such as *She's gone through a lot in life* and *He's without direction in life* [examples given by Kövecses and Benczes (2010: 3–4)]. As seen in these examples, linguistic metaphorical expressions originate from the language or terminology from the source domain. Likewise, the linguistic metaphorical expression *She defended her views well* holds the conceptual metaphor ARGUMENT IS WAR. When talking about an argument as war, as in this case, the source domain facilitates the target domain; in other words, the elements from the domain WAR are mapped onto the target ARGUMENT. Source domains typically feature concrete matters (e.g., constructions, machines, armed conflicts, the human body, organisms), while target domains often feature abstract entities (e.g., theories, life, love, relationships). Domains can be categorized into different hierarchical levels or groups; for instance, 'house' is a subgroup under 'constructions', and 'constructions' is a subgroup under inanimate objects.

Metaphor can be based on knowledge or an image, according to Kövecses and Benczes (2010: 42–44). LIFE IS A JOURNEY and ARGUMENT IS WAR are both examples of metaphors where conceptual elements of knowledge (i.e., travelling, travellers, destinations) are mapped onto a target. However, we also have metaphors where conceptual elements of *image-schemas* are mapped onto the target. Image-schemas structure our experience of our physical surroundings (Boers et al. 2004: 56). Examples of image-schemas include in–out, front–back, up–down, contact, motion and force (Kövecses and Benczes 2010: 43). In the conceptual metaphor MORE IS UP, which underlies linguistic metaphorical expressions such as *Her income is above average*, the image-schema UP–DOWN lends its structure to the abstract target domain MORE.

Conceptual metonymy is another cognitive mechanism closely related to conceptual metaphor. While conceptual metaphor concerns mapping from one domain to another, conceptual metonymy directs attention to one entity (i.e., the 'target entity') through another entity (i.e., the 'vehicle entity'), when both entities belong to the same domain (Kövecses and Benczes 2010: 173). For instance, in the expression *All hands on deck*, 'hands' stands for 'men'. This conceptual metonymy expresses A PART FOR THE WHOLE, or more specifically, AN OBJECT FOR A PERSON.

The CMs that are expressed in language differ across different cultures. According to Kövecses (2010: 215–216), there are variations in the *range* of CMs and metonymies for a given target, in the particular *elaborations* of CMs and metonymies, and in the emphasis on or the use of *metaphor versus metonymy*. Variations in the range of conceptual metaphor are often observed when an emotion is the target domain, such as anger or love. Japanese speakers tend to use the concept *hara* (belly in Japanese) in metaphors describing anger; the metaphor ANGER IS (IN THE) BELLY is unique to Japanese speakers. Likewise, speakers of Zulu often use the heart when describing the anger-emotion in metaphors while English speakers more often apply the hear-metaphor when describing love or affection (Kövecses and Benczes 2010: 215–216). Cultural differences apply more to metaphors where elements of knowledge are mapped onto a target, and less to image-schematic metaphors, since these relate to general physical experiences which affect all people regardless of culture.

Research on metaphor is an important aspect of the Cognitive Linguistic approach for understanding vocabulary, since it explains the way we understand the abstract in

terms of the more concrete. It also offers pathways to understanding symbolic language in idioms and other chunks by categorisation based on conceptual metaphor or conceptual domains, from which we can deduce pedagogical methods.

2.4.4 Vocabulary teaching and SLA from a CL-perspective

Cognitive Linguistics offers many insights that can be used in teacher-instructed L2 vocabulary learning, and research on L2 acquisition and related fields has driven many of the advances made in CL-research, according to Boers, Rycker, et al. (2010b). The authors listed three important common assumptions for these fields. First, research from cognitive psychology provides important insights, including models of perception and attention. Second, L1 acquisition is usage-based and language should be seen as a by-product of *communicative processes*. Third, the lexis-grammar dichotomy is fallacious, meaning that language should be seen as a continuum 'from atomic and specified units to increasingly complex and more schematic ones' (Boers et al. 2010b: 3). This last assumption explains the importance of chunks and the fact that words have 'their own grammar'—that is, words have their own collocational patterns.

In line with the first assumption mentioned above, CL-inspired vocabulary teaching often refers to the basic 'cognitive processing theories' underlying vocabulary acquisition, namely levels of processing theory, dual coding theory, and trace theory (Boers and Lindstromberg 2008b: 11–12; Vasiljevic 2015b: 5).

Levels of processing theory was introduced by Craik and Lockhart (1972; in Vasiljevic 2015b: 6), and this later gave inspiration to the involvement load hypothesis (see section 2.1.7), according to Nation (2013: 100). The theory holds that learning effects depend on the amount of cognitive effort invested in the learning process by the learner. The deeper the level of processing, the richer and more detailed the representations of the information. Deep-processed information is regarded more memorable and accessible for subsequent recall than information processed in a more shallow way, such as rote learning (Vasiljevic 2015b: 11). This theory explains the need for more complex language learner activities, where learners take part in an active way. It also points to benefits from tasks, in which learner awareness and deeper understanding of language is the goal.

Dual coding theory suggests that the formation of mental images facilitates learning: 'Stimulus can be encoded both verbally and visually, and information for which there are mental representations in both modalities can be retrieved more easily than

information mediated through only one modality' (Vasiljevic 2015a). In accordance with this theory, Cognitive Linguistic approaches to vocabulary learning seek to illustrate images of concepts, either by the use of visual stimulus or by focusing on the images that language creates, through discussion or other student-engaging activities.

According to trace theory, linguistic expressions that are repeatedly encountered entrench their traces in our memory, and this entrenchment creates more lasting traces (Baddeley 1990; Cohen, Eysenck, and LeVoi 1986; in Boers and Lindstromberg 2008b: 11–12). The entrenchment process is accompanied by 'detectible chemical and structural changes in the neurons presumed to be associated with the processing and storage of the information in question' (Squire and Kandel 2000; in Boers and Lindstromberg 2008b: 12). Based on this theory, we can assume that individual words or multi-word expressions need to be repeated several times in order to be memorised by learners. This is reflected in the emphasis in CL-inspired teaching on reviewing target items which have already been introduced during a learning lesson (Boers and Lindstromberg 2008a; Lindstromberg and Boers 2008a).

Moving on to the second assumption by Boers, Rycker, et al. (2010b) that underlie both Cognitive Linguistics and SLA, we are told that L1 vocabulary acquisition is usage-based and that language learning is seen as a by-product of communicative processes. L1 vocabulary learners or L2 vocabulary learners in immersion contexts (e.g., living in an L2 environment) have many opportunities for incidental acquisition and incremental uptake of language elements from surroundings, which is regarded as a main ingredient in vocabulary acquisition. However, L2 vocabulary acquisition (in non-immersion contexts) mainly takes place in a classroom-based setting with some additional media language input and provides comparatively little exposure to L2. Therefore, the incidental uptake in L2 classroom settings is restricted to a small number of high-frequency L2 elements (Boers, Rycker, et al. (2010b: 4). In this setting, the teacher-instructed intentional learning is considered necessary to accelerate the learning process.

The third out of the three assumptions that are common to SLA and Cognitive Linguistics regards the fundamental claim that there is no clear distinction between grammar and vocabulary. In other words, language is a continuum ranging from atomic units to more complex ones. This means that multi-word units should not be regarded as mere combinations of single-word units derived on the basis of grammatical rules, but

instead be studied as independent and autonomous units (see the principle of idiomaticity in section 2.2.3). Thus, from a CL-perspective, formulaic language plays an important role in vocabulary learning and teaching, and chunks should be regarded as important learning targets.

While chunks play an important role in CL vocabulary acquisition, there is also the general principle of utility (or usefulness) to consider when choosing learning targets for CL instruction (Boers et al. 2010b: 8). This principle is apparent in three well-established criteria for vocabulary selection, namely frequency, relevance, and ease of learning (Boers and Stengers 2008a: 369). *Frequency* generally correlates with utility, since it is useful to know words that we are likely to need and encounter in written or spoken language. Likewise, archaic expressions such as 'it's raining cats and dogs' [example given by Boers, Rycker et al. (2010b)] should be avoided because they are hardly ever encountered and hence has low utility. *Relevance* refers to the usefulness of learning items for the accomplishment of certain tasks or for specific purposes within a certain field or domain; one example of this is academic vocabulary. Lastly, *ease of learning* relates to the accessibility of an item. For instance, idioms such as 'show someone the ropes' can be easily explained since the literal meaning can be readily illustrated and used to explain the metaphorical meaning.

However, there are additional concerns when selecting learning targets. Boers et al. (2010a: 242) pointed out that even though high-frequency vocabulary is important, there are several reasons why explicit vocabulary instruction should also include medium- and low-frequency vocabulary. First, in order to reach a higher level of language proficiency, a certain amount of medium and low-frequency vocabulary is necessary, simply because there are such a vast number of words and chunks in a language. Second, low- or medium-frequency chunks require explicit instruction since they are less likely to be picked up incidentally through meaning-focused reading. This relates to the 'the noticing problem' that occurs in meaning-focused reading—that is, how readers' primary focus is on meaning rather than (noticing) form. If the reader understands the content of the expression 'commit a crime' from the context, he or she might still have to encounter the whole word string several times before noticing the collocational pattern as well as the form of each single word and incorporate the word string 'commit a crime' as opposed to 'do a crime' into their productive vocabulary. Boers et al. (2010a: 243) explained that

'only chunks of very high frequency appear to meet the conditions for incidental uptake through multiple encounters to become probable', and therefore argue that *medium-frequency chunks* are often the most suitable target items for explicit vocabulary instruction.

2.4.5 Elaboration on form and meaning connections

A number of studies have explored how the explication of linguistic motivation in words or phrases benefits vocabulary acquisition. Some studies measure the effect of explicating the motivation for different senses of polysemous words and the idioms in which they occur (Beréndi et al. 2008), whereas others study the mnemonic effect of alliteration in multi-word units (Boers and Lindstromberg 2005). These studies have a common ground in how they are oriented towards the pedagogical uses of Cognitive Linguistic theory, and how the understanding and awareness of vocabulary can benefit from the explication of motivation. However, while the first study examines the processes of motivation that relates to the *meaning*-aspect of vocabulary, the second study deals with processes of motivation that concern *form*. The Cognitive Linguistic term *elaboration* refers to the examination or mental exploration of both form and meaning connections. As mentioned in the previous section, the levels of processing theory states that the more engaging and effortful the mental work is (i.e., the deeper the mental process is), the greater the chances are that the information is taken up by long-term memory. Likewise, elaboration also builds upon dual coding theory and trace theory, as outlined in section 2.4.2.

Studies on lexis within the Cognitive Linguistic field regularly refer to different types of linguistic motivation with regards to which meaning and form connections are involved in the motivational process. Radden and Panther (2004) proposed three categories of linguistic motivation, or whether the process of motivation involves the following:

(a) meaning-meaning connections

(b) form-meaning connections / meaning-form connections

(c) form-form connections

Elaboration on meaning-meaning connections, also referred to as semantic elaboration, relates to any mental operations regarding the meaning of a word or phrase. This can be promoted by 'linking new vocabulary items to pre-existing vocabulary, fitting the items into meaningful scenarios, or associating items with mental images' (Boers and

Lindstromberg (2008b: 12). Semantic elaboration can be used for pedagogical purposes in several ways. Beréndi, Csábi and Kövecses (2008: 65–66) hypothesized that understanding the function of CMs or metonymies helps learners' comprehension and memorisation of polysemes and idioms. Moreover, learners may understand abstract concepts more easily if they are made aware of the source-target connection in metaphors, since concrete and abstract knowledge are not equally accessible from a vocabulary acquisition point of view (Schmitt 2015: 53).

Indeed, most studies on vocabulary in Cognitive Linguistics have focused on meaning-meaning connections. This includes studies on conceptual metaphor (see section 2.4.4), which primarily focus on the semantic motivation behind figurative words and expressions. According to Boers and Lindstromberg (2008b: 20), studies on meaning-meaning connections have largely focused on polysemic words and figurative idioms. For example, studies on polysemic words have analysed the use of motivated meaning networks of prepositions for pedagogical purposes (Taylor 1988; in Boers and Lindstromberg 2008b: 20) and examined English idioms to detect similarities and differences in metaphorical and metonymical conceptualisation of the human body in English and Hungarian (Csábi 2004). The other significant group of studies in this category deals with the semantic motivation of figurative idioms. Boers and Lindstromberg (2008b: 21) noted that many idioms may be treated as figurative since they have a conventionalized meaning that can be seen as originating in a trope, along with metaphor and metonymy. The motivation behind some idioms become apparent after diachronic exploration—that is, after having examined their etymological origin. According to Boers and Lindstromberg (2008b: 22), this group of studies has contributed greatly to the Cognitive Linguistic field by showing how idioms express CMs which are grounded in human experience.

The second category of linguistic motivation includes *form-meaning connections* and *meaning-form connections*. This refers to connections that are established when form is motivated by meaning or meaning is motivated by form. Examples are studies that examine whether word order in phrases reflects chronological events, or whether the meaning of some lexemes is derived from their phonological features; for instance, words that begin with the sound sequence /sp/ often have negative connotations, like *spam*, *spit* and *spoil* (Boers and Lindstromberg 2008b: 23). Likewise, research on onomatopoeia (i.e.

the sound) determines the form of the word that is used to imitate it. Elaboration on *semantic prosody*, which is the pattern with which words or word strings are steered towards certain meanings, also falls into this category.

The third category of form-form connections pertains to studies on rhyme, assonance, alliteration, and other patterns that shows phonological repetition, particularly in phrases. Many chunks show some kind of phonological motivation in their lexical makeup. According to Boers and Lindstromberg (2008c: 334), this applies to 20% of English idioms. This number is even higher (23%) among frequently used idioms, which may indicate that catchy sound patterns and/or ease of use plays a role in standardizing certain Chunks. Examples of phonological patterns of similarity include rhyme ('pie in the sky'), alliteration ('gas guzzler'), and assonance ('lo and behold'). Alliteration seems to be the most common phonological feature (Boers and Lindstromberg 2008c: 338). Also, a study by Gries (2013) suggested that phonological similarity in terms of *alliteration* is highly frequent in chunks. Using three different procedures Gries compared the frequency of alliteration in chunks to the frequency in the English language overall. One study that explores the phonological similarity of elements on the phonological pole is done by Boers and Lindstromberg (2005). This research paper reported that multi-word units that alliterate are more memorable than multi-word units that have no such salient sound patterns. Another study suggested that approximately 20 % of frequent idioms show a form of phonological motivation, mostly alliteration, but also assonance (cf. Boers and Stengers (2008b). Boers and Lindstromberg (2008c) argued that phonological motivation has been a neglected dimension in Cognitive Linguistics and that elaboration on form-form connections (i.e., structural elaboration) help learners remember multi-word expressions. They also found that alliteration proved to have a significant mnemonic effect. Gries (2013) investigated V-NP$_{DirObj}$ idioms (e.g., *kick the bucket* and *lose one's cool*) and way-construction idioms (e.g., *He fought his way through the crowd*), and he found that the lexical makeup of the idioms is strongly affected by alliteration. The terms *phonological elaboration* (Gries 2013) and *structural elaboration* (Boers and Lindstromberg 2008b) are used to describe mental operations regarding formal properties of a word or phrase. According to Gries (2013) and Boers and Lindstromberg (2008b), structural elaboration can be promoted by the recognition of features such as affixes, peculiarities of spelling, and salient sound patterns (e.g., repetitions as in rhyme).

Most Cognitive Linguistic studies view linguistic motivation from a synchronic perspective. However, there are also studies that view linguistic motivation from a *diachronic* angle. These studies often examine etymology—that is, from historical origins of words or expressions—and the elaboration from this perspective is referred to as *etymological elaboration*. This involves processes such as the identification of loan words and cognates, noticing changes in form or meaning over time, and analysing words or expressions and then breaking them down into meaningful affixes and roots (Boers et al. 2004; Boers and Lindstromberg 2008b). Boers et al. (2004: 58) studied how the recall of idioms can be facilitated by reactivating the literal sense of the expression and tracing the idiom back to its original use or context. In this case, etymological elaboration serves as a particular instance of the more general strategy of semantic elaboration. In an example of a multiple-choice task administered in this study, learners are asked what domain of experience the idiom *show someone the ropes* originates in, and the question offered them the possible options of *boats*, *prison*, *sports*. As a feedback to the task, the learners were given the explanation of the origin of the expression, namely that 'experienced sailors had to teach novice one which ropes they should handle' (Boers et al. 2004: 60).

2.4.6 CL-inspired learning activities

This section presents CL-inspired teaching activities that target chunks. It also accounts for Lindstromberg and Boers' (2008a) three stages of teaching chunks in what they refer to as the Lexical Approach—noticing, memorizing, reviewing.

Teaching inspired by Cognitive Linguistics (i.e., CL-inspired teaching) includes a large number of activities aimed at many kinds of students. A number of these are equally suitable to L1 and L2 students, but in some cases the study singles out activities that would be less suitable for ESL-students.

The list of CL-inspired teaching activities include some of the following exercises:

- locating chunks in text and pictures in exam preparation material, in the students' text books, and in the students' own written work;
- viewing pictures that illustrate the literal meaning of metaphors used in chunks (cf. etymological elaboration in section 2.4.5);
- guessing the meaning of chunks that are linguistic metaphorical expressions;
- listening to and analysing songs and movies where metaphors are used;
- drawing expressions (on paper) based on metaphorical content;

- finding and categorizing expressions which utilizes certain source domains or which highlights certain target domains; and

- discussing the use of English idiomatic language in various situations.

Many of these activities are adapted from Lindstromberg and Boers' (2008a) book *Teaching Chunks of Language* and the Lazar's (2003) book *Meanings and Metaphors*. Others are inspired by some of the activities presented in relevant studies (Hulstijn 1996; Boers et al. 2004; Jones and Haywood 2004; Boers and Lindstromberg 2008b; 2008c; Boers and Stengers 2008b; Kövecses and Benczes 2010).

Noticing refers mainly to recognizing chunks in language, but it also refers to noticing patterns and features in language that makes chunks easier to understand and remember. For instance, informing the learners about the literal sense of one of the main words in a phrase—such as the word *square* in the phrase *back to square one* (Lindstromberg and Boers (2008a: 13)—can facilitate the comprehension and memorisation of the phrase. Likewise, reminding the learners about alliteration in a formulaic sequence, such as 'the more the merrier', can make it easier for learners to remember it. Lindstromberg and Boers (2008a) presented a number of noticing activities that can be used in the first stage of teaching chunks, including (a) reading the text out loud with noticeable pauses where it seems natural which may lead the learner to notice and reflect upon chunks in the text, (b) playing 'Chinese whispers'[6] with slips of paper on which phrases from song lyrics are written, and (c) putting word strings that are part of a conversation into chronological order.

