MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„An analysis of the development of lexical proficiency in the writing of Austrian students of English"

verfasst von / submitted by

Bianca Klein, BA BEd

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2022 / Vienna 2022

Acknowledgements

First of all, I would like to express my gratitude to Univ.-Prof. Mag. Dr. Christiane Dalton-Puffer for supervising my MA thesis. Not only did she provide valuable feedback which I could learn a lot from but she also always had an open ear whenever unforeseen difficulties or questions emerged and was extremely helpful in suggesting solutions.

Moreover, I am very grateful to the students who took the time to fill in my questionnaire and provided me with their academic papers. Without them this empirical study would not have been possible. At this point I would also like to thank all my colleagues and professors that distributed my questionnaire and thus helped me recruit more participants for my study.

I would also like to express gratitude to my parents for their empathy, generosity and encouragement. Without their constant support I would not have been able to pursue my degree.

Furthermore, my deepest gratitude goes to my boyfriend Andreas who has supported me not only concerning this thesis but my entire studies, has encouraged me in stressful situations and has always believed in me.

Lastly, I would like to thank my friends and colleagues who have cheered me on during my studies. Special thanks goes to my colleague and friend Rebecca who has not only read my thesis and provided me with constructive feedback but has also always had an open ear for my questions and concerns.

Thank you!

Table of contents

List of abbreviations

AVL .... Academic Vocabulary List (Gardner & Davies 2014)

AWL .... Academic Word List (Coxhead 2000)

BNC .... British National Corpus

CLIL .... Content and Language Integrated Learning

COCA .... Corpus of Contemporary American English

EE .... Extramural English

EFL .... English as a foreign language

ELF .... English as a lingua franca

ESL .... English as a second language

GSL .... General Service List (West 1953)

L1 .... First language

L2 .... Second language

LFP .... Lexical Frequency Profile (Laufer & Nation 1995)

NAWL .... New Academic Word List (Browne, Culligan & Phillips 2013)

new-GSL .... New General Service List (Brezina & Gablasova 2013)

NGSL .... New General Service List (Browne, Culligan & Phillips 2013)

PS paper .... Proseminar paper

TTR .... Type-token ratio

UWL .... University Word List (Xue & Nation 1984)

VP .... VocabProfile (Cobb n.d.)

WP .... Written production

List of tables

List of figures

1. Introduction

Vocabulary is an area of linguistics that everybody is confronted with, not only unconsciously when learning their first language as a child but also consciously when acquiring a foreign language at school and being required to actively study vocabulary. One peculiarity regarding vocabulary learning that not only many students but also teachers notice is that some seem to memorize new vocabulary items easier whilst others struggle with remembering or selecting the correct words. Despite a vast amount of studies focusing on vocabulary acquisition, research on potential influential factors, particularly regarding academic vocabulary, is rare.

Therefore, the present paper aims to shed some light on the development of lexical proficiency of students of the English Department of the University of Vienna throughout their studies as well as factors that might contribute to a higher lexical proficiency. For this purpose, two written compositions that the students produced in the course of their studies (namely the end-of-term paper for the first seminar in linguistics and their BA or BEd thesis in linguistics) are analysed with regard to lexical proficiency. This analysis will focus on different aspects of lexical proficiency, namely lexical density, lexical variation and lexical sophistication. Moreover, a potential effect of the participants' age, gender and linguistic background, particularly a potential bilingual upbringing or education, their knowledge of other languages, potential stays abroad, actively studying vocabulary and the extent of English being involved in their leisure time was investigated.

Hence, the aim of the present paper is twofold as it not only focuses on which aspects of lexical proficiency improve the most or the least but also seeks to ascertain whether any correlations between the participants' lexical proficiency and their linguistic background can be detected. Thus, the research questions of this paper strive to explore (1) which aspects of lexical proficiency improved the most and which the least between the two written productions; (2) whether the development of lexical proficiency is dependent on its initial level; (3) whether correlations between lexical proficiency and the participants' linguistic background (i.e. bilingual upbringing or education, stays abroad, extent of Extramural English activities or actively studying vocabulary) can be found and (4) whether effects of the participants' age or gender on lexical proficiency can be found.

The first two chapters of this paper are concerned with presenting the most essential research for the present study as well as determine definitions that will be used henceforth. Whilst the first theoretical section (chapter 2) introduces numerous definitions of lexical proficiency, various units of counting and types of vocabulary as well as different aspects of lexical proficiency, the second theoretical section (chapter 3) focuses on factors that might have an effect on lexical proficiency, namely bilingual upbringing and education, time spent abroad, engaging in activities in one's leisure time in which English is used, actively studying vocabulary as well as gender and age. The theoretical sections are followed by a chapter introducing the methodology of the study, in particular the participants as well as the collection, treatment and analysis of the data. Subsequently, the results of the study will be presented (chapter 5) and discussed (chapter 6) with regard to previous research on the topic.

2. Lexical Proficiency

Since this study is concerned with the development of lexical proficiency of students, a more detailed discussion of what lexical competence and lexical proficiency are and how they can be measured is called for. This chapter will, therefore, illustrate different definitions of lexical competence and lexical proficiency. Moreover, word knowledge and the choice of unit of counting will be addressed. Subsequently, the types of vocabulary that can be found in academic writing will be discussed. Lastly, different aspects of lexical proficiency will be introduced followed by measures used to analyse them.

2.1 Definition of lexical proficiency and lexical competence

First of all, it is essential to define what *lexical proficiency* and *lexical competence* are, as looking at the literature on this topic unveils numerous different terms and definitions, which Leńko-Szymańska (2020: 8) has described fittingly as follows: "SLA researchers refer to the lexicon, vocabulary knowledge, lexical competence and lexical proficiency either employing these labels interchangeably or applying them to different, but poorly defined concepts." However, despite this apparent lack of clear and consistent definitions of *lexical competence*, *lexical proficiency*, *vocabulary competence* and *vocabulary proficiency*, a working definition for this study has to be determined. Therefore, the most important aspects regarding lexical competence and proficiency will be outlined in this subchapter.

When discussing lexical competence, Chomsky's distinction between *competence* and *performance* is pivotal and is referred back to by many researchers. According to Chomsky, competence refers to "the speaker-hearer's knowledge of his language" whilst performance is considered "the actual use of language in concrete situations" (Chomsky 1965: 2). However, since the introduction of these terms numerous decades ago, various researchers have used the term competence to refer to different concepts than the one proposed by Chomsky (Taylor 1988: 166). Not only has the notion of competence been extended to include other aspects but it even has been augmented to the extent that it is sometimes equated with performance, rendering a distinction between the two concepts impossible (Taylor 1988: 166). However, it is important to note that performance hardly ever reflects competence or only does so under ideal circumstances (Chomsky 1965: 2) since, as Nation and Webb state, "we produce only a small proportion of what we are capable of producing" (2011: 197). Thus, it is definitely problematic to consider performance and competence to be the same concept. However, the discourse regarding the definition of the concept performance will not be elaborated at this point, as nowadays most researchers do not distinguish between competence and performance but instead between the two terms competence and proficiency. Hence, the focus of this section will be on discussing what proficiency is and to what extent it differs from the concept competence as the two terms are often used interchangeably with little consensus on their meaning. Moreover, at the end of this section the definitions for proficiency and competence with regard to their use in the present study will be determined.

As mentioned in the introduction of this chapter, the term *proficiency* is often used to express different ideas. For instance, Hulstijn's rather general definition of *language proficiency* for both L1 (first language) and L2 (second language) is that it "is the extent to which an individual possesses the linguistic cognition necessary to function in a given communicative situation, in a given modality (listening, speaking, reading, or writing)" (Hulstijn 2011: 242). This linguistic cognition is constituted by not only the knowledge of certain aspects of a language but also the skill required to process this knowledge and use the language (Hulstijn 2011: 242). Similar definitions of proficiency focusing on the ability to communicate were also proposed by other researchers. For instance, Thomas states that proficiency is "a person's overall competence and ability to perform in L2" (Thomas 1994: 330, quoted in Leńko-Szymańska 2020: 11). Thus, Thomas not only considers the

3

ability to use a language to constitute a part of proficiency but also the underlying competence required to communicate in the language. This definition hence bears similarities to the previously mentioned one by Hulstijn as both the more theoretical concepts of knowledge or competence and the skill to use the language is comprised in the notion of proficiency.

Whilst both Hulstijn and Thomas consider proficiency and competence to be interconnected, Davies goes even further to state that knowledge and proficiency are in fact "triggered and determined by [...] competence" (Davies 1989: 169). Moreover, he adds that knowledge and proficiency are often considered parts of a binary system, in a similar way that competence and performance are (1989: 165-166). However, he voices criticism regarding these binary terms, since "distinguishing [...] knowledge *what* from knowledge *how*" (Davies 1989: 166 [original emphasis]) can be problematic. Instead they should be perceived as being connected to each other, as the definitions mentioned above also clearly show a relation between the two concepts.