In the second stage of the CL-teaching—the stage of memorizing—learners should be encouraged to 'engage in rich mental processing that is likely to result in the formation of robust memories for [chunks]' (Lindstromberg and Boers 2008a: 47). This is where the main work of the learning should take place. This includes activities in which students read sound phrases to each other (e.g., *crashing waves*, *distant drums*) and tell each other what they think of the sounds; here, the students can sort figurative idioms by the source domain (see section 2.4.3 which explains the role of source and target domains in conceptual metaphor) and can try to pair verbs and nouns which are collocates. These

---

[6] Chinese whispers is 'a game in which a message is passed on, in a whisper, by each of a number of people, so that the final version of the message is often radically changed from the original' (Collins Dictionary).

activities can be regarded as elements of different kinds of elaboration, as described in section 2.4.5.

The third and final stage in Lindstromberg and Boers' (2008a) design of CL-inspired teaching is reviewing. This echoes the claim put forth by trace theory as described in section 2.4.4, which states that repeated encounters with vocabulary items is important because they 'entrench' them in memory traces in the human mind. According to Lindstromberg and Boers (2008a: 83), it is well established that reviewing is important to long-term recall of the targeted vocabulary items. Reviewing procedures should aim to involve 're-noticing' of particular chunks, provide new context so as to consolidate or extend understanding of the targeted vocabulary items, and make the students hear, say, and write the chunks—not just read them, among other things. One example of the activities proposed by Lindstromberg and Boers (2008a) is the composition of mini-stories with ready-made themes (by the teacher) that goes well with previously learned chunks; for instance, an example given by Lindstromberg and Boers (2008a: 87) involves writing about the first day at work, which goes well with idioms such as *learn the ropes* and *try a new tack*. Other examples of activities include filling in missing words in song lyrics that are already familiar to the students with the help from clues (e.g., keywords or sentences with rearranged word order) and mapping which chunks the students know and which ones they do not know together with the students.

2.5 Summary

This chapter covered theory on vocabulary, chunks, academic vocabulary, and the Cognitive Linguistic approach to vocabulary. The experiment in this thesis essentially measures the effects from teaching of vocabulary. Thus, the broader topic of vocabulary and L2 learners of English was a natural starting point from which to build the framework. This first section addressed different aspects of vocabulary knowledge, L2 vocabulary learning and vocabulary acquisition strategies. Sections 2.2 and 2.3 then narrowed the scope and explored the two sub-categories of chunks and academic vocabulary, which are the specific types of vocabulary items targeted in the experiment. Section 2.2 accounted for the term chunks; the section also presented different ways of classifying the term and elaborated on the connections between chunks and L2 learning. Section 2.3 explained the notion of academic vocabulary and the importance of corpus research in the construction of vocabulary lists, and discussed the role of chunks, specifically idioms

in academic vocabulary. These first three sections represent a shift from a broad scope to a narrower one, from the broad term vocabulary to the narrower kind of vocabulary targeted in the experiment, namely academic chunks. The fourth and last section introduced a new approach to the academic chunks and to the teaching and learning of these, namely the Cognitive Linguistic approach. In this section, the study presented some of the features that characterize the Cognitive Linguistic approach and which separate it from other related approaches. By explaining the importance of the concept of linguistic motivation and the theory revolving around conceptual metaphor, the study aims to establish a conceptual framework for 'CL-inspired' teaching and learning. The study went on further to explain the link between L2 vocabulary acquisition and Cognitive Linguistics as well as the different routes to elaboration on form and meaning from a CL-perspective. Lastly, the study presented some thoughts on how to organize CL-inspired teaching of vocabulary—that is, the different stages in teaching or learning of vocabulary and what kind of activities can be used within these stages.

3. Methods

This chapter presents the methods and methodology used in the present study. The first section addresses the overall methodological approach; the second describes the technical details of how the experiment in this study was conducted; the third explains how the results from the studies were analysed; and, finally, the fourth comments on some of the methodological issues associated with these kinds of study and with this study in particular.

3.1 Methodological approach

According to Janda (2013), the field of Cognitive Linguistics has developed from a mainly qualitative field into a quantitative one. Between 1990 and 2007, most articles used a qualitative methical approach, but this changed from 2008 onwards, according to Janda (2013: 4). She describes this change as the 'quantitative turn', since the majority of studies after this point employed applied quantitative methods.

The present study falls in line with other studies that reflect this current trend towards the quantitative approach. According to Dörnyei (2007: 32), one feature of quantitative studies is that they are 'centred around numbers'. As such, the testing in the present experiment measures *how many* multi-word units the participants highlight in a text handed out, *how many* words in expressions they are able to fill out correctly, and *how many* collocations and idioms they are able to translate. Another feature that is more common in quantitative research is *a priori categorisation*;[7] for example, in the present test, each test item is defined as a certain type of multi-word unit *a priori* to the testing. A third feature of quantitative research is the use of statistical methods in the presentation of findings, which aims at generalisability. The test results are presented as scores in tables, and I make use of descriptive and inferential statistics to interpret the results. Finally, the findings in this study apply only to groups of learners that share the characteristics of the participants who are tested.

Classroom studies such as this one are often quasi-experiments that stand in contrast to 'true experiments' because the random assignment of the participants in the study can be difficult to achieve (Dörnyei 2007: 117). The students are part of set classes,

---

[7] Some would argue that a priori categorisation is as much a feature of qualitative research, demonstrated by qualitative procedures such as Stepwise-Deductive Induction, by Tjora (2018).

and if treatment and testing are to be administered during regular teaching hours, it is often difficult from a practical point of view to divide the students into new groups. According to Zoltán Dörnyei, using non-equivalent participant-groups has become 'an accepted research methodology' (Dörnyei 2007: 117) in fields in which randomising the participants is not possible. However, the effects of these initial group differences must be taken into account when making causal claims based on quasi-experimental studies.

In quasi-experimental designs, as in true experiments, the use of a control group is essential. One of the greatest concerns with such a study is whether a potential effect, that is, a quantitative difference between the pre-test and the post-test, should be attributed to students' ability to memorise a set of answers over a limited time period or to more lasting learning benefits which give the students tools to develop cognitively over time. After the pre-test, the teacher goes through the correct answers together with all three groups. Obviously, having already tried once – and having been presented with the correct answers, many of the students will score higher on tasks 3 and 4 simply from remembering the correct answers or looking up the correct answers themselves. The use of a control group can potentially reveal such a bias, since it provides a means of comparing post-test scores between the experiment groups and the control group.

3.2 About the experiment

This section goes through the different elements of the experiment undertaken in this thesis: the participants, the vocabulary items that are targeted for teaching and testing (hereafter referred to as target items), how the target items are taught, and how the participants' knowledge of these target items is tested. First, I present the participants. Subsequently, I account for the different categories of target items, and I explain why these specific target items were chosen. Third, I describe the treatments administered in the present experiment: five teaching sessions based on CL-inspired teaching principles, as introduced in Chapter 2. Finally, the testing procedure consists of four different kinds of tasks and aims to enable a comparison of treatment effects across different aspects of knowledge.

3.2.1 Sampling

The participants in this study are 33 Norwegian students in upper secondary school who are attending vocational lines of studies in the program subjects of electrical engineering

(12 students, treatment group 1), building and construction (13 students, treatment group 2), and technical and industrial production (8 students, control group). Norwegian is the first language of all students. English is a mandatory subject in Norwegian upper secondary schools, and in vocational lines of studies, it is completed over a two-year-period, with a possible final written exam at the end of the second year. All participants in this study are second-year-students.

All the students who participated are males. 29 of the 33 students are 17 or 18 years old. There are two older students in treatment group 2, at 23 and 26 years old, and two older students in the control group, at 21 and 23 years old. The participants are thus younger than the participants in many similar studies (Laufer and Goldstein 2004; Boers and Lindstromberg 2008a; Peters and Pauwels 2015), who are typically college or university students. On the one hand, learning benefits can be detected more easily because the students may be unaccustomed to the kind of academic language they are presented with. They are probably also less well-travelled in English-speaking areas and consequently less familiar with English native speakers. Thus, they form a 'purer'[8] group of L2-language users. On the other hand, there are challenges related to younger students, namely, that they are overall less mature and less proficient in English compared to students at a college or university level. Students in the vocational lines of study can also be assumed to be less motivated for further English studies since they have chosen to attend a programme that does not aim at further academic studies. Hypothetically, such a relative lack of motivation might have a negative impact on tasks that require productive language skills, since these are known to require higher motivation and greater overall language proficiency than tasks that require receptive language skills (Nation 2001: 31–33). This might result in shorter answers in the essay-writing test component and less willingness to engage in oral elaboration in front of the rest of the class during the learning period.

---

[8] The term 'pure' is used here in parallel to how it is sometimes used in works of dialectology, i.e. to describe a dialect with little or no interference with other dialects. An example of this is the following passage from J.C. Wells' Accents of English 2 (1982: 395): 'Hence the received wisdom that the purest English is spoken in Inverness; in Inverness, Scots has never been in general use, since there Gaelic was displaced directly by Standard English'.

3.2.2 Target items

Earlier studies on L2-vocabulary learning in Cognitive Linguistics have used suitable sets of (metaphorically related) figurative idioms (Boers and Lindstromberg 2008a) or sets of polysemous words (Csábi 2004) as target items. The experiment in this thesis tests students' understanding and awareness of academic chunks that are particularly high-frequent or 'useful' in academic writing. As explained in Section 2.2, corpus is a valuable tool for research on chunks, and the target items are gathered from studies which have constructed vocabulary lists based on computer-generated corpora. In addition to the academic chunks, there are some non-academic chunks included in the testing as well. These are idioms that are particularly suitable for CL-inspired teaching of vocabulary and typically contain figurative meaning or phonological features such as alliteration or rhyme. These idioms have already been used as target items in earlier CL-inspired teaching experiments (cf. Boers and Lindstromberg 2008a). This enables a comparison between non-academic and academic chunks, which in turn might provide information on how suitable academic chunks are to the teaching and learning strategies in this experiment.

In all, there are 148 target items that are used variably in the test components, presented in Tables 3.1, 3.2 and 3.3. Table 3.1 shows 20 academic idioms, which 'lend themselves particularly well to an academic context, based primarily on their semantic content but also partly on frequency', according to Simpson and Mendis (2003: 435).

Table 3. 1: List of particularly useful idioms for English for Academic Purposes Curricula.

bottom line	hand in hand	the big picture
hand-waving	carrot and stick	in a nutshell
chicken-and-egg question	ivory tower	come into play
litmus test	draw a line between	on the same page
get a grasp of	play devil's advocate	get a handle on
shift gears	get to the bottom of things	split hairs
go off on a tangent	thinking on my feet	

Note: Adapted from Simpson and Mendis (2003)

Table 3.2 shows the next 56 target items, which are some of the most common academic formulas in written English. In a study by Simpson-Vlach and Ellis (2010), a group of 20 experienced EAP-instructors (instructors of courses on English for Academic Purposes) rated the *formula teaching worth* (FTW) of the most common formulas in written English based on 'an empirically derived psychologically valid measure of utility' (Simpson-Vlach and Ellis 2010: 488). The EAP-instructors judged the following three parameters, using a scale from 1 to 5:

[1] whether the phrase constituted a formulaic expression

[2] whether the phrase had a cohesive meaning as a phrase

[3] whether the phrase was worth teaching as a 'genuine' phrase

Based on the added values on these three variables, they arrived at one *formula teaching worth (FTW)* for each phrase. The 56 target items in Table 3.2 are the academic formulas with a FTW-value above 1, as reported by Simpson-Vlach and Ellis (2010).

Table 3. 2: List of common academic formulas in written English with a formula teaching worth (FTW) above 1

on the other hand	the United Kingdom	is likely to be
due to the fact that	on the basis	wide range of
on the other hand the	there is no	the same way as
it should be noted	over a period of	due to the fact
it is not possible	as a result of the	in accordance with
a wide range of	can be seen in	it is necessary
a number of	a wide range	the other hand
such a way	there are a number	can be seen
take into account the	it is interesting to	it is likely
as can be seen	it is impossible to	such a way that
it is clear that	it is obvious that	carry out
take into account	it is possible to	it is possible
can be used to	it is not possible	with respect to
in this paper we	been carried out	give rise to
likely to	can be found in	carried out by
next section	it is important to	whether or not
large number of	was carried out	present study

| should be noted | the other hand the | his or her |
| be carried out | does not appear | |

The last set of target items is shown in Table 3.3. These represent a selection of idioms from a study by Boers and Lindstromberg's (2008a) in an account of a Cognitive Linguistic approach to teaching idioms whose conceptual metaphors can be traced to specific source domains. The first 21 idioms contain conceptual metaphors that belong to the domain of BOATS/SAILING, and the next 51 idioms contain conceptual metaphors that belong to the domain of WAR/AGGRESSION. I refer to these target items as 'CL-suitable idioms' hereafter.

Table 3. 3: List of idioms from highly productive source domains that provide CL-inspired vocabulary teaching

clean bill of health	the tip of the iceberg	close ranks
take something on board	in the wake of	step out of line
a close call	be up in arms	on someone's watch
a loose cannon	a baptism of fire	be on your guard
steer clear	drop a bombshell	off guard
stay the course	in the front line	keep your head down
show your true colours	come under fire	a Trojan horse
with flying colours	in the line of fire	a body blow
dead in the water	fight a rearguard action	put the boot in
clear the deck(s)	a last ditch attempt	bite the dust
in the doldrums	stick to your guns	cloak and dagger
on an even keel	burn your boats/bridges	throw down the gauntlet
leave someone high and dry	the cut and thrust	rattle your sabre
break the ice	steal a march	to the hilt
pass muster	give lock, stock, and barrel	a hit list
leading light	the standard bearer	hit and miss
show someone the ropes	gain ground	show your true colours
(all) at sea	stand shoulder to shoulder	with flying colours
a shot across someone's bows	break ranks	a loose cannon

a shot across someone's bows	take it on the chin	no holds barred
be at loggerheads	fight your corner	stick your neck out
head-to-head	not pull your punches	up to scratch
ride roughshod over someone	a slap in the face	be on the ropes
lower your guard	drop of a hat	throw in the towel

Note: Adapted from Boers and Lindstromberg 2008a: 389–391

3.2.3 Treatment of Norwegian L2 learners in the form of CL-inspired teaching

The explicit CL-inspired teaching of the two treatment groups made use of many of the activities presented in section 2.4.6. Five 70-minutes-teaching sessions were set aside for explicit CL-inspired teaching for the two experiment groups. These are discussed in the following.

Teaching session 1

The first session begins with an introduction to the distinction between 'words' and 'expressions' or 'chunks of words'. Many words in various texts are part of fixed expressions and cannot be understood properly without the awareness of this 'fixedness'. The teacher elaborates on expressions such as 'come on', 'to be square', 'cry over spilt milk', which many of the students know and which have Norwegian equivalents with which they are familiar. The teacher explains that many terms are intended to mean something other than what the literal interpretation of the expression would indicate.

The next step in this teaching session is to have the students create a 'chunk bank', in which they classify expressions that are mentioned in these teaching sessions into several columns for different categories. The first category is 'expression'; the second is 'translation [of each single word into Norwegian]'; the third is 'extensional definition in English' (the English *extensional definition* is translated into the Norwegian 'overført betydning'). The students then 'store' the expressions in this chunk bank. Finally, the students are asked to create some columns to the right and leave them empty for later use.

Teaching session 2

After a brief review of some of the main points of teaching session 1, the teacher begins session 2 by introducing the term 'metaphor'. Most of the students have heard the word before; some of them remember it from their Norwegian classes, while others cannot remember where they have heard of it. The teacher explains the term and uses examples from songs and movies in which metaphor use is salient. The teacher highlights several expressions in song lyrics and in movie titles or subtitles, and discusses the meaning which can be extracted from a literal interpretation versus the intended or figurative meaning(s) (various extensional definitions).

The teacher presents idioms in which a metaphor is easily spotted, such as 'hit the nail on its head' and 'show someone's true colours'. Then he or she presents pictures that illustrate the literal interpretation of the expressions. Furthermore, the students are asked to guess the extended or figurative meaning of the expressions. In the end, the teacher provides suggestions for correct answers.

The teacher introduces the terms 'concrete' and 'abstract' and asks the students to add a new category in their chunk bank: 'source'. The teacher explains that the metaphors used in the expressions can originate from a source. In the song 'You are the sunshine of my life', the reflecting light from the closest star in our galaxy is the concrete thing highlighting the abstract thing that the author wants to describe, that is, a person who brings joy to his or her life in some way.

At the end of the teaching session, the students collect new English expressions from the ones presented by the teacher to store in their 'chunk bank'.

Teaching session 3

In teaching session 3, the teacher explains the meaning of 'academic' words and expressions. Examples of academic texts are given, some of which originate from previously given exam preparation material and some from the students' own English textbooks. The students are presented with a list of statements about different topics and are asked to come up with arguments for and against one or more of these statements. To do so, they are given a list of 'useful expressions for debating' that they can use to improve their texts. Some examples of the expressions presented in the list are 'from my point of view', 'we are on the same page', and 'this argument does not hold water'.

The students add the expressions they use from the list of useful expressions for debating to their chunk bank. Many of the students find it difficult to fill out all the fields

in their chunk bank, since the kinds of expressions dealt with this time are mostly *collocations*, in which metaphorical content is more difficult to spot.

Teaching session 4

Teaching session 4 begins with the introduction of another type of chunk, namely, English phrasal verbs. This is followed by a discussion in the class of what counts as a phrasal verb. The difference between phrasal verbs and prepositional verbs is briefly mentioned in this context. The teacher presents pairs of sentences in which a phrasal verb is used metaphorically in one of the sentences and literally in the other. The teacher uses sentence pairs such as 'the dog dug up an old bone' / 'they dug up some interesting facts' and 'two planes were shot down' / 'each proposal was shot down'. The teacher asks the students to identify which sentence contains the metaphorical expression and which one contains the literal expression. Then, the teacher hands out a list of phrasal verbs arranged by their preposition (group 1: 'pull off, take off, clear off'; group 2: 'give in, hand in, hold in', and so on). The students are asked to choose three phrasal verbs and create sentence pairs like the ones presented by the teacher.