A somewhat different model has been suggested by Bulté et al. (2008), which consists of three levels. The first one is the "theoretical level of cognitive constructs" (Bulté et al. 2008: 279) which consists of lexical competence including size, width and depth of knowledge as well as procedural competence. Whilst the former three are considered "lexical 'knowledge'" (Bulté et al. 2008: 278), the latter "is more a matter of skill and control over knowledge and refers to how strongly linguistic information is stored in lexical memory, which in turn determines how well learners can access, retrieve and encode/decode relevant lexical information in real time" (Bulté et al. 2008: 279). The second level of Bulté et al.'s model is the "observational level of behavioural constructs" (Bulté et al. 2008: 279) concerned with lexical proficiency, which can be observed in oral or written production and consists of lexical diversity, sophistication, complexity, productivity and fluency. The last level is the "operational level of statistical constructs" (Bulté et al. 2008: 279) focusing on the statistical measures and measurements that can be used, such as type-token ratio (TTR), frequency-based lexical measures or the number of word types (Bulté et al. 2008: 278-279).

This model also bears similarities to the definitions proposed by Hulstijn and Thomas. Firstly, Bulté et al.'s concept of *lexical knowledge* can be considered similar to Hulstijn's concept of *linguistic cognition* concerned with knowledge of a language as well as Thomas' notion of

competence. Secondly, the *procedural competence* introduced by Bulté et al. is somewhat similar to the *skill* or *ability* mentioned by both Hulstijn and Thomas which is required to use the language. However, it also ought to be mentioned that the procedural competence is ascribed to the "theoretical level" (Bulté et al. 2008: 278) and there is an additional "observational level" focusing on lexical proficiency which manifests and can be observed in L2 production. Hence, despite the previously mentioned definitions of proficiency bearing some resemblance to Bulté et al.'s model, the latter is considerably more extensive. Moreover, a highly interesting aspect of this model is the inclusion of the *operational level* including statistical measures. This aspect is not only intriguing because it is hardly ever being mentioned in other definitions but it is also essential for the present study. Due to the three levels of the model which differentiate between the concepts of proficiency and competence remarkably clearly as well as the inclusion of the highly useful operational level, Bulté et al.'s model and their definitions of proficiency and competence were selected for the present study: whilst lexical competence is considered a "theoretical construct[...] which [is] not open to direct observation or measurement" (Bulté et al. 2008: 279), lexical proficiency is the "behavioural correlative of lexical competence in actual L2 use" (Bulté et al. 2008: 279).

As this definition indicates, it is lexical proficiency that is being measured when analysing the lexical aspects of a written or oral production. Although lexical competence cannot be measured or observed directly, it can - at least to some extent - be "inferred from [...] the behavioural manifestations of the underlying cognitive constructs in actual L2 performance" (Bulté et al. 2008: 279), i.e. the lexical proficiency demonstrated in L2 production. The notion of proficiency being what is measured is not only supported by Bulté et al. but also by other researchers, such as Leńko-Szymańska who states that "[t]he use of vocabulary (i.e. lexical performance) is the observable manifestation of lexical proficiency" (Leńko-Szymańska 2020: 39). Therefore, the term *lexical proficiency* will be used in the present study when referring to measuring and analysing learners' written productions.

To conclude, it can be said that there are numerous different definitions of the concepts competence, proficiency and performance with the introduction and comparison above showing merely a fraction of the discourse on the topic. However, the aim of this section was not to provide a full account of the variety of different definitions but to compare the

most relevant ones in order to determine which definition ought to be selected for the present study. Despite lexical proficiency and lexical competence by no means being identical, the two concepts are closely connected to each other. Although it is lexical proficiency that can be directly observed and hence also measured, it might be possible to draw inferences regarding the more theoretical construct of lexical competence based on a person's proficiency as demonstrated in L2 production.

2.2 Word knowledge

In the context of vocabulary knowledge and acquisition, word knowledge is frequently mentioned. Hence, in the following, the different definitions of word knowledge and aspects involved in knowing a word will be illustrated.

Richards (1976), for instance, has provided a rather detailed account of assumptions that prevail regarding word knowledge and has demonstrated that knowing a word can mean anything from knowing the semantic value or the syntactic behaviour to "knowing the network of associations between that word and other words in the language" (Richards 1976: 83). Similarly, Nation (2013) proposes that there are nine aspects to be considered when striving to know a word, namely "spoken form", "written form", "connecting form and meaning", "concept and referents", "associations", "grammatical functions", "collocations", "constraints on use" as well as "item knowledge and system knowledge" (Nation 2013: 65-85).

This enumeration already indicates that there are numerous different aspects involved in word knowledge. When researching literature on the topic of word knowledge, one could easily be misled into thinking that a person needed to know all aspects of a word in order to be able to state that they really know a word. However, numerous researchers have called attention to the fact that there are many different degrees of word knowledge (Laufer & Paribakht 1998: 366-367) instead of an "'all-or-nothing' phenomenon, as is commonly assumed by laymen" (Leńko-Szymańska 2020: 19). Hence, one should not think of a person knowing or not knowing a word in terms of an all or nothing state, as "knowledge of a word exists on various levels" (Taylor 1990: 1).

Consequently, more often than not it is unnecessary to know all aspects of a word in their entirety (Leńko-Szymańska 2020: 19), such as the ones enumerated by Richards (see

Richards 1976: 83 for more details) or Nation (2013), in order to be able to use it effectively. Thus, when referring to a learner knowing a word in the present study, the ability to use the word in a meaningful way is meant rather than knowing every single aspect of a word.

2.3 Unit of counting

Another aspect of importance when discussing the meaning of *word knowledge* is the unit of counting that is used when measuring and assessing lexical proficiency in a written production. One possibility would be counting the words of a text. Nation (2013) makes a distinction between counting tokens and counting types. When employing the former strategy, all tokens, i.e. all orthographical words, are counted no matter how often they occur. In contrast, when using the latter strategy words that occur multiple times are only counted once (Nation 2013: 9-10). However, this method can be rather difficult as it is not always entirely clear what counts as a separate word (Meara & Miralpeix 2016: 15), particularly with regard to testing a person's vocabulary size (Read 2000: 16). Hence, there are different viewpoints regarding several issues, for instance, whether merely content words should be regarded, whether inflected forms or derivatives should be considered separate words or how homographs should be treated (Read 2000: 18-20).

Thus, some researchers prefer to use alternative units of counting. One of these is counting word families, which can be defined as "a headword, its inflected forms and its closely related derived forms" (Nation 2013: 11). An example provided by Read is "*leak*, with the inflected forms *leaks*, *leaking* and *leaked* as well as these derivatives: *leaky*, *leakiness*, *leakage* and *leaker*" (Read 2000: 19). Nation and Webb have added that when measuring vocabulary knowledge, the unit of counting has to be selected according to what exactly, e.g. receptive or productive knowledge, is analysed. For instance, selecting the word family as the unit of counting is particularly advisable when investigating receptive knowledge (Nation & Webb 2011: 136) as when some of the members of the word family are already known, the learning effort required to learn the other family members as well is relatively low (Nation & Webb 2011: 136).

In contrast to the relatively low learning effort of learning more members of a word family in order to know them receptively, it is much more time-consuming to learn how to use them in oral or written productions (Nation & Webb 2011: 136). Hence, for measuring productive

knowledge, for instance in the form of written compositions, a different unit of counting is more advisable, namely lemmas (Nation & Webb 2011: 201), which can be considered the "base and inflected form of a word", such as "*society, societies, society's* and *societies*'" (Read 2000: 18 [original emphasis]). However, the effort required to learn lemmas can vary: if a learner knows how to, for instance, form the plural of a word, learning the word *books* when already knowing the word *book* will not require a tremendous effort. Although the same is also the case with forming the past tense of verbs, irregular forms feature a higher learning effort than regular forms (Nation 2013: 10). This has led to the question of whether irregular forms should be considered part of the lemma of their base word or as separate lemmas (Nation 2013: 10). As this is only one of the issues that have to resolved, it can be seen that various different considerations have to be taken into account when choosing lemmas as the unit of counting.

To sum up, it can be said that the choice of unit of counting (e.g. tokens, types, word families or lemmas) is highly dependent on two factors in particular: researchers do not only have to decide which unit is most appropriate for the purpose of the study but also for the level of knowledge (Nation 2013: 11) that can be expected of the study's participants. As in the present study productive knowledge is analysed, lemmas might be considered the most appropriate unit of counting. However, word lists will be used in order to investigate the learners' Lexical Frequency Profiles and the lists most suitable for the purpose of this study have used modified lexemes. This unit of counting differs from the above-mentioned ones in that modified lexemes "count the headword in all its various parts of speech and include all inflected forms" (Browne 2014: 6). The example that Browne provides is the modified lexeme *list*, which consists of *lists*, *listed*, *listing*, and *listings* (Browne 2014: 6). As this unit of counting was selected for the present study, it will be discussed in more detail in chapter 2.6 on word lists.