Teaching session 5

The fifth and final session begins with a summary of the topics presented in the previous sessions. The teacher elaborates on the different categories in the chunk bank. The students present some of the expressions they have stored. The teacher goes on to present a text from the preparation material from a previously administered exam (not the one used in the test phase). The students are asked to highlight new expressions in this text, and some of them are discussed in class. The students read through their own texts (the two essays they wrote during the pre-test), and they are asked to replace words or expressions with entries from their chunk bank where it may be appropriate.

3.2.4 Testing

All the participant groups take part in two identical tests – one before and one after the treatment period, that is, the period from the beginning to the end of the treatment. This test design draws heavily on a study by Jones and Haywood (2004), which, among other methods, also used an essay-writing task, a highlighting task, and a modified cloze test ('gap fill-task'). In this study, however, the testing also consists of a translation task, and the participants write two essays instead of one. The test design aims to examine several

kinds of vocabulary knowledge, as discussed in Section 2.1. This includes receptive versus productive skills (pp. 8–10) and breadth versus depth of vocabulary knowledge (pp. 10–12). The test comprises the following four components:

(1) Essay writing, parts 1 and 2 (70 x 2 minutes): The participants write one essay on the topic 'Bullying' (part 1) before the other three test components. At the end of the day, they write another essay, on the topic of either 'DNA-modification and future technology' or 'the police and the Afro-American community in the USA' (in the pre-test, the students choose which topic they want to write about; in the post-test, they write about the topic they did not choose the first time). As an introduction to these topics, the students watch (before both pre-test and post-test) the documentary *Bully (2011)* and season 4, episode 1 ('Unnatural Selection'), and season 5, episode 5 ('Black and Blue'), of the TV-series *Vice* (*2014-present*). The teacher also talks about the documentary and the TV-series in class and answers questions about them. This test component measures the frequency with which participants use the entire list of target items in text production.

(2) Highlighting (30 minutes): In an excerpt from a text from the preparation material attached to a previously administered written exam (spring 2017), the participants are asked to highlight words or phrases which would be useful to know for students who are less proficient than themselves. This task measures awareness of multi-word units, or possible chunks, in text containing academic vocabulary. When referring to the underlined words in this task, I refer to multi-word units rather than chunks, because I do not evaluate whether the highlighted words are in fact chunks (cf. the distinction between chunks and multi-word units in Section 2.2.3). All instances in which two or more words are underlined are counted, except if the underlined words form several clauses in a sentence. This exception ensures that the participants are actively selecting multi-word units. If a participant, for instance, highlights an entire paragraph, it is difficult to know whether he or she is aware of any multi-word units or if he or she is actually highlighting all the single words in that paragraph.

(3) Gap filling (40 minutes): The participants are asked to fill in the missing letters of 20 chunks in which the last part of the words have been removed in the same two texts as in task 2. This task measures the participants' ability to recognise and produce chunks in academic texts. The answers to each of the 20 gap-fill alternatives receives a score of 1–3 points based on how close it is to the correct phrase.

(4) Translation (70 minutes): The participants are asked to translate 30 chunks from English into Norwegian based on how they are used in example sentences that are provided (one example sentence for each chunk). The 30 chunks are selected from each type of target item and include 10 'academic collocations', 10 'academic idioms', and 10 'CL-suitable idioms'. This task measures the participants' knowledge of English chunks.

After this pre-test, the treatment period begins, in which the two experiment groups receive *explicit CL-inspired teaching* and the control group receives normal teaching. In short, normal teaching consists of reading texts and doing exercises (speaking, listening, and writing tasks) from the textbook *Skills* (Lokøy et al. 2013). This textbook is normally used by the students in the treatment groups as well. Other activities of normal teaching include watching suitable, fact-based documentaries or movies with a fictional plot and using the internet for educational purposes. When the treatment period is complete, the participants go through the post-test, that is, the exact same test as the one they went through before treatment.

3.3 Analysing data from the experiment

As already mentioned, using tables and statistics is necessary to present data from quantitative studies. I employ both *descriptive* and *inferential* statistical measures to provide a complete account of the results from the experiment I have undertaken. Dörnyei (2007: 209) explains that while descriptive statistics are useful for describing the results of participants in a sample, inferential statistical procedures are necessary to draw conclusions from the results regarding the wider population.

The descriptive statistical values are found by manual calculation. To find the *inferential statistical* values, however, I enter the scores of the participants in the data editor in the statistical software SPSS. The measurement procedure ANCOVA is suitable

for calculating significance (*p*-values) when using a control group, since it offers the option of using a *covariate*, allowing for control over any differences in scores between the participants in the control group and the two experiment groups in the pre-test. However, the data material must meet certain assumptions in order to provide reliable and significant values in ANCOVA. These assumptions are *univariate normality*, *homogeneity of variance*, and *homogeneity of regression slopes* (Field 2017: 580–584). If these assumptions are not met, the results from the ANCOVA might still be used, if interpreted with caution (Grace-Martin; Field 2017).

3.3.1 Descriptive statistics

I measure the mean and standard deviation in my data, which are the most common measures of central tendency and variability, according to Dörnyei (2007: 214). The mean is the average of the scores and considers all the scores; however, the disadvantage is that extreme scores skew it considerably. Standard deviation indicates the average distance between the scores and the mean and can therefore be used to balance the impression provided by the mean. The standard deviation is high for heterogenous samples with extreme scores and low for homogenous samples with many scores close to the mean.

3.3.2 Inferential statistics

This study measures the *effects* of treatment (CL-inspired teaching) on a dependent variable (vocabulary knowledge and awareness) in a quasi-experiment. These effects are interesting to the extent that they apply to a wider population, which, in this case, is young-adult L2-speakers of English. It is therefore essential to find out whether the observed effects in the experiment are *significant*, that is, whether they occur so consistently that we may conclude that a random sample from the relevant wider population would probably also reveal the same effects. The findings of the experiment can be formulated as a yes- or no-answer to a null hypothesis and an alternative hypothesis:

(1) 'Null hypothesis': CL-inspired teaching has no effects on knowledge and awareness of chunks compared to traditional teaching

(2) 'Alternative hypothesis': CL-inspired teaching has effects on knowledge and awareness of chunks compared to traditional teaching

Measures of *statistical significance* provide an answer to whether the results gathered in the samples are 'true' for the relevant wider population, that is, whether they are generalisable, by testing the null hypothesis. The result of a significance test is a *p*-value, which indicates the *probability* that we would obtain results this far from the null-hypothesis in a random sample from the relevant wider population. If, on the one hand, this *p*-value is equal to or lower than the *significance level*, which is normally 0.05 (5%) in applied linguistic research (Gass 2015: 137), the results can be considered significant. If, on the other hand, the results are non-significant, we 'cannot be certain that [they] did not occur in the particular sample[s] only because of chance' (Dörnyei 2007: 210). Statistical significance is a function not only of the magnitude of the result but also of the size of the sample. A result which deviates only slightly from the null hypothesis might still prove significant if the sample is large enough. This feature provides valuable information, but it can also be seen as a methodological weakness, because there is always some difference between a sample and the whole population, and, consequently, if we test a large enough sample, we will most certainly get significant results. Plonsky and Oswald (2014: 879) explains this as follows: 'the null hypothesis is always a priori false, even though one may not have enough data to reach that conclusion empirically'. Another problem with statistical significance is that the judgement of a hypothesis requires a more nuanced look at the practical value of the results than that of a dichotomous yes-or-no outcome. Finally, the set significance level is completely arbitrary, and there is no reason why the significance level should not be 0.06 rather than 0.05 in many cases (Plonsky and Oswald 2014: 880).

For these reasons, *effect size* is often used as a supplementary measure. The effect size provides information about 'the magnitude of an observed problem' (Dörnyei 2007: 212), and it is not dependent on sample size like *p*-values are. However, the shortcoming of effect size is that there is no single universally accepted means of measuring it. Rather, there is a whole range of different estimates of effects size; the APA lists more than a dozen different alternatives (Dörnyei 2007: 212). In this study, I refer to the 'partial eta square' that is produced by SPSS when performing an ANCOVA. By examining the amount of variation in the groups (control group versus experiment groups), we gather information about the consistency of the scores and, hence, the strength of the tendency we seek to measure (Coe and Robert 2002). It is difficult to say exactly what constitutes

a large effect size; however, Draper (2011) provides some alternative guidelines for deciding the different levels when using partial eta squared as a measure of effect size: 0.01 may be considered small, 0.06 medium, and 0.14 large. However, Plonsky and Oswald (2014) argue that effect sizes should be set higher in L2 research. They only refer to *d*-values, another kind of measure of effect size, but their conclusion applies to effect sizes in general. Thus, the corresponding partial eta squared-levels should probably be set closer to 0.03 (small), 0.1 (medium), and 0.2 (large).

3.4 Methodological issues and limitations

In this section, I present several methodological issues that concern this study in general and the testing components in particular. I use the terms *reliability*, *research validity*, and *measurement validity* in accordance with Dörnyei (2007: 51–53). As such, reliability concerns the extent to which the testing is carried out in such a way that it is possible to reproduce the test results. Research validity relates to the entire study, that is, whether the results are indeed caused by the treatment ('internal validity') and whether the results can be generalised to populations that are similar to the participant groups ('external validity'). Finally, measurement validity concerns the interpretation of the test results, that is, whether the test measures what it is supposed to measure.

3.4.1 Research validity

Research validity concerns the overall quality of the research project (Dörnyei 2007: 52), in terms of both internal and external validity. Dörnyei (2007) explains that the findings from a study are *internally valid* if the outcome is a function of the variables that are measured and *externally valid* if the results can be generalised to a larger group than the sample involved in the study.

There are several threats to internal validity, including practice effects, boredom effects, and the Hawthorne effect (Dörnyei 2007: 53). Practice effects relate to the experience participants have gained by having already gone through the test, that is, if the post-test is identical to the pre-test. Since the participants have test experience in the post-test but not in the pre-test, we may expect higher scores among the participants in the post-test. Boredom effects, on the other hand, may cause the participants to score lower in the post-test than they did in the pre-test: The participants may become bored from extensive testing, particularly if there is little variation in the tasks. The third type of threat

to internal validity is the Hawthorne effect, which relates to the increase in effort (by the participants) due to the presence of the researcher(s). This threat is believed to be particularly salient in research within applied linguistics, as it is known to interfere in the use of spontaneous language (Mellow 1996: 334; in Dörnyei 2007: 53).

To control for these unwanted effects, we may implement a control group. If both the control group(s) and the treatment group(s) are subjects to the same testing procedure, we may assume that the differences between the groups are caused by the treatment variable, which is the variable which we want to isolate and measure. In the present study, test components 3 and 4 (the 'gap-fill task' and the 'translation task'; see Section 3.2.4) were particularly susceptible to practice effects, since the exercises in these test components are answered either correct or incorrect. The participants could have memorised the answers and answered correctly in the post-test regardless of any treatment they would receive between pre- and post-test. Task 2 (the 'highlighting task') measured awareness rather than memorisation and was likely more resistant to practice effects since there were no 'correct' answers to this task.

One aim with this study was to generate findings that can possibly apply to larger groups of learners of English, or at least to Norwegian L2-learners of English in upper secondary school. This concern of *external validity* was important when selecting participants for the study. Ideally, as is the case in all studies in which a sample of participants represents an entire population, a larger and more diverse sample of participants would be preferable. In this study, the fact that all participants attended vocational lines of studies in an upper secondary school left some open questions: Do students who choose to attend vocational lines of study differ systematically from other kinds of students with regard to certain skills, cognitive attributes, or preferences? If so, how does this influence their responses to the kind of treatment described in this study? How would older students, or students in other lines of study, respond to the kind of treatment presented in this study? To apply the findings in this study to other kinds of learners or younger and older learners, we would need to know the answers to these questions.

3.4.2 Measurement validity

Another form of validity is measurement validity. This term is used by Dörnyei (2007), whereas other researchers, such as Shadish, Cook et al. (2002), use the term construct

validity. Both terms relate to the question of whether the test measures what it is supposed to measure (Dörnyei 2007: 52). Shadish, Cook et al. (2002: 65) explain that the term construct validity refers to the twin problem of understanding constructs and measuring them.

The aforementioned inclusion of a control group made it possible to eliminate certain threats related to research validity, enabling us to measure the effects of the treatment more precisely. However, this tool does not provide more information about what the actual effects of the treatment are or, alternatively, what treatment we actually measure. Put in another way, when we are finally able to (after the inclusion of a control group) isolate the variable measured by the differences in scores between the pre-test and the post-test, do we then measure CL-inspired teaching or something else? Can we attribute the differences in scores to new and increased understanding and awareness of chunks generated by CL-inspired teaching or not? In other words, did the treatment sessions generate long-term learning effects, or did they simply reinforce the sheer memorisation of the chunks presented to them during and after the pre-test?

We should also ask whether those who were able to learn or memorise many chunks might also be more inclined to highlight many multi-word units rather than single words, not because these lexical units were the focus of attention during the period they participated in the experiment. In other words, were the results from test component 2 caused by 'awareness' or by a temporary exposure to the presented input.

The researcher faces a similar concern in a test such as task 1, the essay task as presented in Section 3.2.4: Does the participants' use of the target items actually measure their knowledge of academic chunks, or is the pool of target items too small to reveal such knowledge? No doubt, a larger list of target items would be preferable, and is maybe necessary, to achieve valid results. However, since chunks can be defined in so many ways and there are no 'complete' lists to use as a reference for this (cf. Read and Nation 2004: 24–25), a more limited examination of the production of chunks was chosen as test component. Even though this test component only measures the use of the chunks in the three target items lists, it may nevertheless serve as a rough indicator of the participants' ability to produce chunks in text.

3.4.3 Reliability

In task 2, the participants' interpretation of the instructions given obviously affects the way they perform the task and may therefore affect the test's *reliability*. They are supposed to 'imagine that they are asked to give advice to first year students who want to improve their academic writing in English'. After the treatment period, the participants should be more aware that the English language has many chunks and thus highlight more multi-word units. However, some participants highlight many words in general, maybe entire sentences or paragraphs, while others highlight only a few words. This pattern can be seen in both the pre-tests and the post-tests. Also, some participants choose to highlight low-frequency words, while others highlight high-frequent words. The reasons for these differences are complex, and the term awareness can be used ambiguously.

3.4.4 Practical limitations

The treatment in this study had to be administered within the time frame of the students' normal English classes. These classes amount to one 70-minute lesson per week. It was important not to overshadow the regular English teaching sessions during this period, since the CL-inspired teaching of chunks was primarily meant as a supplement. Furthermore, some English classes were cancelled for various reasons, like practical work related to their line of study (organised by the school), public holidays or various school arrangements. It was therefore difficult to ensure the kind of continuity that would be ideal for such a teaching project. Consequently, students sometimes forgot the main ideas presented in one session before they were presented with new ones.

3.4.5 Ethical considerations

Dörnyei (2007: 63–64) notes that ethical concerns are often more acute in qualitative, social studies, in which the researchers are sometimes interested in people's personal views and target sensitive or intimate matters. Nevertheless, there are several important ethical considerations worth mentioning with regards to quantitative studies, such as this one.

It is important to ensure that the *participants do not suffer any loss and, ideally, benefit from the study* in some manner. In this study, treatment was only given to two out of three groups. In cases in which treatment proves to be more effective than 'no treatment', participants in the control group should ideally receive some form of treatment

after the study, so that all participants eventually benefit from partaking in the study. In this study, the members of the control group were also given some 'CL-inspired teaching' after the experiment had ended as recompense for the lack of treatment during the experiment period. While not an issue in this study, such compensatory treatment of the control group would obviously interfere with a prospective follow-up 'post-post-test' if this had been part of the research design.

Another important concern is that of *the anonymity of the participants and the storage of confidential information*. In this study, each participant was given a random number, and, in addition to the participants' age, this is the only information published, thereby maintaining the participants' anonymity. Furthermore, the storage of the participants' answers to tasks and their names are in accordance with guidelines from the Norwegian Centre for Research Data.

Researcher integrity is also an important ethical concern. This relates to, among other things, how the researcher treats and presents her own and others' findings; for example, the researcher should refrain from data fabrication and from misrepresenting other authors' ideas or findings as a means of achieving certain findings.

4. Results

In this chapter, I present data results from the four tasks included in the testing phase. I also analyse the findings from each of the tasks. The average score among the participants (the mean) and the amount of variation in the score of all the participants in the group (the standard deviation) are calculated for all tasks. In task 1, I have also included the total number of occurrences among all the participants (the total score). Furthermore, I present the p-values and effect sizes, measured by partial eta square, obtained through the statistical procedure ANCOVA.

4.1 Test component 1: Essay writing

Tables 4.1–4.3 present the number of academic collocations, academic idioms, and idioms suitable for CL-teaching ('CL-suitable idioms') that the participants used in their essays in the pre-test and in the post-test, as well as the total number of words used. The tables also show *which* expressions are used, as well as how many times these are used – if used more than once (in parentheses). Table 4.5 shows the results from the ANCOVA, indicating the p-value and effect size.

Table 4.1 shows the results of the essay writing task.

Table 4. 1: Results of the essay writing task for experiment group 1

	Pre-test	Post-test
Participant 1	1 ('on the other hand') /709	0/713
Participant 2	0/590	1 ('likely to')/984
Participant 3	0/986	1 ('likely to')/921
Participant 4	0/617	0/930
Participant 5	1 ('the same way as') /829	4 ('there is no', 'on the other hand' (2), 'it is necessary')/1062
Participant 6	0/753	0/928
Participant 7	0/939	13 ('a wide range of' (6), 'take into account', 'on the basis of', 'it is important to', 'on the other hand' (2), 'come into play', 'the tip of the iceberg') /893
Participant 8	0/801	1 ('there is no')/874
Participant 9	1 ('there is no') /776	5 ('it is possible' (2), 'on the other hand' (2), 'there is no')845
Participant 10	2 ('on the other hand' (2)) /828	0/981
Participant 11	4 ('it is possible to' (2), 'it is impossible to', 'it is important to') /699	0/688
Participant 12	0/618	3 ('on the other hand' (2), 'there is no')/880
Total score	**9/9145**	**29/10699**
Mean	**0.75/762.08**	**2.33/891.58**
Standard deviation	**1.22/124.42**	**3.77/106.69**

Table 4.1 demonstrates that the students use very few chunks from the target items list in both pre- and post-test, except for participant 7, who appears to have acquired quite a few of them and demonstrates that is able to use them in essay writing in the post-test. This participant uses chunks from all the target items list, while the other participants use only chunks from the *academic collocations* list.

Table 4.2 shows the results from the essay task for experiment group 2.