2.4 Types of vocabulary in academic writing

Similar to many other concepts that have already been discussed, there is no one undisputable definition of *academic vocabulary* consistently used in literature and consequently, various different definitions can be found (Coxhead 2020: 99). However, before investigating what academic vocabulary is, it appears necessary to introduce general concepts of how vocabulary can be categorised since not only academic vocabulary can be

found in academic writing. An often-cited division of vocabulary into three areas which has been created by Nation (2001) is to categorise words into *high-frequency words*, *low-frequency words* and *specialised vocabulary* (which contains *academic* and *technical vocabulary*) (Nation 2001: 13-21). In a later edition of his book, which was published in 2013, Nation added the category of *mid-frequency words* (Nation 2013: 25). In the following, these categories will be discussed in more detail.

2.4.1 High-, mid- and low-frequency words

High-frequency words can be considered words that "cover a very large proportion of the running words in spoken and written texts and occur in all kinds of uses of the language" (Nation 2013: 22). Although many researchers use Nation's term *high-frequency words*, sometimes different terms can be found in literature. One of them is *core vocabulary* as used by Paquot, who also mentions the terms *basic or nuclear vocabulary* (Paquot 2010: 10). Despite the difference in terminology when compared to Nation, Paquot defines these words as being "of high frequency in most uses of the language" (Paquot 2010: 10) and hence, his definition is relatively similar to that of the term *high-frequency words*. Another term that can be found - albeit rather rarely - is *general service words* which can primarily be ascribed to West compiling the *General Service List* (GSL) in 1953, consisting of 2,000 word families that are "considered suitable as the basis of vocabulary for learning English as a foreign language" (West 1953: vii) (for more details see chapter 2.6 on word lists). Since the word families in the GSL were, amongst other criteria, selected due to their high frequency (West 1953: vii), the term *general service words* can be considered equivalent to that of *high-frequency words* (Read 2000: 227). Regarding the number of high-frequency words, some researchers, such as Nation (2013: 23) as well as Chung and Nation (2003: 104), have orientated themselves to the original GSL by West (1953) and therefore, traditionally, 2,000 word families are assumed to be in this group. However, sometimes different numbers are used (Nation 2013: 23), for instance by Nation and Waring who state that learners ought to know "the 3,000 or so high frequency words of the language" (Nation & Waring 1997: 11). Similarly, Schmitt and Schmitt suggest that considering 3,000 words as high-frequent words would be "more pedagogically useful" as they "represent an important milestone in language development" (Schmitt & Schmitt 2014: 492), and other researchers have followed their lead (Nation & Anthony 2013: 7).

When adding proper nouns, transparent compounds and marginal words (i.e. words like "*aah, gosh, sshh* which are common in spoken language but are not dictionary entry words" (Nation 2013: 20)) to these 3,000 word families, 95% of most texts are covered (Nation 2013: 18). Hence, this group of word families is essential for comprehending a language and using it proficiently.

Despite many researchers merely mentioning high- and low-frequency words, others include a frequency band in between these two, namely *mid-frequency words*. This group can be considered of consisting of "7,000 word families from the third to the ninth 1,000" (Nation 2013: 25-26) or the "gap between the 3,000 and 9,000 levels" (Schmitt & Schmitt 2014: 494) depending on how many words are considered to belong to the high-frequency group. Schmitt and Schmitt argue that despite mid-frequency words usually not being taught, they should definitely play a bigger role in teaching as they enable students "to engage with English for authentic purposes" (Schmitt & Schmitt 2014: 495).

Although low-frequency words constitute the largest number of words in a language, words belonging to this group are the scarcest in a text (Nation 2013: 18). Nation as well as Schmitt and Schmitt (2014: 494) consider low-frequency words those that are "beyond the most frequent 9,000 words of English" (Nation 2013: 28) as this number of words covers 98% of the majority of texts (Nation & Anthony 2013: 7). This group of words consists of proper names, technical vocabulary but also simply words that are very rarely used in written texts, which can be due to several reasons, such as that they "represent a rarely expressed idea, they may be similar in meaning to a much more frequent word or phrase, they may be marked as being old-fashioned, very formal, belonging to a particular dialect, or vulgar, or they may be foreign words" (Nation 2013: 29). Due to the scarcity of this group of words, a large number of researchers seems to agree on the fact that teaching low-frequency words should not be prioritized (Nation & Waring 1997: 10). Instead, teachers should focus on introducing vocabulary strategies (Nation & Waring 1997: 10; Nation 2013: 10) and should only teach low-frequency words when they are essential for a specific text or when they are part of a technical vocabulary the students ought to know (Nation 2013: 29).

2.4.2 Specialised vocabulary

In this sub-chapter *specialised vocabulary*, which Nation divides into *academic* and *technical vocabulary*, will be discussed (Nation 2001: 17). Despite the categorisation by Nation being a popular and often-cited one, not all researchers subsume academic and technical words under the term specialised vocabulary but use different terms instead (e.g. Baumann & Graves 2010; Paquot 2010). However, when investigating the literature on academic vocabulary, it becomes evident that researchers usually define academic vocabulary as either "general academic vocabulary", which consists of "words that appear in texts across several disciplines or academic domains" or as "domain-specific academic vocabulary", i.e. vocabulary used in certain domains, such as chemistry or medicine (Baumann & Graves 2010: 6). For both of these concepts, numerous different terms can be found in the literature. Whilst the more general academic vocabulary is sometimes called *sub-technical vocabulary* (Cowan 1974: 391, quoted in Martin 1976: 91) or *discourse-organizing vocabulary* (Paquot 2010: 9), the latter is by some also referred to as *technical vocabulary* (Martin 1976: 91). Similar to Baumann and Graves, Nation considers academic words as being "common in different kinds of academic texts" (Nation 2013: 19), whereas technical words are "closely related to the topic and subject area of the text" (Nation 2013: 19) and hence "highly subject-specific" (Paquot 2010: 13). Gardner and Davies subdivide academic vocabulary into *academic core words* (which can be found "in the vast majority of the various academic disciplines") and *academic technical words* (which are primarily used in very specialised contexts and disciplines) (Gardner & Davies 2014: 312). Whilst the former tend to constitute approximately 9% of a text, the latter can even make up 20% to 30% of a text (Nation 2013: 19-20).

Furthermore, it also ought to be mentioned that, in addition to academic and technical vocabulary, general high-frequency words can be found in all registers and texts, including academic ones (Gardner & Davies 2014: 312). Similarly, technical vocabulary is distributed between belonging to high frequency words, the *Academic Word List* (which will be discussed in more detail in chapter 2.6) or to a group of words that can merely be found in a very specialised field (Nation & Meara 2010: 37-38) and can hence also belong to low-frequency words (Nation 2013: 31; Chung & Nation 2003: 108). The words *cabotage* and *chest*, for instance, can both be considered technical vocabulary, but whilst the former is

seldom found outside its very specialised field, the latter is frequently used in everyday English (Nation 2013: 31). Conversely, words that are considered to belong to the general or subtechnical group might also have a technical meaning, which may lead to them being included in technical lists as well (Read 2000: 230) just as a word can be considered "academic vocabulary in some contexts and general vocabulary in others" (Coxhead 2020: 99). An example for a word belonging to multiple categories or groups of vocabulary is the word *area*, which, as Coxhead points out, can be found in the category of general vocabulary but also in the academic field of mathematics (Coxhead 2020: 99). These sometimes fuzzy boundaries between core words, academic and technical vocabulary can lead to difficulties in the compilation of exclusively academic or exclusively technical word lists (Paquot 2010: 13-15).

To conclude, it can be seen from the discussion above that not all researchers categorise academic and technical vocabulary as belonging to the group of specialised vocabulary in the way that Nation does. Nevertheless, it could also be shown that most researchers divide academic vocabulary into a more general academic vocabulary, which can be found in all disciplines, and into a more specific academic vocabulary, which is merely used in certain domains and could therefore be considered technical vocabulary. Hence, despite different terms being used, the categorisations often do not vary significantly. For this paper, Nation's categorisation into academic and technical vocabulary will be used consistently as the terms coined by him are one of the most widely used and referred to ones.

2.5 Measuring lexical proficiency

As Leńko-Szymańska states, researchers only started investigating vocabulary in the 1970s since this aspect of language was often considered to belong to grammar (see Leńko-Szymańska 2020: 6-7 or Meara 1996: 35 for a more detailed discussion) and even in foreign language teaching, other aspects of language were considered more important than vocabulary (Taylor 1990: 1). However, nowadays there are numerous ways of analysing and measuring both receptive and productive lexical proficiency (Milton 2009: 20). The latter, which is of more importance for the present study, can, for instance, be analysed with the help of various different measures of lexical richness (Laufer & Nation 1995: 308-310), Lexical Frequency Profiles, the proportions of different word classes used in a text or word association tasks requiring students to write down words they associate with a stimulus

word (Milton 2009: 140-141). For this study, the two most important concepts are lexical richness measures and Lexical Frequency Profiles, which will be presented and discussed with regard to their suitability in more detail in the following subchapters.