Table 4. 2: Results of the essay task for experiment group 2

	Pre-test	Post-test
Participant 1	1 ('there is no') /842	2 ('it is possible to' (2))/765
Participant 2	0/1083	1 ('likely to')/1481
Participant 3	0/407	0/607
Participant 4	0/342	0/649
Participant 5	0/790	0/255
Participant 6	0/499	0/353
Participant 7	1 ('over a period of') /878	3 ('there is no', 'it is possible', 'over a period of')/841
Participant 8	1 ('whether or not') /843	2 ('there is no', 'his or her')/978
Participant 9	3 ('on the other hand' (2), 'there is no') /840	2 ('it is possible to' (2))/782
Participant 10	0/326	0/346
Participant 11	0/192	0/87
Participant 12	0/359	0/329
Participant 13	0/364	0/440
Total score	**6/7765**	**10/7913**
Mean	**0.46/597.31**	**0.77/608.69**
Standard deviation	**0.88/287.39**	**1.09/370.33**

Table 4.2 also shows very little use of chunks from the target items list. The few chunks that are used are from the list of *academic collocations*.

Table 4. 3: Results of the essay task for control group

	Pre-test	Post-test
Participant 1	0/621	1 ('it is impossible to')/742
Participant 2	0/528	0/387
Participant 3	0/849	0/712
Participant 4	0/603	0/617
Participant 5	0/642	0/690
Participant 6	1 ('there is no') /670	0/484
Participant 7	0/890	0/297
Participant 8	0/709	2 ('on the other hand' (2))/867
Total score	**1/5512**	**3/4796**
Mean	**0.125/689.00**	**0.38/599.50**
Standard deviation	**0.35/123.58**	**0.74/193.84**

Similar to Tables 4.1 and 4.2, Table 4.3 shows very little use of chunks from the target list. There are only a few academic collocations that are used.

The three tables demonstrate that the participants use more target items in their post-test essays than in their pre-test essays; however, the total numbers of occurrences (total scores) are still extremely low. In experiment group 1, the participants go from using 9 in the pre-test essays to using 29 in the post-test essays. The same numbers increase from 6 to 10 and from 1 to 3 in experiment group 2 and the control group, respectively. The high standard deviation value in experiment group 1 in the post-test results can be explained by a very high score of participant 7. This participant uses 13 target items in the post-test compared to 0 in the pre-test.

By comparing the increase in percentage from pre-test to post-test in the mean values of the four groups, we can form an impression of the effects of the treatment in this kind of tasks. Table 4.4 shows this comparison.

Table 4. 4: The increase in percentage in mean values in essay writing

Treatment group 1	Treatment group 2	Treatment group 1+2/2	Control group
165.54	64.26	114.90	249.38

As seen in Table 4.4, this increase is significantly higher in the control group. This is also reflected in Figure 4.1 below. The control group has the highest increase in mean score value in this test component, an increase of 249.38% compared to an increase of 114.90% in the treatment groups combined.

Figure 4. 1: Bar chart illustrating the increase in percentage in mean values from pre-test to post-test in essay writing
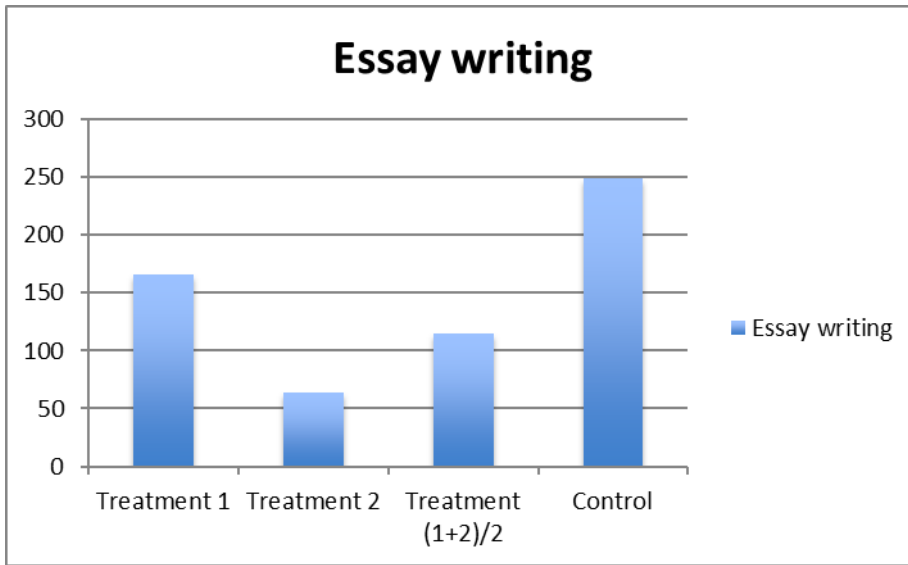
The results of this task do not meet the assumptions for univariate normality and homogeneity of variance. This can be attributed to unequal variance across the groups, possibly because of small sample sizes. The data meet the assumption of homogeneity of regression slopes, however, which means that there is a (sufficiently) similar pattern in the distribution of values across the groups. The ANCOVA for the essay-task (Table 4.5) produces a *p*-value of 0.177, which is much higher than the alpha level of 0.05 and therefore suggests that there are *not* significant differences in post-test essay-task scores by group, while controlling for pre-test scores. The effect size (η_p2) is 0.11 for this task.

Table 4. 5: ANCOVA for the essay task post-test scores by group while controlling for pre-test scores

Term	SS	df	F	p	η_p^2
Pre-test	0.00	1	0.14	.712	.01
Group	0.00	2	1.84	.177	.11
Residuals	0.00	29			

There are several reasons why the low scores in this test component are not surprising. First, the students were not made aware that these chunks were being assessed, in either the pre-test or post-test. Second, the target items list contains *academic* vocabulary, which is not frequently used among upper secondary students. Indeed, the students initially gave

the impression that a large number of these chunks were unknown to them. Third, the target items list is extremely small compared to the size of the vocabulary that we might expect the students to possess. Thus, if the list had included high-frequent words or expressions instead of academic ones, the students would probably still score very low. Fourth, and maybe most importantly, the essay task measures productive skills. As mentioned in Section 2.1.3, productive skills are supposedly more difficult to master than receptive skills. Any use of these chunks in the post-tests would require the student to know the form as well as the meaning of the chunk. It would also require the students to proactively use these expressions instead of alternative expressions that they might be more accustomed to using.

As mentioned above, only one participant in all the groups uses target items significantly more in the post-test compared to in the pre-test, namely, participant 7 in experiment group 1. One explanation for this might be that this participant learned more effectively than the others, either for motivational reasons or reasons related to his or her preference for learning method. If we see this increase in score as a proper reflection of an increase in level of productive skills, we could argue that learning here has happened in steps, not gradually. In other words, the participants need to achieve a certain level of understanding before being able to use their knowledge when producing text, such as in the case of participant 7. Another explanation is that the student in question understood what the test measured, that he or she 'saw through' the test design, and, unlike the other participants, that the student proactively used as many of these chunks as he or she could remember.

4.2 Test component 2: Highlighting

Tables 4.6–4.8 present the number of highlighted multi-word units compared to the overall number of words highlighted (including the words that are part of multi-word units).

Table 4. 6: Results of the highlighting task for experiment group 1

	Pre-test	Post-test	Pre-test	Post-test
Participant 1	6/38	1/25	15.8	4.0
Participant 2	1/5	1/20	20,0	5.0
Participant 3	3/81	9/66	3.7	13.6
Participant 4	1/14	6/47	7.1	12.8
Participant 5	1/43	11/79	2.3	13.9
Participant 6	1/21	3/40	4.8	7.5
Participant 7	0/5	13/39	0	33.3
Participant 8	0/18	6/39	0	15.4
Participant 9	0/17	9/67	0	13.4
Participant 10	0/11	1/38	0	2.6
Participant 11	2/25	8/34	8.0	23.5
Participant 12	1/6	15/79	16.7	19.0
Total score	**16/284**	**83/573**	**78.4**	**164.0**
Mean	**1.33/23.67**	**6.9/47.83**	**6.53**	**13.67**
Standard deviation	**1.72/21.77**	**4.78/20.10**	**7.22**	**8.76**

In Table 4.6, we can see that experiment group 1 highlights significantly more multi-word units in the post-test compared to the pre-test. Moreover, most participants highlight more words in total (more than double). Therefore, even though the participants highlight almost five times as many multi-word units, the *percentage* of multi-word units highlighted is only about twice the percentage highlighted in the pre-test.

Table 4. 7: Results of the highlighting task for experiment group 2

	Pre-test	Post-test	Pre-test	Post-test
Participant 1	0/18	0/16	0	0
Participant 2	9/38	9/35	23.7	25.7
Participant 3	4/65	6/68	6.2	8.8
Participant 4	0/25	0/24	0	0
Participant 5	0/15	0/13	0	0
Participant 6	0/6	0/2	0	0
Participant 7	1/31	0/28	3.2	0
Participant 8	2/20	7/64	10.0	10.9
Participant 9	1/12	2/32	8.3	6.3
Participant 10	1/19	1/20	5.3	5.0

Participant 11	1/33	2/59	3.0	3.4
Participant 12	1/16	0/23	6.3	0
Participant 13	2/48	12/99	4.2	12.1
Total score	**22/346**	**39/483**	**70.2**	**72.2**
Mean	**1.69/26.62**	**3.00/37.15**	**5.40**	**5.55**
Standard deviation	**2.46/16.29**	**4.10/27.42**	**6.42**	**7.52**

As seen in the table above, the difference between pre-test and post-test scores in treatment group 2 is less pronounced than it was in treatment group 1, and it represents only a slight increase when measured as a percentage of total words, from 5.40% to 5.55%.

Table 4. 8: Results of the highlighting task for the control group

	Pre-test	Post-test	Pre-test	Post-test
Participant 1	6/128	14/156	4.6	9.0
Participant 2	7/189	3/32	3.7	9.4
Participant 3	0/11	0/13	0	0
Participant 4	16/98	6/89	16.3	6.7
Participant 5	0/9	0/19	0	0
Participant 6	1/48	1/43	2.1	2.3
Participant 7	10/216	2/63	4.6	3.2
Participant 8	3/33	11/239	9.1	4.6
Total score	**43/732**	**37/654**	**40.4**	**35.2**
Mean	**5.38/91.5**	**4.63/81.75**	**5.05**	**4.40**
Standard deviation	**5.60/80.16**	**5.29/78.64**	**5.41**	**3.70**

As illustrated above in Table 4.8, the participants in the control group actually highlighted fewer multi-word units in the post-test than in the pre-test, both in total and as a percentage of the total number of words highlighted. The mean percentage is 5.05% in the pre-test, compared to 4.40% in the post-test. A comparison among the different groups regarding the increase in their mean values from pre-test to post-test is presented below in Table 4.9.

Table 4. 9: Increase in percentage in mean values in the highlighting task

Treatment group 1	Treatment group 2	Treatment group 1+2 /2	Control group
109.34	2.77	56.06	-12.87

This table demonstrates that the treatment groups experience a 56.06% increase, while the control group experiences a *decrease* of 12.87%. This comparison is also illustrated below in Figure 4.2.

Figure 4. 2: Bar chart illustrating the increase in percentage in mean values from pre-test to post-test in the highlighting task
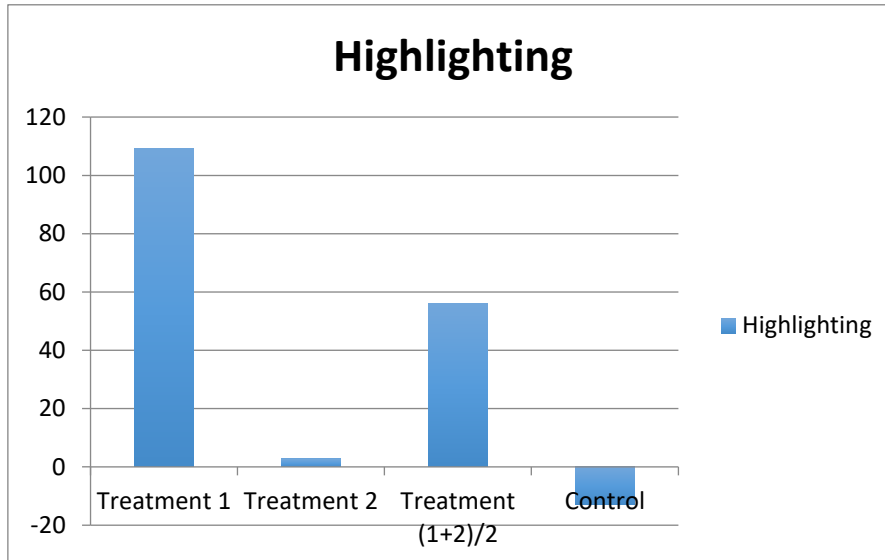
As we can see in Figure 4.2, treatment groups 1 and 2 differ significantly with respect to the increase in mean values from pre-test to post-test.

In this task, several of the participants score significantly lower in the post-test than they did in the pre-test, for example, participant 1 and 2 in experiment group 1 (Table 4.6), and participants 4 and 8 in the control group (Table 4.8). As such, the results are 'skewed'. This is verified by the fact that two of the assumptions for the ANCOVA—univariate normality and homogeneity of regression slopes—are not met. However, the assumption of homogeneity of variance is met. All in all, this suggests that the ANCOVA findings should be interpreted with a level of caution.

Table 4. 10: ANCOVA for highlighting task post-test scores by group while controlling for pre-test scores

Term	SS	df	F	p	η_p^2
Pre-test	124.40	1	2.42	0.131	0.08
Group	508.35	2	4.94	0.014	0.25
Residuals	1493.62	29			

The results of the ANCOVA are presented in Table 4.10. The *p*-value is 0.014, which suggests that *there are significant differences* in post-test highlighting-task scores by group while controlling for pre-test scores. The effect size is also relatively large here (0.25). Overall, this indicates that there are significant and important differences between the scores in the treatment groups and the scores in the control group in the post-test, when adjusted for the pre-test-scores.

While essay writing (test component 1) measures productive skills, the highlighting task measures receptive skills. As mentioned in Section 2.1.3, tasks that require receptive skills are often regarded as easier than tasks that require productive skills (cf. Laufer and Goldstein 2004). The participants did not have to produce text themselves, only mark those words that they considered important. One noticeable drawback with this task is the unclear instructions given to the students. It was up to each participant to interpret these instructions. Thus, some might have emphasised high-frequency words while others emphasised difficult words, or some might have emphasised single words while others emphasised multi-word units or both.

4.3 Test component 3: Gap filling

Tables 4.11–4.13 present the scores from the gap-fill task. All three groups score significantly higher in the post-tests. However, the participants in treatment group 1 made the greatest progress, scoring on average 7.83 higher in the post-test than in the pre-test. The participants in treatment group 2 scored on average 5.31 higher, whereas the control group scored on average 4.37 higher in the post-test. From the standard deviation values, we see that there are, on average, greater differences between the scores of the participants in experiment group 2 than there are in the two other groups. Despite a lower average score in this group, as many as four participants scored 54 points or higher in the post-test as opposed to only one participant in experiment group 1 and none in the control group.

Table 4.11 presents the results from treatment group 1. The mean value of the pre-test is 35.50, while in the post-test it is 43.33.

Table 4. 11: Results of the gap filling task for experiment group 1

	Pre-test	Post-test
Participant 1	42	50
Participant 2	34	45
Participant 3	34	41
Participant 4	9	20
Participant 5	29	28
Participant 6	35	44
Participant 7	48	51
Participant 8	38	40
Participant 9	33	56
Participant 10	50	50
Participant 11	37	42
Participant 12	37	53
Total score	**426**	**520**
Mean	**35.50**	**43.33**
Standard deviation	**10.34**	**10.46**

As we see in Table 4.11, the participants score relatively evenly, except for one participant (4), who has a pre-test score of 9. However, this student's score increased to 20 in the post-test. The standard deviation values reflect this relatively even distribution of scores.

Table 4.12 shows the scores by treatment group 2 in the pre-test and post-test. We can see that the mean values in these two tests are somewhat below the mean values from treatment group 1. However, the standard deviation values are significantly higher.

Table 4. 12: Results of the gap filling task for experiment group 2

	Pre-test	Post-test
Participant 1	51	55
Participant 2	57	58
Participant 3	39	42
Participant 4	27	29
Participant 5	51	36
Participant 6	9	28
Participant 7	49	54
Participant 8	41	56
Participant 9	24	43
Participant 10	15	18
Participant 11	38	36
Participant 12	8	16
Participant 13	30	37
Total score	**439**	**508**
Mean	**33.77**	**39.08**
Standard deviation	**16.45**	**14.05**

Table 4.13 presents the test results by the control group in the gap filling task.

Table 4. 13: Results of the gap filling task for control group

	Pre-test	Post-test
Participant 1	24	29
Participant 2	27	19
Participant 3	32	47
Participant 4	31	37
Participant 5	33	36
Participant 6	28	33
Participant 7	18	32
Participant 8	18	13
Total score	**211**	**246**
Mean	**26.38**	**30.75**
Standard deviation	**5.93**	**10.65**

As we can see from Table 4.13, the scores are quite evenly distributed. This is particularly apparent in the pre-test, in which the standard deviation value is as low as 5.93.

The comparison of the overall increase in mean values from pre-test to post-test is shown below in Table 4.14.

Table 4. 14: The increase in percentage in mean values in the highlighting task

Treatment group 1	Treatment group 2	Treatment group 1+2 /2	Control group
22.06	15.72	18.89	16.57

As evident in Table 4.14, the increase in mean values is greater in the treatment groups combined than in the control group, an 18.89% increase in the treatment groups as opposed to a 16.57% increase in the control group. The nuances between these differences in increases are highlighted in the bar chart in Figure 4.3 below as well.

Figure 4. 3: Bar chart illustrating the increase in percentage in mean values from pre-test to post-test in the gap filling task
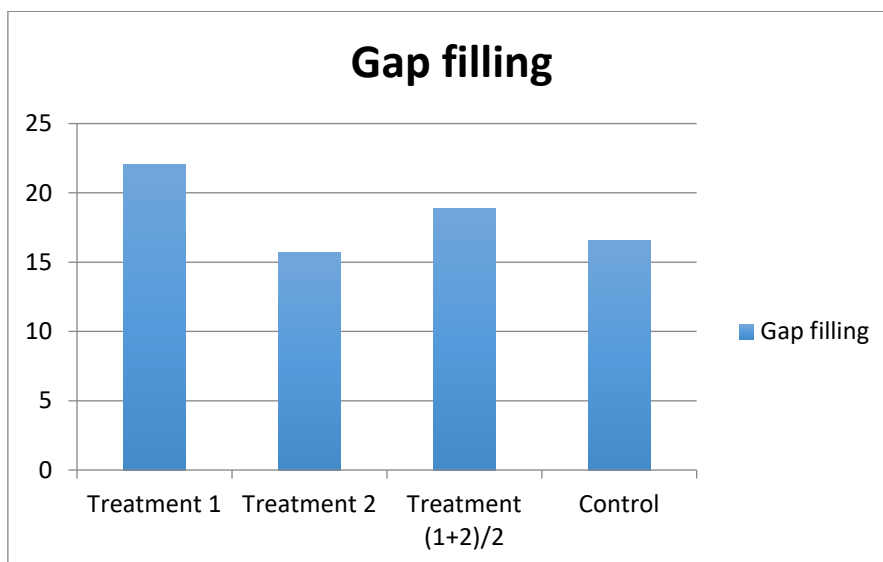
Treatment group 1 has the highest increase in mean values from pre-test to post-test, but the differences here are relatively small. We can see from the figure that treatment group 2 has the lowest increase.

All the assumptions required for the ANCOVA for the gap-fill task were met, indicating that we can be quite certain that the results from ANCOVA provide accurate measures of significance and effect size. The results of the ANCOVA are presented in Table 4.15.

Table 4. 15: ANCOVA for gap-fill task post-test scores by group while controlling for pre-test scores

Term	SS	df	F	p	η_p^2
Pre-test	2677.81	1	46.02	<0.001	0.61
Group	146.95	2	1.26	0.298	0.08
Residuals	1687.29	29			

As we can see in the table above, the *p*-value is 0.298, suggesting that there were *not* significant differences in post-test close-task scores by group while controlling for pre-test scores. The partial eta squared (effect size) is quite small (between 'small' and 'medium'), suggesting that the differences are not important.

4.4 Test component 4: Translation

Tables 4.16, 4.17 and 4.18 show the scores in the translation task. These tables show not only the participants' *total scores* in all three categories of chunks, but their scores in each of the respective categories as well – (1) academic collocations, (2) academic idioms, and (3) idioms that are 'suitable for CL-teaching'. As we can see, most participants score highest in the first category and lowest in the third category – in both pre-test and post-test, with some few exceptions. Looking at the total scores, we see that the participants in experiment group 1 score, on average, the highest. In this group, we can also see the greatest improvement in scores from pre-test to post-test: 13.08 points, as opposed to 4.69 in experiment 2 and 6.37 in the control group.

Table 4.16 shows the test results by treatment group 1 in the translation task.

Table 4. 16: Results of the translation task for experiment group 1

	Pre-test				Post-test			
	Academic collocations	Academic idioms	CL-suitable idioms	Total score	Academic collocations	Academic idioms	CL-suitable idioms	Total score
Participant 1	26	19	12	**57**	29	23	25	**77**
Participant 2	9	10	7	**26**	15	16	8	**39**
Participant 3	14	5	2	**21**	17	12	8	**37**
Participant 4	8	6	2	**16**	10	7	1	**18**
Participant 5	24	20	3	**47**	25	22	16	**63**
Participant 6	24	11	6	**41**	29	18	12	**59**
Participant 7	27	20	13	**60**	24	22	21	**67**
Participant 8	22	12	6	**40**	20	17	13	**50**
Participant 9	17	18	4	**39**	24	23	7	**54**
Participant 10	18	6	6	**30**	25	15	7	**47**
Participant 11	17	15	2	**34**	21	22	11	**54**
Participant 12	28	12	7	**47**	27	15	8	**50**
Total score	234	154	70	**458**	266	212	137	**615**
Mean	19.50	12.83	5.83	**38.17**	22.17	17.67	11.42	**51.25**
Standard deviation	6.78	5.56	3.66	**13.54**	5.82	5.02	6.63	**15.39**

As we can see in this table, treatment group 1 experiences a significant increase in mean value, from 38.17 in the pre-test to 51.25 in the post-test.

The next table (Table 4.17) highlights the test results from experiment group 2.

Table 4. 17: Results of the translation task for experiment group 2

	Pre-test				Post-test			
	Academic collocations	Academic idioms	CL-suitable idioms	Total score	Academic collocations	Academic idioms	CL-suitable idioms	Total score
Participant 1	18	10	8	**36**	22	13	12	**47**
Participant 2	26	24	15	**65**	28	28	23	**79**
Participant 3	23	9	3	**35**	23	13	0	**36**
Participant 4	18	12	5	**35**	10	15	7	**32**
Participant 5	21	13	7	**41**	17	6	1	**24**
Participant 6	5	5	0	**10**	7	8	0	**15**
Participant 7	23	17	6	**46**	24	19	18	**61**
Participant 8	28	15	8	**50**	27	16	14	**57**
Participant 9	21	22	2	**45**	24	16	7	**47**
Participant 10	12	3	0	**15**	12	1	0	**13**
Participant 11	20	16	7	**43**	18	17	8	**43**
Participant 12	3	0	0	**3**	2	4	0	**6**
Participant 13	11	9	6	**26**	22	16	13	**51**
Total score	229	155	67	**450**	236	172	103	**511**
Mean	17.62	11.92	5.15	**34.62**	18.15	13.23	7.92	**39.31**
Standard deviation	7.71	7.02	4.24	**17.26**	8.10	7.11	7.69	**21.05**

In Table 4.17 there are several test scores that deviate in an obvious manner from the mean value from the combined test categories. For example, participant 2's total score of 65 clearly contrasts participant 12's total score of 3 in the pre-test. The standard deviation value of 17.26 confirms the heterogenous test results illustrated by these two examples.

Table 4.18 display the test results of the control group in the translation task.

Table 4. 18: Results of the translation task for control group

	Pre-test				Post-test			
	Academic collocations	Academic idioms	CL-suitable idioms	Total score	Academic collocations	Academic idioms	CL-suitable idioms	Total score
Participant 1	17	9	1	**27**	10	8	6	**24**
Participant 2	10	16	3	**29**	15	9	6	**30**
Participant 3	16	5	7	**28**	23	12	9	**44**
Participant 4	17	14	7	**38**	18	18	16	**52**
Participant 5	12	9	3	**24**	6	7	10	**23**
Participant 6	18	15	0	**33**	22	11	8	**41**
Participant 7	7	6	3	**16**	16	6	3	**25**
Participant 8	4	11	7	**22**	13	8	8	**29**
Total score	101	85	31	**217**	123	79	66	**268**
Mean	12.63	10.63	3.88	**27.13**	15.38	9.88	8.25	**33.50**
Standard deviation	5.24	4.10	2.80	**6.73**	5.76	3.83	3.81	**10.78**

As is shown in Table 4.18, there are no really high numbers in the scores in either pre-test nor post-test. However, none of the participants score as low as the lowest-scoring participants of treatment group 2 either.

When comparing the three tables, we also see that the average improvement in score is not the same for the three categories of chunks. The improvement is lowest in category 1, academic collocations, and highest in category 3, the CL-suitable idioms. We can find this pattern in all three groups; however, it is more significant in the control group, which has an average improvement of 4.37 in the third category of chunk as opposed to -0.75 in the second category (academic idioms) and 2.75 in the first category (academic collocations).

It is also worth mentioning that we find the lowest and the highest total scores among participants in treatment group 2 and that the standard deviation value is higher in this group than in the two other groups. This result adds to the pattern that we find in test component 2 and 3, that is, a higher standard deviation value in experiment group 2. Table 4.19 presents a comparison of the increase in mean (total) values between the groups from pre-test to post-test.

Table 4. 19: The increase in percentage in mean values in the translation task

Treatment group 1	Treatment group 2	Treatment group 1+2 /2	Control group
34.27	13.55	23.91	23,48

As we can see from Table 4.19, the increases in the mean values from pre-test to post-test are minimal if we compare the combined value from the treatment groups to the control group. However, here, we can see a huge gap between the two treatment groups. This is also visible in the bar chart below (Figure 4.4).

Figure 4. 4: Bar chart illustrating the increase in percentage in mean values from pre-test to post-test in the translation task
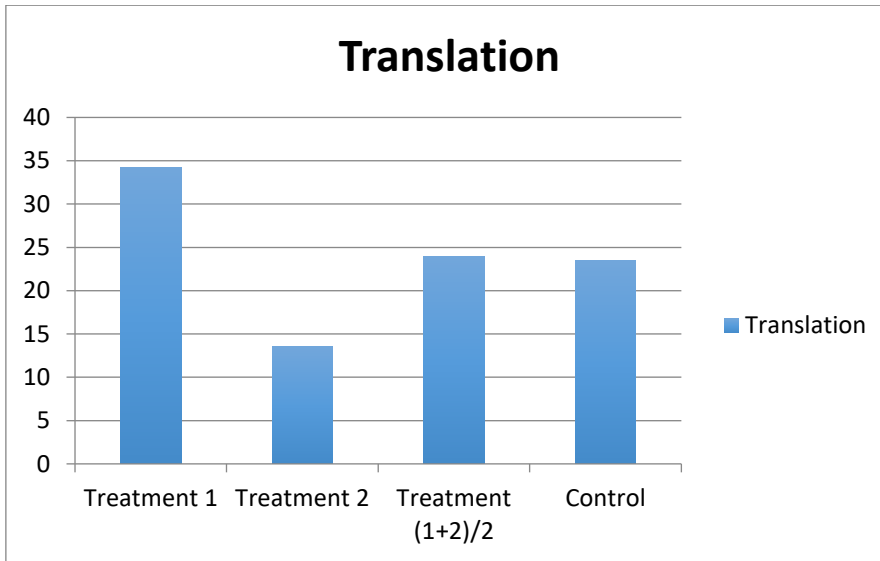
We can see from Figure 4.4 that the increases in mean values from pre-test to post-test are very different in the two treatment groups regarding this test component.

The data from the translation task met all three assumptions required for the ANCOVA, suggesting reliable results from the ANCOVA measurement procedure because of consistent patterns in score between the experiment groups and the control group. The results of the ANCOVA are presented in Table 4.20. The *p*-value obtained by the ANCOVA is 0.067, which indicates that the differences in post-test translation-task scores by group—while controlling for pre-test scores—were not significant. However, 0.067 is not far from the alpha level (0.05), and the effect size is 0.17, which can be defined as relatively large.

Table 4. 20: ANCOVA for the translation task post-test scores by group while controlling for pre-test scores

Term	SS	df	F	p	η_p^2
Pre-test	6734.39	1	97.52	<0.001	0.77
Group	420.85	1	3.05	0.063	0.17
Residuals	2002.63	2			

In addition to the total mean values in pre-test and post-test (in bold) in the three tables 4.16–4.18, we have category-specific mean values: mean values in scores for academic

collocations, academic idioms, and CL-suitable idioms. Table 4.21 illustrates the increase from pre-test to post-test in the category-specific mean values for the respective groups.

Table 4. 21: Increase in category-specific mean values from pre-test to post-test in the translation task for the experiment groups and the control group.

Group	Category 1 (academic collocations)	Category 2 (academic idioms)	Category 3 (Idioms 'suitable' for CL-inspired teaching)
Experiment group 1	13.7%	37.7%	95.9%
Experiment group 2	3.0%	11.0%	53.8%
Control group	21.8%	-3.6%	112.6%

Table 4.21 shows that all scores increase from pre-test to post-test, except the score of the control group in category 2.

4.5 Summary

By comparing the mean scores of the two experiment groups (means added together and divided by two) to the mean score in the control group, we find that in 3 out of 4 test components, the difference in pre-test scores and post-test scores follow quite a similar pattern. Table 4.22 and Figure 4.5 below show as percentages the groups' increase from pre-test to post-test in mean score values in the respective test components.

Table 4. 22: Overall increase in mean values from pre-test to post-test in all test components

	Essay writing	Highlighting	Gap filling	Translation
Treatment 1	165.54	109.34	22.06	34.27
Treatment 2	64.26	2.77	15.72	13.55
Treatment 1+2 /2	114.90	56.06	18.89	23.91
Control	249.38	-12.87	16.57	23.48

Table 4.22 shows several patterns in the increases in mean values. Treatment group 1 scores consistently higher than treatment group 2. Furthermore, the increases in mean values in the control group varies substantially depending on which test component we examine. Even though this is a useful illustration, presenting these numbers side by side can lead to some misinterpretation. It is important to note that the test components are

very different and that a large value in one of the tasks may not be considered a large number in another task. For instance, the increase of 249.38 percent for the control group in the essay writing task would be virtually impossible to achieve in some of the other tasks. This must also be noted when examining Figure 4.5 below, which illustrates these numbers.

Figure 4. 5: Bar chart illustrating the increase in percentage in mean values from pre-test to post-test in all test components
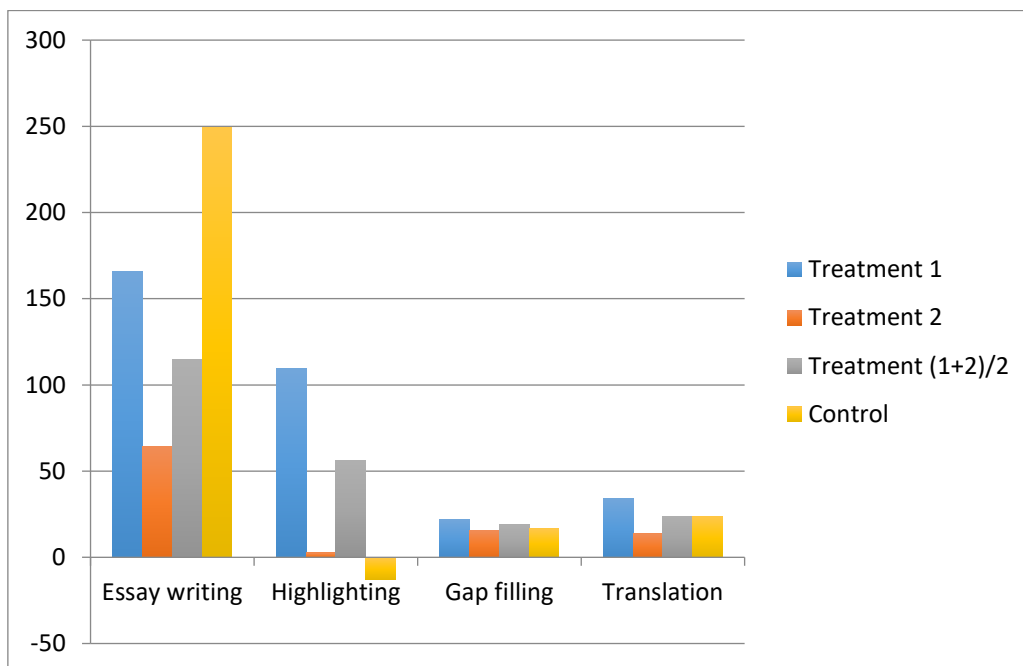
We see that the blue bars, belonging to treatment group 1, are taller than the orange ones all across the board. The yellow bars vary the most, given the high scores on the essay writing task. The highlighting task clearly stands out among these numbers. The results from this task indicate a significant increase from pre-test to post-test in the treatment groups and a *decrease* in the control group. The results indicate that the effects of CL-inspired teaching on the participants' awareness of multi-word units are significant. However, the results from the three other test components show an equally large increase among the treatment groups and the control group, indicating that the treatment effects are not significant.

Overall, the results from experiment groups 1 and 2 differ significantly. The isolated scores by the participants in experiment group 1 display high mean values in all

test components, especially in the highlighting task. Here, experiment group 1 has an increase of more than a 100 percent, while experiment group 2 experiences no significant increase and the control group actually experiences a *decrease* in the number of highlighted words.

The scores from the participants in experiment group 2 display a tendency to be *more heterogenous* than those of the participants in experiment group 1 and the control group. This is reflected in higher standard deviation values in test components 2, 3 and 4. In test component 1, the essay task, the standard deviation values for both pre-test and post-test are higher for experiment group 1. However, this may be explained by the fact that there are extremely low scores in all groups, and that the slightly higher scores by experiment group 1 creates a potential for deviance which is not present for the other groups. The exceptionally high score by participant 7 in the post-test adds greatly to this general impression.

5. Discussion

5.1 What kind of language proficiency benefits the most from CL-inspired vocabulary teaching?

The four components of the pre- and post-test measure different sets of skills in language proficiency. The first three test components require the participants to decide on the form of the vocabulary, while the translation task requires retrieval of meaning – what Laufer and Goldstein (2004) refer to as passive recall. Although the essay-writing, the gap-filling, and the translation task all require productive skills from the participants, the essay task is the most open and comprehensive of these. The highlighting task requires no productive skills, but it can arguably indicate something about the participants' ability to 'recognise form' in vocabulary (cf. Laufer and Goldstein 2004, see section 2.1.3). On the basis of the test results, we can draw some conclusions as to what kinds of learning skills CL-inspired vocabulary teaching promotes. Since the highlighting task was the only one which indicated significant results from the treatment, one might claim that CL-inspired teaching is most effective in creating understanding and awareness of L2 form.

5.2 Are academic chunks 'suitable' for CL-inspired vocabulary teaching?

According to Boers and Lindstromberg (2008a), CL-inspired teaching has proved effective when applied to certain 'suitable' idioms. The translation task (test component 4) in my experiment measures the students' knowledge of academic chunks in the form of academic collocations (category 1), academic idioms (category 2), and idioms that are listed as suitable for CL-inspired teaching by Boers and Lindstromberg (2008a) (category 3). The comparison between these three categories can provide some answers as to whether academic vocabulary is as suitable for CL-inspired teaching as non-academic vocabulary and whether collocations are as suitable for CL-inspired teaching as idioms.

Table 4.21 shows a greater increase from pre-test to post-test in the mean values for the third category (CL-suitable idioms) compared to the first two categories. In experiment group 1, there is a 95.9 percent increase, in experiment group 2 there is a 53.8 percent increase, and in the control group there is a 112.6 percent increase. The fact that this increase is significant yet equal in the experiment groups and the control group might indicate that this kind of chunk is, in general, easier to retain. One interpretation of the

increase in mean value from pre-test to post-test in the control group is that the mere introduction of these 'suitable' figurative idioms in the pre-test prompted the participants to look up the items and reflect on them, so that they achieved higher scores in the post-test. Alternatively, an unknown or untested factor, such as increased attention from the teacher or the fact that the post-test was administered later on in the school year, might have led the participants from all groups to perform better in the post-test. Another possibility is that the participants in the experiment groups were more motivated because of the CL-inspired teaching, in which case the participants in the control group actually learned more from traditional teaching than from CL-inspired teaching, since they performed quite similarly in the post-test. However, the participants in the experiment groups may have 'suffered from boredom': Since they were frequently introduced to various kinds of chunks, they might have been less motivated, in which case the CL-inspired teaching promoted learning but the participants simultaneously became less motivated due to the learning burden. In any case, the present experiment does not confirm the claim that the kind of chunks included in category 3 is more suitable for CL-inspired teaching than academic collocations and academic idioms.