2.5.1 Lexical richness

Lexical richness can be considered an umbrella term for various characteristics displaying "the quality of lexical knowledge that is demonstrated in a text" (Nation & Webb 2011: 246). All of the measures of lexical richness can be said to indicate "the degree to which a writer is using a varied and large vocabulary" (Laufer & Nation 1995: 307) and hence displays "effective vocabulary use" (Read 2000: 200). Amongst these are lexical density, lexical variation, lexical sophistication and lexical originality, which will be defined and discussed concerning their usefulness and reliability in the following sections.

Lexical density

Lexical density measures the percentage of lexical words (which carry the majority of the meaning) in relation to the amount of grammatical words (Read 2000: 200). In written language, the number of content words is usually higher than that of grammatical words, and vice versa in spoken language (Read 2000: 200). However, Laufer and Nation also stress the issue that lexical density might not be an indicator of a person's lexical proficiency but that it rather reveals information about syntactic and cohesive aspects of a written production (Laufer & Nation 1995: 309). Hence, lexical density can, for instance, give information about the text being spoken or written and also about the genre of the text (Read 2000: 209) rather than the lexical proficiency of a person. Nevertheless, combined with other measures, lexical density can be a beneficial tool as the proportion of lexical words can indeed provide valuable insight into the lexical proficiency demonstrated in a written production.

Lexical variation

Lexical variation, which is also referred to as type-token ratio (TTR) (Read 2000: 200), shows whether words are frequently used repetitively or whether the author has a larger amount of words at their disposal and is therefore able to avoid frequent repetitions (Read 2000: 200). According to Nation and Webb, lexical variation, along with lexical sophistication, belongs to the most discussed aspects of lexical richness (Nation & Webb

2011: 249), which is particularly due to it being a highly controversial issue as will be discussed in this section.

The type-token ratio is drastically affected by the length of a written production as with the increasing length of a text, the number of words that have already been used increases automatically, leading to a lower type-token ratio (Read 2000: 201; Nation & Webb 2011: 250). Therefore, it can also be said that the type-token ratio only yields meaningful results when analysing texts of (approximately) the same length (Nation & Webb 2011: 199). In the present study, the texts under analysis will thus be adapted with regard to their length. Furthermore, it also has to be mentioned that there are numerous different measures of lexical variation that are "mathematical variations of the type-token ratio" (Nation & Webb 2011: 199). An enumeration and description of these measures can be found in Kojima & Yamashita (2014: 24) and Nation & Webb (2011: 250-251). Interestingly, the latter state that the various measures do not yet provide valid results for all text lengths, as, for instance, the D index can be considered to yield reliable results for texts of 100-400 tokens but none of the indices described calculate reliable comparisons for both shorter and longer texts (Nation & Webb 2011: 251). The various different ways to measure the type-token ratio as well as their advantages and disadvantages, particularly with regard to the present study, will be discussed in more detail in chapter 4.3.

Another problematic aspect about the measure of the type-token ratio is that there is no distinction between low-frequency and high-frequency words, meaning that two students might have very similar results even though one uses very sophisticated vocabulary and the other uses more high-frequency vocabulary (Laufer & Nation 1995: 310). This shows that the type-token ratio, despite illustrating the variation in a written composition, cannot be the sole indicator of a person's lexical proficiency. Due to these problematic aspects of lexical richness measures, particularly concerning lexical variation, a vast amount of research has been conducted with the aim of finding and improving alternative measures which are more reliable (see, for instance, Nation & Webb 2011: 249-250 for an enumeration of these studies).

Lexical sophistication

Lexical sophistication is concerned with the question of whether a high number of "low-frequency words that are appropriate to the topic and style of the writing" (Read 2000: 200) are used as opposed to everyday vocabulary (Read 2000: 200). However, this raises the question of which words are considered "advanced" (Laufer & Nation 1995: 309). As Laufer and Nation point out, this highly depends on the level of proficiency of the person who has produced the text as, for instance, a word that is considered advanced for a student with a rather low proficiency might not be an advanced word for a university student (Laufer & Nation 1995: 309-310). Hence, when comparing written productions from numerous learners, it is essential to establish a standardised definition of which words are considered low-frequency or advanced. Furthermore, it ought to be ensured that the learners' proficiency does not differ tremendously, i.e. that a first year learner is not compared to a very proficient learner as this could distort the results considerably.