As a matter of fact, Table 4.21 shows that, while the experiment groups score similarly to the control group in both category 1 and category 3, category 2 differs in this regard. The results seem to indicate that CL-inspired teaching has little effect on the retention of academic collocations and 'suitable' idioms, but significant effects on the retention of *academic idioms*. In this category, the mean value for the control group scores actually decreases from pre-test to post-test, while the mean values for the experiment groups increase. The reason for this could be that the retention of academic idioms requires more effort than academic collocations because they are less transparent and compositional (cf. transparency in chunks, Section 2.2.4). At the same time, they do not have the obvious figurative content that many of the suitable idioms have, which makes it more difficult to guess their meaning without guidance. These features may then lead to a greater need for the guidance provided by the CL-inspired teaching, which may explain why the treatment groups did better than the control group.

5.3 Differences between homogenous and heterogenous groups in CL-inspired teaching

We established in Section 4.5 that the differences in test results between the two treatment groups are significant, considering that treatment group 1

(1) has higher mean scores

(2) is more homogenous with regard to score values (i.e., low standard deviation values)

(3) has a higher mean increase (from pre-test to post-test)

The differences between the groups regarding the two first variables are not surprising in themselves. After all, students, similar to people everywhere, have different skills, interests, and backgrounds, and it is only natural that there are systematic differences in scoring patterns between groups that are composed randomly (often by name or local address), such as in this case. However, the differences regarding the third variable, mean increase in scores, are equally high between these groups. This may lead us to question whether the participants in the two groups may have reacted differently in response to the treatment, that is, CL-inspired learning activities, based on their starting (pre-test) levels or based on the level of homogeneity in the groups. Could it be that learners who are part of more proficient or more homogenous groups in the first place respond better to the kinds of learning activities introduced in the present experiment? After all, the learning activities required the students to engage actively in the teaching sessions, and the participants of the experiment might very well have benefitted from the collaborative abilities among their peers.

5.4 Reasons for deviating scores by participant

A closer look at one of the 33 participants' scores in the four tasks might provide some interesting findings. The scores by participant 7 in treatment group 1 deviate substantially from the general scoring patterns in the groups.

In the post-test of the essay task, participant 7 in experiment group 1 is the only one who uses the target items to a 'noticeable' extent, namely, 13 times. This participant does not use any target items in the pre-test, and in the highlighting task, the participant's scores follow the same pattern there as well; he or she highlighted 0 target multi-word

units in the pre-test, compared to 13 multi-word units in the post-test. In the two last test components, however, the participant scored quite similar in the pre-test and post-test, at 48 and 51 in the gap-fill task and 60 and 67 in the translation task.

Participant 7's scores in the essay-task and the highlighting task could possibly indicate that he or she did not make any effort or had a 'bad day at the office' in the pre-test. However, this is contradicted by the results in the two other tasks, where the participant scores well above average. Given this fact, a more plausible explanation might be that the student reached a higher awareness of MWUs, which became evident in these two test components, and that students who score consistently high on the two other tasks are likely to achieve such awareness. However, there are several participants whose scoring patterns contradict such a hypothesis. Participant 2 in experiment group 2 has the highest average scores of all participants in the two last tasks: 57 and 58 in the pre-test and post-test of the gap-fill task (Table 4.12), and 65 and 79 (Table 4.17) in the translation task, respectively. The same participant highlights 9 multi-word units in both pre-test and post-test, that is, about 25% of the total number of words in both instances. However, in the essay task, the participant uses none of the target items in the pre-test and only one target expression in the post-test. Participant 1 in experiment group 1 is another example. He or she scored 1 and 0 in the pre-test and the post-test of the essay task and 6 and 0 in the highlighting task, respectively. The same participant scored 42 and 50 in the pre-test and post-test of the gap-fill task and 57 and 77 in the translation task, respectively. The scoring pattern of these participants show that a high score in the two last tasks does not necessarily point to an increase in scores in the two first tasks.

Another explanation for the deviating test results is that the participant became aware of the MWUs because he or she saw through the inherent deception in the task instructions. In both tasks, the instructions did not, or at least were not supposed to, give away the real research goal. In the essay task, the participants were merely asked to write essays based their opinions about two documentaries and in the highlighting task the participants were instructed to highlight words *or* expressions that would be useful for a younger student. If participant 7, unlike the other participants, understood that the measured dependent variable in both tasks was the number of MWUs used, his or her test results are not influenced by the 'deception variable' that the other participants' test

results are. If this was the case, it would certainly explain the participant's deviating results in the two first tasks.

5.5 Comparison to similar studies

This study is inspired by and also replicates several of the tests undertaken by Jones and Haywood (2004). Similar to their study, my study tests the production of chunks in an essay, awareness of chunks in a highlighting task and form recall in a (modified) Cloze test (i.e., a gap-fill test). However, there are some important differences between the studies. There is no translation task in the Jones and Haywood's study, but Jones and Haywood did include an interview with some of the participants, which is not included in the present study. Another difference concerns the selection of vocabulary test items: Jones and Haywood tested different items in the pre-test and post-test, while in the present study, the same items were tested. There is also a difference in how the essay-task was tested. Jones and Haywood had five experienced teachers of EAP (English for Academic Purposes) review the student's essays and decide how many chunks they used. This is, of course, a more thorough test method which provides a more complete picture than the test method used in the present study.

Because of these differences, the two studies faced different challenges. A translation task might have provided useful information to the Jones and Haywood-study, whereas an interview of the test participants could have proved useful in this study. Furthermore, if the test items in Jones and Haywood's post-tests turned out to be easier or more difficult than the ones in the pre-test, the participants might score differently because of this. In the present study, the fact that the test items were the same probably made a significant impact on the scores as well. However, both studies included a control group in order to detect these potential biases.

The scores by the participants in the experiment groups in Jones and Haywood's (2004) study indicated a notable increase in awareness. Six out of the ten participants even highlighted more sequences than words in the post-test when they had highlighted more words than sequences in the pre-test. This coincides with the significant result in the present study.

The results of Jones and Haywood's (2004) essay task also coincide with the results of the present study. With one exception, none of the participants experienced noticeable progress from pre-test to post-test. According to Jones and Haywood (2004:

285), there is 'a disappointing lack of apparent improvement in terms of the use of phraseology in the students' essays'.

6. Conclusion

This chapter summarises the preceding chapters, and revisits the research question posed in the introduction of the thesis. Furthermore, I examine possible pedagogical implications of the findings in this study. Finally, I present some thoughts about potential topics and concerns for further studies.

6.1 Summary

Cognitive Linguistics offers numerous avenues to research vocabulary learning and teaching. In the preceding chapters, I have argued that this relatively new linguistic approach provides the rationale for targeting chunks in teaching and learning. Moreover, the rather frequent use of certain chunks in academic discourse and the pedagogical benefits of the genre-specific approach to learning chunks (cf. Section 2.3.4) demonstrate the relevance of academic vocabulary. Academic vocabulary is also particularly relevant to the group of learners that was targeted in this experiment, namely, ESL students in upper secondary school.

From this framework, I set out to explore some of the pathways to 'chunk-learning' offered by cognitive linguistics. I reviewed research that shows that linguistic motivation can explain a large part of the semantic and structural features of chunks, which opens a number of strategies to teaching and learning different kinds of chunks. The CL theory revolving around conceptual metaphor further introduces new ways of categorising chunks based on source domains or target domains, as explained in Section 2.4.3. There are also important cognitive processing theories that underline the effectiveness of 'deep processing', 'entrenchment', and the use of different learning input (see Section 2.4.4). These underlying principles are apparent in elaboration on different meaning- and form-connections, as outlined in Section 2.4.5. Additionally, the principles are reflected in the organisation of vocabulary teaching into the three stages *noticing*, *memorising*, and *reviewing* (cf. Lindstromberg and Boers 2008a).

The theory on chunks, academic vocabulary, and Cognitive Linguistics that I had reviewed in Chapter 2, then, indicated that we can apply principles of Cognitive Linguistics in teaching and learning academic chunks by engaging the students in elaboration on form- and meaning-connections, and by making sure that the students notice, memorise, and review chunks according to the three-stage teaching model by

Lindstromberg and Boers (2008a). Activities in the first stage pertain to creating awareness of word strings and patterns of semantic or structural features (e.g., call attention to collocative patterns, alliteration, or figurative meaning). The next stage involves elaboration on different features, such as sorting idioms based on their source domains or examining metaphor in familiar text (e.g., song lyrics). Finally, it is important to consolidate the knowledge by reviewing the chunks repeatedly and using different kinds of input.

The next step was to implement these learning activities in a series of teaching sessions and, subsequently, to analyse the outcome of this pedagogical undertaking. The teaching sessions were carried out in three classes of Norwegian L2-students in upper secondary school. To document potential effects of the teaching sessions, tests were carried out before and after the teaching sessions in two treatment groups and one control group. As discussed in Chapter 3, the research design of this teaching experiment presented several methodical issues, specifically with regards to validity. Given the fact that chunks are 'ubiquitous' in language (cf. Schmitt 2015: 117, see also Section 2.2.2), 148 target items was probably not enough to measure the participants' productive formulaic language adequately in test component 1. Furthermore, while it is possible to control for the fact that the participants had gone through the tasks in pre-test before the post-test, by the inclusion of a control group, it is difficult to judge with certainty what treatment the participants picked up in the course of the treatment of CL-inspired teaching. Since the CL-teaching introduced some of the same target items that were tested in the post-test, it is not certain whether the treatment effects represent awareness from deep-learning (see Sections 2.4.4–2.4.5) or simply the memorisation of chunks.

The results of the experiment were presented in tables, and the average scores were calculated for each test component for each group (treatment groups and control group). The analysis of the test results also presented probability values (p-values) to indicate statistical significance. The p-values, in this case generated by the use of ANCOVA-tests in SPSS, estimate the probability that the null-hypothesis (see Section 3.3.2) is correct. By convention, p-values of less than 0.05 indicate statistically significant test results in applied linguistics. The summary of the results indicated the following scoring patterns:

(1) The treatment groups had consistently higher scores than the control group in test component 2 (highlighting); the increase in scores from pre-test to post-test were found to be statistically significant.

(2) The treatment groups did not have consistently higher scores than the control group in test component 1 (essay writing), 3 (gap filling), and 4 (translation); the test results were not found to be statistically significant.

(3) Treatment group 1 had consistently higher scores than treatment group 2 and the control group in test components 2, 3 and 4.

(4) Treatment group 2 had a more heterogenous scoring pattern than that of treatment group 1 and the control group in test components 2, 3 and 4.

(5) The scores by one particular participant (participant 7, treatment group 1) deviated highly from the others in test components 1 and 2.

These scoring patterns were discussed in Chapter 5, first concerning what kind of language proficiency benefits from CL-inspired teaching. It was pointed out that the only test component that displayed a statistically significant increase in scores from pre-test to post-test was the highlighting task. Since this task measured mainly receptive language skills and the tasks in the other test components also measured productive language skills, it was suggested that CL-inspired teaching possibly favours the learning of receptive language skills.

The second section discussed whether academic chunks were suitable for CL-inspired teaching in view of potential evidence from the test results. This was done by comparing the increase in scores from pre-test to post-test in the different target item categories in test component 4, the translating task. The comparison showed that academic idioms was the only category in which a statistically significant increase in scores from pre-test to post-test was found. This led to the conclusion that academic chunks were equally suitable, or, in the case of academic idioms, more suitable, than non-academic chunks used in previous CL studies.

The third discussion concerned the potential effects from the level of group homogeneity on CL-inspired teaching. The test results indicated that the scores by the participants in treatment group 1 were more homogenous (indicated by consistently lower values of standard deviation in this group) than the scores by the participants in treatment group 2. Since this treatment group 1 also had a consistently higher increase in scores

from pre-test to post-test, it could be argued that groups that are homogenous with respect to level of proficiency benefit more from CL-inspired teaching than groups that are heterogenous.

The fourth discussion examined some of the score values of participant 7 in treatment group 1, which were interesting because they deviated noticeably from the overall scoring patterns. It was suggested that this participant possibly 'saw through' the somewhat misleading instructions that were given to the students in the two first test components while the other students did not. Therefore, he or she might have used more multi-word units or chunks than usual.

The fifth discussion compared the test results in this test to the ones in other, similar studies. In the case of Jones and Haywood's (2004) study, the test results coincided significantly in the essay task and the highlighting task.

6.2 Answer to the research question

The preceding chapters aimed to bring forth the necessary insights to answer the research question. An attempt was made to describe the characteristics of and theories related to second language vocabulary acquisition, academic vocabulary, chunks, and Cognitive Linguistics. Furthermore, some of the pedagogical implications derived from intersecting areas of these approaches were explored through literature, and CL-inspired teaching strategies were introduced. The experiment conducted here, then, sought to employ these teaching strategies. The discussion of the test results from the experiment has already highlighted some important findings in this study. I elaborate more on these findings in the following endeavour to answering the research question posed in the introduction:

How does teaching inspired by Cognitive Linguistics compare to traditional teaching, in terms of recognition, understanding and use of academic chunks among Norwegian ESL students in upper secondary school?

The statistical inferential analysis indicates that CL-inspired teaching benefits the awareness of academic chunks (as shown in a highlighting task) more than traditional teaching does. At the same time, CL-inspired teaching does not seem to aid the retention of academic chunks or benefit the skill of using them (as shown in essay writing, gap filling, and translation) more than traditional teaching.

The findings also suggest that participants who are more proficient (have higher mean test scores) and are part of groups that are more homogenous regarding proficiency level (as indicated by lower standard deviation values in test scores) seem to experience greater benefits from CL-inspired teaching (higher increase in mean test score values from pre-test to post-test). However, the test design of this experiment does not provide the tools to determine whether there is not only a correlating relationship but also a causal relationship between these variables. Therefore, we cannot with certainty claim that participants in homogenous and proficient groups benefit more from CL-inspired teaching.

This study also indicates that chunks from academic discourse, specifically, academic idioms, make for a suitable kind of target vocabulary in CL-inspired teaching. Academic collocations and idioms claimed to be suitable in previous studies, were found to be less suitable. Section 5.2 discussed reasons for this, which suggested that academic idioms may be more difficult to learn than the two other categories while at the same time having great potential elaborative tasks, as might be provided in the CL-inspired teaching.

The findings from this teaching experiment leaves much to be discussed. Chapter 3 presented some of the methodological issues regarding validity. Test component 1, essay task, suffers from validity threats specifically linked to construct validity. It is highly uncertain whether this test component provided valuable information about the students' ability to produce academic chunks, since the test measured the use of such a small number of vocabulary items compared to the overall number of chunks in the English vocabulary. This is an important note, because the test results from this component differ greatly from those from the other components; that is, the isolated results from tests 2, 3 and 4 indicate greater benefits from CL-inspired teaching.

Overall, the findings from this experiment do not indicate significant differences between CL-inspired teaching and traditional teaching in terms of the use of academic chunks among Norwegian upper secondary ESL students. The findings do, however, indicate that there are some benefits from CL-inspired teaching regarding these students' awareness of academic chunks.

6.3 Pedagogical implications

The findings might indicate that the treatment in the form of CL-inspired teaching did not stimulate the students' *use* of vocabulary sufficiently. Indeed, many of the activities in

the teaching sessions ended up as conversations between the teacher and the students, and sometimes between the students. There is a possibility that the focus on 'elaboration' did not include enough exercises relevant to the training of productive language skills. A solution to this could be to add more writing tasks or speaking tasks to the student activities in the teaching sessions.

It could also be argued here that the short amount of time spent on the teaching sessions was not sufficient to induce learning effects pertaining to the possibly more difficult aspects of language mastery, such as productive language skills. These five teaching lessons might have been enough to give the students some basic understanding of what chunks are, which led to the small increase in recognition of possible chunks. Thus, the five teaching lessons could be suitable as an introduction to the topic, with additional lessons focussing more on productive language skills.

The findings from this study suggested that homogenous groups of learners and proficient learners benefitted the most from CL-inspired teaching. This may suggest that the general lines of studies in upper secondary school might prove more successful than teaching in the vocational lines of studies (as in this study). It could also point to the benefits of using differentiated classes when carrying out CL-inspired teaching sessions.

6.4 Suggestions for future research

As pointed out in Section 6.2, the findings indicate a correlative relationship between homogenous student groups, proficient student groups, and learning benefits from CL-inspired teaching. However, one would have to use a more suitable research design than the one in this study to decide whether there are causal relationships between these three variables. The finding of such causality would imply that group composition is a more important factor in planning CL-inspired teaching than in traditional teaching.

Another aspect of interest is the individual differences related to the use of chunks among learners. A study by Dörnyei et al. (2004) points to the fact that the use of certain chunks, such as colloquial phrases, varies highly among different individuals, and these differences do not always relate to the skill level of the individuals. Further studies on the different 'styles' or preferences among different learners could provide interesting knowledge on acquisition of academic language, and academic chunks.

References

Alali, F. A. and Schmitt, N. (2012) Teaching Formulaic Sequences: The Same as or Different From Teaching Single Words? *TESOL Journal* 3 (2): 153–180.

Albrechtsen, D., Haastrup, K., Henriksen, B. and Cumming, A. H. (2008) *Vocabulary and writing in a first and second language: Processes and development.* Houndmills, Basingstoke, Hampshire, New York, N.Y.: Palgrave Macmillan.

Alderson, J. C. (2006) *Diagnosing foreign language proficiency: The interface between learning and assessment.* London, New York: Continuum.

Andrews, R. (2003) *Research questions.* London, New York: Continuum.

Batstone, R. and Ellis, R. (2009) Principled grammar teaching. *System* 37 (2): 194–204.

Baumann, J. F. and Graves, M. F. (2010) What Is Academic Vocabulary? *Journal of Adolescent & Adult Literacy* 54 (1): 4–12.

Beréndi, M., Csábi, S. and Kövecses, Z. (2008) Using conceptual metaphors and metonymies in vocabulary teaching. In F. Boers and S. Lindstromberg (eds). *Cognitive linguistic approaches to teaching vocabulary and phraseology*, 65–98. Berlin: Mouton de Gruyter.

Bobrova, L. and Lantolf, J. P. (2012) *Metaphor and pedagogy.* http://calper.la.psu.edu/sites/default/files/pubfiles/CALPER_WP11_Metaphor_and_Pedagogy.pdf. Accessed 7/25/2017.

Boers, F., Deconinck, J. and Lindstromberg, S. (2010a) Choosing motivated chunks for teaching. In S. d. Knop and F. Boers (eds). *Fostering language teaching efficiency through cognitive linguistics*, 239–256. Berlin: De Gruyter Mouton.

Boers, F., Demecheleer, M. and Eyckmans, J. (2004) Etymological elaboration as a strategy for learning idioms. In P. Bogaards and B. Laufer-Dvorkin (eds). *Vocabulary in a second language: Selection, acquisition, and testing / edited by Paul Bogaards, Batia Laufer*, 53–78. Amsterdam, Philadelphia: John Benjamins Pub.

Boers, F. and Lindstromberg, S. (2005) Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. *System* 33 (2): 225–238.

Boers, F. and Lindstromberg, S. (2008a) From empirical findings to pedagogical practice. In F. Boers and S. Lindstromberg (eds). *Cognitive linguistic approaches to teaching vocabulary and phraseology*, 375–393. Berlin: Mouton de Gruyter.

Boers, F. and Lindstromberg, S. (2008b) How cognitive linguistics can foster effective vocabulary teaching. In F. Boers and S. Lindstromberg (eds). *Cognitive linguistic approaches to teaching vocabulary and phraseology*, 1–51. Berlin: Mouton de Gruyter.

Boers, F. and Lindstromberg, S. (2008c) Structural elaboration by the sound (and feel) of it. In F. Boers and S. Lindstromberg (eds). *Cognitive linguistic approaches to teaching vocabulary and phraseology*, 329–353. Berlin: Mouton de Gruyter.

Boers, F. and Lindstromberg, S. (2009) *Optimizing a lexical approach to instructed second language acquisition*. Basingstoke: Palgrave Macmillan.

Boers, F., Rycker, A. D. and Knop, S. d. (2010b) Fostering language teaching efficiency through cognitive linguistics: Introduction. In S. d. Knop and F. Boers (eds). *Fostering language teaching efficiency through cognitive linguistics*, 1–26. Berlin: De Gruyter Mouton.

Boers, F. and Stengers, H. (2008a) A quantitative comparison of the English and Spanish repertoires of figurative idioms. In F. Boers and S. Lindstromberg (eds). *Cognitive linguistic approaches to teaching vocabulary and phraseology*, 355–374. Berlin: Mouton de Gruyter.

Boers, F. and Stengers, H. (2008b) Adding sound to the picture: Motivating the lexical composition of metaphorical idioms in English, Dutch and Spanish. In M. S. Zanotto, L. Cameron and M. d. C. Cavalcanti (eds). *Confronting metaphor in use: An applied linguistic approach / edited by Mara Sophia Zanotto, Lynne Cameron, Marilda C. Cavalcanti*, 63–79. Amsterdam, Philadelphia: John Benjamins Publishing.

Brezina, V. and Gablasova, D. (2015) Is There a Core General Vocabulary?: Introducing the New General Service List. *Applied Linguistics* 36 (1): 1–22.

Carter, R. (2012) *Vocabulary: Applied linguistic perspectives / Ronald Carter*. Milton Park, Abingdon, Oxon, New York: by Routledge.

Coe and Robert (2002) *It's the effect size, stupid: what effect size is and why it is important*. https://www.leeds.ac.uk/educol/documents/00002182.htm. Accessed 12/5/2018.

Collins Dictionary *Chinese whispers*.
https://www.collinsdictionary.com/dictionary/english/chinese-whispers. Accessed 5/2/2019.

Coxhead, A. (2000) A New Academic Word List. *TESOL Quarterly* 34 (2): 213–238.

Croft, W. and Cruse, D. A. (2004) *Cognitive linguistics*. Cambridge, U.K., New York: Cambridge University Press.

Csábi, S. (2004) *Alternative conceptualization in English and Hungarian Idioms*. Doctoral thesis. Eötvös Loránd University, Budapest.

Cummins, J. (1980) The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Issue. *TESOL Quarterly* 14 (2): 175–187.

Cummins, J. (2008) BICS and CALP: Empirical and Theoretical Status of the Distinction. In N. H. Hornberger (ed). *Encyclopedia of Language and Education*, 487–499. Boston, MA: Springer US.

Davies, M. (2018) *The iWeb Corpus*. https://www.english-corpora.org/iweb/. Accessed 5/11/2019.

Davis, P. and Kryszewska (2012) *The Company Words Keep: Lexical Chunks in Language Teaching*. Peaslake: Delta Publishing.

Deignan, A. (2005) *Metaphor and corpus linguistics*. Amsterdam, Philadelphia: J. Benjamins Pub.

DeKeyser, R. M. (1995) *Learning Second Language Grammar Rules*. https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/article/learning-second-language-grammar-rules/15D76EF2F0D1DC4BC654E7053DBD7FA3. Accessed 5/10/2019.

Draper, S. (2011) *Effect Size*. http://www.psy.gla.ac.uk/~steve/best/effect.html. Accessed 1/14/2019.

Dörnyei, Z. (2007) *Research Methods in Applied Linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Dörnyei, Z., Durow, V. and Zahran, K. (2004) Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (ed). *Formulaic sequences: Acquisition, processing, and use / edited by Norbert Schmitt*. Amsterdam, Great Britain: John Benjamins Pub.

Ebeling, J. and Ebeling, S. O. (2013) *Patterns in contrast*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Education First (2018) *English Proficiency Index: A comprehensive ranking of countries by English skills*. https://www.ef.no/epi/. Accessed 5/11/2019.

Elgort, I. (2011) Deliberate Learning and Vocabulary Acquisition in a Second Language. *Language Learning* 61 (2): 367–413.

Ellis, R. (2008) *The study of second language acquisition*. Oxford: Oxford University Press.

Erman, B. and Warren, B. (2000) The idiom principle and the open choice principle. *Text - Interdisciplinary Journal for the Study of Discourse* 20 (1).

Evans, V. (2007) *A glossary of cognitive linguistics*. Edinburgh: Edinburgh University Press.

Evans, V. and Green, M. J. (2006) *Cognitive linguistics: An introduction / Vyvyan Evans and Melanie Green*. Edinburgh: Edinburgh University.

Falck, M. J. and Gibbs, R. W. (2013) Embodied motivations for metaphorical meanings. In L. A. Janda (ed). *Cognitive linguistics: The quantitative turn the essential reader*, 81–102. Berlin: De Gruyter Mouton.

Field, A. (2017) *Discovering statistics using IBM SPSS statistics*. Thousand Oaks CA: SAGE Publications.

Gardner, D. and Davies, M. (2014) A New Academic Vocabulary List. *Applied Linguistics* 35 (3): 305–327.

Gass, S. (2015) Experimental Research. In B. Paltridge and A. Phakiti (eds). *Research Methods in Applied Linguistics: A Practical Resource / edited by Brian Paltridge and Aek Phakiti*, 130–148. London, New York: Bloomsbury Academic.

Geeraerts, D. (2006) A rough guide to Cognitive Linguistics. In D. Geeraerts (ed). *Cognitive linguistics: Basic readings / edited by Dirk Geeraerts*, 1–67. Berlin, New York: Mouton de Gruyter.

Geeraerts, D. and Cuyckens, H. (2010) Introducing Cognitive Linguistics. In H. Cuyckens and D. Geeraerts (eds). *The Oxford handbook of cognitive linguistics*, 1–22. Oxford, New York: Oxford University Press.

Gibbs, R. W. (2007) Idioms and Formulaic Language. In D. Geeraerts and H. Cuyckens (eds). *The Oxford handbook of cognitive linguistics*, 697–725. Oxford: Oxford University Press.

Grace-Martin, K. *Ancova Assumptions: When Slopes are Unequal*. https://www.theanalysisfactor.com/ancova-assumptions-when-slopes-are-unequal/. Accessed 12/7/2018.

Gries, S. T. (2013) Phonological similarity in multi-word units. In L. A. Janda (ed). *Cognitive linguistics: The quantitative turn the essential reader*, 274. Berlin: De Gruyter Mouton.

Hellekjær, G. O. (2005) *The Acid Test: Does Upper Secondary EFL Instruction Effectively Prepare Norwegian Students for the Reading of English Textbooks at Colleges and Universities?* Doctoral. The University of Oslo.

Hulstijn, J. H. (1996) Mnemonic methods in foreign language vocabulary learning: Theoretical considerations and pedagogical implications. In J. Coady and T. Huckin (eds). *Second Language Vocabulary Acquisition*, 203–224. Cambridge: Cambridge University Press.

Hummel, K. M. (2010) Translation and short-term L2 vocabulary retention: Hindrance or help? *Language Teaching Research* 14 (1): 61–74.

Hyland, K. and Tse, P. (2007) Is There an "Academic Vocabulary"? *TESOL Quarterly* 41 (2): 235–253.

Janda, L. A. (ed) (2013) *Cognitive linguistics: The quantitative turn the essential reader*. Berlin: De Gruyter Mouton.

Jones, M. and Haywood, S. (2004) Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (ed). *Formulaic sequences: Acquisition, processing, and use / edited by Norbert Schmitt*. Amsterdam, Great Britain: John Benjamins Pub.

Krashen, S. (1989) We Acquire Vocabulary and Spelling by Reading: Additional Evidence for the Input Hypothesis. *The Modern Language Journal* 73 (4): 440–464.

Krashen, S. (2012) Direct Instruction of Academic Vocabulary: What About Real Reading? *Reading Research Quarterly* 47 (3): 233–234.

Kövecses, Z. and Benczes, R. (2010) *Metaphor: A practical introduction / Zoltán Kövecses ; exercises written with Réka Benczes … [et al.]*. New York, Oxford: Oxford University Press.

Langacker, R. W. (2011) Semantic motivation of the English auxiliary. In K.-U. Panther and G. Radden (eds). *Motivation in grammar and the lexicon*, 29–48. Amsterdam, Philadelphia: John Benjamins Pub. Co.

Laufer, B. (1992) How Much Lexis is Necessary for Reading Comprehension?: 126–132.

Laufer, B. and Goldstein, Z. (2004) Testing Vocabulary Knowledge Size, strength and Computer Adaptiveness. *Language Learning* 54 (3): 399–436.

Laufer, B. and Hulstijn, J. (2001) Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics* 22 (1): 1–26.

Lazar, G. (2003) *Meanings and Metaphors: Activities to practise figurative language*. Stuttgart, Cambridge: Klett; Cambridge Univ. Press.

Lervåg, A. and Aukrust, V. G. (2010) Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of child psychology and psychiatry, and allied disciplines* 51 (5): 612–620.

Lightbown, P. and Spada, N. M. (2013) *How languages are learned*. Oxford: Oxford University Press.

Lindstromberg, S. and Boers, F. (2008a) *Teaching chunks of language: From noticing to remembering*. [London]: Helbling Languages.

Lindstromberg, S. and Boers, F. (2008b) The Mnemonic Effect of Noticing Alliteration in Lexical Chunks. *Applied Linguistics* 29 (2): 200–222.

Lokøy, G., Lundgren, H., Langseth, J. and Hellesøy, S. (2013) *Skills: Bygg- og anleggsteknikk engelsk for yrkesfag vg1/vg2*. Oslo: Gyldendal undervisning.

Malec, W. (2010) On the Asymmetri of Verb-Noun Collocations. In J. Arabski and A. Wojtaszek (eds). *Neurolinguistic and psycholinguistic perspectives on SLA*, 126–144. Bristol: Multilingual Matters.

Martinez, R. and Schmitt, N. (2015) Vocabulary. In D. Biber and R. Reppen (eds). *The Cambridge handbook of English corpus linguistics*, 439–459. Cambridge: Cambridge Univ. Press.

McIntosh, C., Francis, B. and Poole, R. (2009) *Oxford collocations dictionary: For students of English / chief editor, Colin Mclntosh ; editors, Ben Francis, Richard Poole*. Oxford: Oxford University Press.

Merriam Webster (2018) *Merriam Webster's Dictionary*. https://www.merriam-webster.com/dictionary/vocabulary. Accessed 6/12/2018.

Moon, R. (1998) *Fixed expressions and idioms in English: A corpus-based approach / Rosamund Moon*. Oxford: Clarendon Press.

Myhill, D., Jones, S. and Watson, A. (2013) Grammar matters: How teachers' grammatical knowledge impacts on the teaching of writing. *Teaching and Teacher Education* 36: 77–91.

Nagy, W. and Townsend, D. (2012) Words as Tools: Learning Academic Vocabulary as Language Acquisition. *Reading Research Quarterly* 47 (1): 91–108.

Nation, I. S. P. (2001) *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2013) *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. and Kyongho, H. (1995) Where would general service vocabulary stop and special purposes vocabulary begin? *System* 23 (1): 35–41.

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

Noroozi, I. and Salehi, H. (2013) The Effect of the Etymological Elaboration and Rote Memorization on Learning Idioms by Iranian EFL Learners. *Journal of Language Teaching and Research* 4 (4).

Olsen, S. (1999) Errors and compensatory strategies: A study of grammar and vocabulary in texts written by Norwegian learners of English. *System* 27 (2): 191–205.

Pawlak, M. (2014) *Error Correction in the Foreign Language Classroom*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Peters, E. and Pauwels, P. (2015) Learning academic formulaic sequences. *Journal of English for Academic Purposes* 20: 28–39.

Plonsky, L. and Oswald, F. L. (2014) How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning* 64 (4): 878–912.

Radden, G. and Panther, K.-U. (2004) Introduction: reflections on motivation. In G. Radden and K.-U. Panther (eds). *Studies in linguistic motivation*, 1–46. Berlin: DE GRUYTER.

Read, J. and Nation, I. S. P. (2004) Measurement of formulaic sequences. In N. Schmitt (ed). *Formulaic sequences: Acquisition, processing, and use / edited by Norbert Schmitt*. Amsterdam, Great Britain: John Benjamins Pub.

Rindal, U. (2014) *What is English?* https://www.journals.uio.no/index.php/adno/article/view/1137. Accessed 5/11/2019.

Schmitt, N. (2015) *Researching Vocabulary: A Vocabulary Research Manual*. Houndmills, Basingstoke, Hampshire, New York, NY: Palgrave Macmillan.

Schmitt, N. and Carter, R. (2000) *Lexical Phrases in Language Learning*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.707.4953&rep=rep1&type=pdf.

Schmitt, N. and Carter, R. (2004) Formulaic Sequences in action: An Introduction. In N. Schmitt (ed). *Formulaic sequences: Acquisition, processing, and use / edited by Norbert Schmitt*. Amsterdam, Great Britain: John Benjamins Pub.

Schmitt, N., Dörnyei, Zoltán, Adolphs, Svenja and Durow, V. (2004a) Knowledge and acquisition of formulaic sequences: A longitudinal study. In N. Schmitt (ed). *Formulaic sequences: Acquisition, processing, and use / edited by Norbert Schmitt*. Amsterdam, Great Britain: John Benjamins Pub.

Schmitt, N., Grandage, S. and Adolphs, S. (2004b) Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (ed). *Formulaic sequences: Acquisition, processing, and use / edited by Norbert Schmitt*. Amsterdam, Great Britain: John Benjamins Pub.

Schmitt, N. and Underwood, G. (2004) Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (ed). *Formulaic sequences: Acquisition, processing, and use / edited by Norbert Schmitt*. Amsterdam, Great Britain: John Benjamins Pub.

Shadish, W. R., Cook, T. D. and Campbell, D. T. (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Simpson, R. and Mendis, D. (2003) A Corpus-Based Study of Idioms in Academic Speech. *TESOL Quarterly* 37 (3): 419.

Simpson-Vlach, R. and Ellis, N. C. (2010) An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics* 31 (4): 487–512.

Simpson-Vlach, R. and Ellis, N. C. (2018) *The Academic Formulas List (AFL): Common academic word sequences*. https://www.eapfoundation.com/vocab/academic/afl/. Accessed 3/24/2019.

Singleton, D. M. (1999) *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.

Skoglund, D. E. (2006) *A Comparison of Norwegian and American Pupils' English Vocabulary Usage in Upper Secondary Schools*. Master in English Didactics. University of Oslo.

Skoufaki, S. (2008) Conceptual metaphoric meaning clues in two idiom presentation methods. In F. Boers and S. Lindstromberg (eds). *Cognitive linguistic approaches to teaching vocabulary and phraseology*, 101–132. Berlin: Mouton de Gruyter.

Suzuki, Y. and DeKeyser, R. (2017) The Interface of Explicit and Implicit Knowledge in a Second Language: Insights From Individual Differences in Cognitive Aptitudes. *Language Learning* 67 (4): 747–790.

Swain, M. (2005) The Output Hypothesis: Theory and Research. In E. Hinkel (ed). *Handbook of research in second language teaching and learning*, 471–484. London: Routledge.

Taylor, J. R. (2003) Meaning and context. In G. Radden and H. Cuyckens (eds). *Motivation in language: Studies in honor of Günther Radden / edited by Hubert Cuyckens … [et al.]*, 27–48. Amsterdam: John Benjamins.

Timmis, I. (2015) *Corpus linguistics for ELT: Research and practice / Ivor Timmis*. London: Routledge.

Tjora, A. (2018) *Qualitative research as stepwise-deductive induction*. London: Routledge.

Underwood, G., Schmitt, N. and Galpin, A. (2004) The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (ed). *Formulaic*

sequences: Acquisition, processing, and use / edited by Norbert Schmitt. Amsterdam, Great Britain: John Benjamins Pub.

Utdanningsdirektoratet (2006) *Læreplan i engelsk: Kompetansemål etter Vg1 – studieforberedende utdanningsprogram og Vg2 - yrkesfaglige utdanningsprogram.* https://www.udir.no/kl06/ENG1-03/Hele/Kompetansemaal/kompetansemal-etter-vg1-%E2%80%93-studieforberedende-utdanningsprogram-og-vg2---yrkesfaglige-utdanningsprogram.

Vasiljevic, Z. (2015a) Imagery and Idiom Teaching: Effects of Learner-Generated Illustrations and Etymology. *International Journal of Arts & Sciences* 8 (25-42).

Vasiljevic, Z. (2015b) Teaching and Learning Idioms in L2: From Theory to Practice. *MEXTESOL Journal* 39 (4).

Vincent, C. (1996) *Singing to a star: The school meanings of second generation Salvadorean students.* George Mason University, Fairfax, VA.

Wells, J. C. (1982) *Accents of English: 2: The British Isles*. Cambridge: Cambridge University Press.

Wray, A. (2004) 'Here's one I prepared earlier': Formulaic language learning on television. In N. Schmitt (ed). *Formulaic sequences: Acquisition, processing, and use / edited by Norbert Schmitt*. Amsterdam, Great Britain: John Benjamins Pub.

Wray, A. (2008) *Formulaic language: Pushing the boundaries / Alison Wray*. Oxford: Oxford University Press.

Appendices

Appendix 1: Information letter to the participants

Forespørsel om deltakelse i forskningsprosjekt

Informasjonsskriv om masterprosjekt, høsten 2017

" Cognitive linguistic approaches to teaching English idiomatic expressions"

Bakgrunn og formål

Prosjektet er del av en masterstudie ved Universitetet i Bergen. Formålet med studien er å prøve ut lærings- og undervisningsstrategier som er basert på nyere lingvistisk teori, og undersøke hvor effektive disse er med tanke på innlæring av engelske idiomatiske uttrykk.

Følgende problemstillinger analyseres i studien:

-        Hvilke implikasjoner kan trekkes ut fra kognitiv lingvistisk teori, om hvordan en lærer engelske idiomatiske uttrykk?

-        Hvordan kan disse implikasjonene anvendes på undervisning av idiomatiske uttrykk for elever med norsk som morsmål og engelsk som andrespråk ved videregående skole?

Hva innebærer deltakelse i studien?

Deltakelse innebærer testing før og etter undervisningsopplegg som går over flere økter. I testene skal du som deltaker lese tekster, markere uttrykk i tekster, fylle inn manglende bokstaver i ord, oversette engelske uttrykk, og skrive engelske tekster. I undervisningsopplegget blir du som deltaker presentert for idiomatiske uttrykk bl.a. gjennom ulike kategorier og ved bruk av bilder og lyd.

Noen av deltakerne vil bli intervjuet. Valg av intervjuobjekter skjer på bakgrunn av besvarelsene i undersøkelsen.

Til sammen tar gjennomføring av testing og undervisningsopplegg anslagsvis 10 undervisningsøkter, hver på 70 minutter.

Hva skjer med informasjonen om deg?

Alle personopplysninger og eventuelle lydopptak vil bli behandlet konfidensielt. Opplysningene som samles inn om hver deltaker er navn, alder, kjønn. Hvert deltakernavn blir koblet til en kode, som gjør at ingen deltakere vil kunne gjenkjennes i publisert materiale. Datamaterialet vil bli oppbevart konfidensielt på ubestemt tid, ved at navneliste og koblingsnøkkel holdes adskilt fra øvrige data.

Prosjektet skal etter planen avsluttes 01.06.2019. Personopplysninger og eventuelle intervjuopptak anonymiseres ved prosjektslutt.

Frivillig   deltakelse

Det er frivillig å delta i studien, og deltakeren kan når som helst trekke sitt samtykke uten å oppgi noen grunn. Dersom deltakeren trekker seg, vil alle opplysninger bli anonymisert.

Dersom du har spørsmål til studien, ta kontakt med Bjørnar Meling, tlf. 99 15 54 04. Veileder for studien er Dagmar Haumann, professor ved Universitetet i Bergen, tlf. 55 58 23 46.

Studien er meldt til Personvernombudet for forskning, NSD - Norsk senter for forskningsdata AS.

Samtykke til deltakelse i studien

Samtykke innhentes muntlig og det er mulig å samtykke til kun deler av studien.

Samtykke innebærer å delta i testing og undervisningsopplegg som er del av studien, samt til at opplysninger om deg kan innhentes fra klasselærer/skoleregister og at ikke-gjenkjennbare personopplysninger kan publiseres.

Appendix 2: Instructions to test components

(Test component 1a)

ESSAY ON 'BULLYING'

Write 500 words about the topic 'bullying':

What is bullying?

Where should we draw the line between bullying and "kids being kids", as one of the individuals in the documentary called it.

What are the consequences of bullying for the individual and for society?

What should be done to stop bullying?

How is bullying today different from what it was when your parents grew up?

Use Word for Mac or PC.

(Test component 1b)

Write 500 words about ONE of the topics below.

EITHER:

'Technology and research on human genetics':

How far has technology come today? Is new technology always a good thing? Why/why not? Today some people can go to other countries to get health treatment they cannot receive in Norway. What should we do about this?

Future scenario: Will the government force people to use this technology in the future, in order to reduce the costs of public health services (force people to choose healthy genes for their children)? Should the government do this?

OR

'The police and the Afro-American community in the USA':

Describe the conflicts between the black and the white community as they are described in the documentary. What are the reasons for and the consequences of this conflict? In your opinion, how can such conflicts be avoided?

Are these problems relevant to Norway? Why/why not?

One police officer referred to others who meant that black people were predisposed to become criminals. What, in your opinion, decides if a person becomes a criminal?

(Test component 2)

Imagine that you are asked to give advice to first year students who want to improve their academic writing in English.

In text 1 and 2, underline the words/phrases that would be useful for them to learn.

(Test component 3)

In text 3 and 4 there are 20 expressions of which the last half of each of the words are missing. Try to fill in the missing letters to complete these expressions.

(Test component 4)

Translate the English expressions into Norwegian. Some expressions may have multiple meanings, but you should translate them according to the meaning reflected in the sentence examples.

If you cannot think of a proper translation, but you think you know what the expression means anyway, explain it in your own words.

Appendix 3: Handout for test component 2, 'highlighting'

TEXT 1

Speech by Prime Minister Kevin Rudd to the Parliament

13 February 2008

Mr Speaker, there comes a time in the history of nations when their peoples must become fully reconciled to their past if they are to go forward with confidence to embrace their future. Our nation, Australia, has reached such a time. That is why the parliament is today here assembled: to deal with this unfinished business of the nation, to remove a great stain from the nation's soul and, in a true spirit of reconciliation, to open a new chapter in the history of this great land, Australia. The time has come, well and truly come, for all peoples of our great country, for all citizens of our great Commonwealth, for all Australians—those who are Indigenous and those who are not—to come together to reconcile and together build a new future for our nation.

To the stolen generations*, I say the following: as Prime Minister of Australia, I am sorry. On behalf of the government of Australia, I am sorry. On behalf of the parliament of Australia, I am sorry. I offer you this apology without qualification. We apologise for the hurt, the pain and suffering that we, the parliament, have caused you by the laws that previous parliaments have enacted. We apologise for the indignity, the degradation and the humiliation these laws embodied. We offer this apology to the mothers, the fathers, the brothers, the sisters, the families and the communities whose lives were ripped apart by the actions of successive governments under successive parliaments. In making this apology, I would also like to speak personally to the members of the stolen generations and their families.

I know that, in offering this apology on behalf of the government and the parliament, there is nothing I can say today that can take away the pain you have suffered personally. Whatever words I speak today, I cannot undo that. Words alone are not that powerful; grief is a very personal thing. I ask those non-Indigenous Australians listening today who may not fully understand why what we are doing is so important to imagine for a moment that this had happened to you. I say to honourable members here present: imagine if this had happened to us. Imagine the crippling effect. Imagine how hard it would be to forgive.

We embrace with pride, admiration and awe these great and ancient cultures we are truly blessed to have among us—cultures that provide a unique, uninterrupted human thread linking our Australian continent to the most ancient prehistory of our planet. Growing from this new respect, we see our Indigenous brothers and sisters with fresh eyes, with new eyes, and we have our minds wide open as to how we might tackle, together, the great practical challenges that Indigenous Australia faces in the future. Let us turn this page together: Indigenous and non-Indigenous Australians, government and opposition, Commonwealth and state, and write this new chapter in our nation's story together.

Mr Speaker, I commend the motion to the House.

TEXT 2

Companies are composed of people, and people make mistakes. Even Apple, the world's largest company can get things very wrong.

Of course, that is what happened when Apple launched the newest version of its Maps app, and faced an avalanche of disastrous reviews and negative attention.

However, Tim Cook, Apple's CEO, chose not to sweep the negative press under the rug. Instead, he confronted it, and issued an apology. That is certainly a good move, one all company leaders can learn from.

To our customers,

At Apple, we strive to make world-class products that deliver the best experience possible to our customers. With the launch of our new Maps last week, we fell short on this commitment. We are extremely sorry for the frustration this has caused our customers and we are doing everything we can to make Maps better.

While we're improving Maps, you can try alternatives by downloading map apps from the App Store like Bing, MapQuest, and Waze, or use Google or Nokia maps by going to their websites and creating an icon on your home screen to their web app. Everything we do at Apple is aimed at making our products the best in the world. We know that you expect that from us, and we will keep working non-stop until Maps lives up to the same incredibly high standard.

Time will tell if this apology will help Apple's tarnished reputation.

Appendix 4: Handout for test component 4, 'gap-fill' (with key)

TEXT 3

Saying "I'm sorry" and making amends in the English-speaking world

We probably all know how difficult it can be to go to a friend, a relative, an employer or a customer and apologise for mak**ing**___ _**a**__ mis**take**___ or for using harsh words.

To say "I'm sorry" in English can have many different meanings, all depe**nding**___ **on**___ th**e**___ context. Saying you are sorry can be a necessary, formal, heartfelt, insincere or even ironic statement. Th**e**___ ot**her**___ si**de**___ **of**___ th**e**___ "I'm sorry" coin is of course the recipient of the apology. If the statement is only formal or insincere, the recipient might feel even more offended, frustrated or angry. O**n**___ th**e**___ ot**her**___ ha**nd**___, a genuine apology can le**ad**___ t**o**___ forgiveness and reconciliation. Sometimes, saying sorry is absolutely necessary for an individual, a group of people or even a nation t**o**___ be___ ab**le**___ t**o**___ move on.

Use your preparation period to reflect on the ways and effects of apologising, and on situations where this might be necessary, for instance to mo**ve**___ **on**___ by mak**ing**___ **up**___ for pa**st**___ an**d**___ pre**sent**___ mistakes.

You should study the texts below, find new information, and revise relevant material you have worked with during your English course. This could be in your textbook, news stories, films, literature and other material in your education programme. You may also find new, suitable material. It is a good idea to ma**ke**___ a no**te**___ of useful key words and phrases. Finally, remember to note down your sources.

TEXT 4

What is social control?

Social control ref**er**___ t**o**___ th**e**___ wa**y**___ **in**___ wh**ich**___ people's thoughts, feelings, appearance, and behaviour are regulated in society. Social control **can**___ b**e**___ achi**eved**___ through socialization, a process whereby people come to identify with a social system and its values and norms, and therefore want to maintain them. Sometimes though, it is achieved through regulations or coercion, like imprisoning those who com**mit**___ a cri**me**___ or administering drugs to make people more manageable.

Sociologists identify two bas**ic**___ me**ans**___ of enforcing social control:

1 Informal means of social control. This is ba**sed**___ **on**___ learning norms, rules, and values thro**ugh**___ a proc**ess**___ known as socialization, in which children and adults are taught acceptable behaviour. The violation of such rules can be met by a vari**ety**___ **of**___ mild reactions. Examples of informal means of control are rolling one's eyes, sighing, politely explaining why certain behaviour is not acceptable, telling somebody off, using social media to express one's opinions, naming and shaming, or exclusion from a group.

2 Formal means of social control. This is based on external sanctions through laws and regulations. Exam**ples**___ **of**___ formal means of control are fines for minor offences,

suspensions, being fired, or imprisonment. It should be kept in mind that laws and social norms vary between societies, communities, belief systems, and times. How to raise children, discrimination, racism, working conditions, and lifestyles are issues that are subjected to social control, but in diff**erent**___ wa**ys**___ in different places and times.

Appendix 5: Handout for test component 4, 'translation' (with key)

1. **point of view**

Example:

 "From a purely practical **point of view**, the house is too small."

Translate **point of view** into Norwegian / Oversett **point of view** til norsk:

__synspunkt, perspektiv, ståsted, _____

2. **on the other hand**

Example:

 "My husband likes classical music – I, **on the other hand**, like all kinds."

Translate **on the other hand** into Norwegian / Oversett **on the other hand** til norsk:

_____derimot,   på   den   annen

side_____

3. **due to the fact that**

Example:

 "This is **due to the fact that** many citizens are not yet in a position to afford a new car."

Translate **due to the fact that** into Norwegian / Oversett **due to the fact that** til norsk:

_____skuldast   det   faktumet   at,   fordi   at,   på   grunn   av   at,
_____

4. **on the basis of**

Example:

"Students were selected **on the basis of** their grades and test scores."

Translate **on the basis of** into Norwegian / Oversett **on the basis of** til norsk:

_____på   grunnlag   av,   på   bakgrunn   av,   med   utgangspunkt   i,
_____

5. **a wide range**

116

Example:

 "We offer **a wide range** of products to our clients."

Translate **a wide range** into Norwegian / Oversett **a wide range** til norsk:

_____breidt   utvalg,   mange   ulike   typar,

_____

6. **with respect to**

Example:

 "The two groups were similar **with respect to** age, sex, and diagnoses."

Translate **with respect to** into Norwegian / Oversett **with respect to** til norsk:

_____med tanke på, _____

7. **give rise to**

Example:

"His speech **gave rise** to a bitter argument."

Translate **give rise to** into Norwegian / Oversett **give rise to** til norsk:

_____gav   grobunn   for,   nørte   opp   under,

_____

8. **it should be noted**

Example:

 "**It should be noted** that as a condition of employment, you will be required to sign an agreement."

Translate **It should be noted** into Norwegian / Oversett **It should be noted** til norsk:

_____Ein bør merka seg, Ein bør notere seg, Det er viktig å legge merke til_____

9. **carry out**

Example:

"They have to **carry out** a number of administrative duties."

Translate **carry out** into Norwegian / Oversett **carry out** til norsk:

_____utføre, _____

10. **take into account**

Example:

"I hope my teacher will **take into account** the fact that I was ill just before the exams when she marks my paper."

Translate **take into account** into Norwegian / Oversett **take into account** til norsk:

_____ta   i   betraktning,   merka   seg,   legge   seg   på minnet_____

11. **bottom line**

Example:

"The **bottom line** is that the great majority of our kids are physically unfit"

Translate **bottom line** into Norwegian / Oversett **bottom line** til norsk:

_____Hovedbudskapet,

Saken_____

12. **hand in hand**

Example:

"In a film, the images and sounds go **hand in hand**."

Translate **hand in hand** into Norwegian / Oversett **hand in hand** til norsk:

_____er   to   sider   av   samme   sak,   fyller   ut   kvarandre,   må   sjåast   ilag
_____

13. **ivory tower**

Example:

"The book was written by some college professor who had spent his entire professional life in an **ivory tower**."

Translate **ivory tower** into Norwegian / Oversett **ivory tower** til norsk:

_____(elfenbenstårn),   i   si   eiga   verd,   i   si   eiga   boble,
_____

14. **come into play**

Example:

«In the summer months a different set of climatic factors **come into play**»

Translate **come into play** into Norwegian / Oversett **come into play** til norsk:

_____trer i kraft, tar til å gjelda, blir sett i kraft_____

15. on the same page

Example:

"I think we're all **on the same page**"

Translate **on the same page** into Norwegian / Oversett **on the same page** til norsk:

_____på bølgelengde, (stort sett) enige/samde_____

16. get a handle on

Example:

" I can't really **get a handle on** the situation here. What's happening?"

Translate **get a handle on** into Norwegian / Oversett **get a handle on** til norsk:

_____få (situasjonen) under kontroll, forstår (ikkje) kva som skjer/foregår_____

17. shift gears

Example:

«Cancer research could **shift** into a higher **gear** thanks to these new findings»

Translate **shift gears** into Norwegian / Oversett **shift gears** til norsk:

_____(komma til å) gjera store framsteg, få vann på mølla_____

18. split hairs

Example:

"Don't **split hairs**. You know what I'm talking about"

Translate **split hairs** into Norwegian / Oversett **split hairs** til norsk:

_____kveruler, driv ordkløveri_____

19. go off on a tangent

Example:

"Our teacher would occasionally **go off on a tangent** totally unrelated to the textbook"

Translate **go off on a tangent** into Norwegian / Oversett **go off on a tangent** til norsk:

_____spore   av,   brått/plutselig   avvike   frå temaet_____

20. **thinking on my feet**

Example:

"You have to **think on your feet** in this job."

Translate **thinking on my feet** into Norwegian / Oversett **thinking on my feet** til norsk:

_____vere   snarrådig   /rådsnar   /   snartenkt
_____

21. **when the chips are down**

Example:

"We are at our best **when the chips are down**."

Translate **when the chips are down** into Norwegian / Oversett **when the chips are down** til norsk:

_____når   det   verkeleg   gjeld   /   står   om   noko   viktig
_____

22. **pass the buck**

Example:

"When it comes to teaching kids about risk, many parents are tempted to **pass the buck** to schools and other organizations."

Translate **pass the buck** into Norwegian / Oversett **pass the buck** til norsk:

_____legge ansvaret over på _____

23. **a red herring**

Example:

"A sighting of the missing woman in London turned out to be a **red herring**"

Translate **red herring** into Norwegian / Oversett **red herring** til norsk:

_____falskt   sport,   villspor,   avleiande   manøver
_____

24. no holds barred

Example:

"Jones had a **no-holds-barred** approach to the game of football."

Translate **no holds barred** into Norwegian / Oversett **no holds barred** til norsk:

_____frilynt,   pragmatisk   (uinnskrenka,   utan   restriksjonar)   ,

_____

25. at the drop of a hat

Example:

 "More people should sort out their own minor problems and stop calling the police **at the drop of a hat**"

Translate **at the drop of a hat** into Norwegian / Oversett **at the drop of a hat** til norsk:

_____for den minste lille ting, _____

26. clean bill of health

Example:

«At the end of that intensive study, the chemical industry got an environmental **clean bill of health**»

Translate **a clean bill of health** into Norwegian / Oversett **a clean bill of health** til norsk:

_____blei erklært skuldfri_____

27. in the doldrums

Examples:

"I was bored and my career was **in the doldrums.**"

Translate **in the doldrums** into Norwegian / Oversett **in the doldrums** til norsk:

_____langt nede, på hell, _____

28. on an even keel

Example:

 "She sees it as her role to keep the family **on an even keel** through its time of hardship"

Translate **on an even keel** into Norwegian / Oversett **on an even keel** til norsk:

_____på rett kjøl, _____

29. **on the skids**

Example:

"My marriage was **on the skids**"

Translate **on the skids** into Norwegian / Oversett **on the skids** til norsk:

_____var i ferd med å rakna_____

30. **pass muster**

Example:

"He spoke French and Spanish and could just about **pass muster** in Italian"

Translate **pass muster** into Norwegian / Oversett **pass muster** til norsk:

_____klarte   seg   såvidt   på   italiensk,   kunne   såvidt   gjera   seg

forstått_____