Due to these limitations, lexical sophistication might not be the most suitable measure when comparing groups from very different educational systems (Laufer & Nation 1995: 310). However, when considering the above-mentioned aspects, namely determining a standardised definition and ensuring that the participants can be allocated to a similar educational background (as is the case in the present study), lexical sophistication is a useful measure to shed more light on the learners' lexical proficiency, particularly when combined with other measures of lexical richness.

Lexical originality

Lexical originality refers to the percentage of words in a written composition that is used by only one person of the group of participants (Laufer & Nation 1995: 309). However, this also means that lexical originality is extremely dependent on the group of participants whose texts are analysed. Since lexical originality is not measured with regard to one specific text but with regard to the composition in relation to the texts of the entire group, the results of the measure can be considered somewhat unreliable (Laufer & Nation 1995: 309). Therefore, lexical originality is only mentioned for the sake of completeness but will not be included and measured in the empirical part of this study.

Errors

Lastly, another measure of lexical richness that can be found in literature is counting and categorizing lexical errors (Nation & Webb 2011: 245). Just as a text that features less repetition and more low-frequency words can be considered more lexically rich, "[t]exts that contain fewer misspelled words are richer than those that have a greater number of these errors" (Nation & Webb 2011: 246). For this reason, errors occurring in a text are often counted and categorised (Engber 1995: 146). That being said, this measure will not be attended to in more detail as applying it would extend the scope of this study due to the lack of computerised programmes that could be used for counting errors.

As this section has shown, most measures of lexical richness have their weaknesses, although it ought to be noted that these primarily take effect when selecting merely one of the measures and striving to infer generalisable results from them. When combining them, however, they can provide meaningful insights into the lexical proficiency of a person. Hence, in the present study several measures will be selected in order to ensure the findings will not be distorted due to potential weaknesses of individual measures.

2.5.2 Lexical Frequency Profiles

As a result of the above-mentioned limitations of some measures of lexical richness, another way of assessing certain aspects of lexical richness was developed, namely frequency measures that distribute the words of a text into different bands (Laufer & Nation 1995: 311). The two most frequently used and quoted ones are the Lexical Frequency Profile (LFP) by Laufer and Nation (1995) as well as the P_Lex by Meara and Bell (2001) (see for instance Crossley, Cobb & McNamara 2013: 966-967). In this section the two measures will be introduced as well as compared concerning their applicability for learners with different proficiency levels.

Laufer and Nation suggest two different LFP measures depending on the level of proficiency of the learners (1995: 311). For advanced learners, multiple levels, such as the first 1,000 most frequent words, the second 1,000 most frequent words, the *University Word List* (*UWL*, which will be discussed in more detail in chapter 2.6 on word lists) and less frequent words are used. Subsequently, the percentage of word families that belong to each of the aforementioned frequency bands or levels are calculated (Laufer & Nation 1995: 311-312).

As the LFP is sensible to varying lengths of the written productions (Read 2000: 204), a standard length is crucial for the analysis to be informative and insightful. According to Laufer and Nation, the LFP has multiple advantages over other measures of lexical richness. Amongst these are an increased objectivity due to the LFP's independence of the learner's environment, its independence of syntax and text cohesiveness and the lack of subjective decisions regarding definitions (Laufer & Nation 1995: 313). Moreover, as already mentioned in chapter 2.5.1, texts might feature a similar type-token ratio but show differences in the distribution of the vocabulary into the various frequency levels. As this is, of course, a reflection of the lexical proficiency of the learners, a distinction between these texts has to be made. This is essential as people with a higher lexical proficiency and vocabulary size use an increased number of low-frequency words (Nation & Webb 2011: 251). Hence, it could be the case that even though the type-token ratio of the texts is similar, one person uses a high number of high-frequency words whilst the other person (with a higher lexical proficiency and larger vocabulary size at their disposal) uses a higher number of low-frequency words. This example shows very clearly that although the type-token ratio is a valid indicator of a person's lexical proficiency, it should never be the sole measure of lexical richness employed.

Despite the advantages of the LFP as a lexical richness measure, it also has some limitations, which have to be taken into consideration, such as being sensitive to the length of a text as the results of written productions that are shorter than 200 words are relatively unstable (Laufer & Nation 1995: 314). For this reason, Meara and Bell developed an alternative, namely the P_Lex, which can also be used for shorter texts without a decrease in reliability (Meara & Bell 2001: 9). In order to accomplish this, all words of a written production that can be found in Nation's first 1,000 word list (1984) as well as "proper nouns, numbers and geographical derivatives" (Meara & Bell 2001: 9) are considered "easy words", whilst all other words are categorised as "hard words" (Meara & Bell 2001: 9). Subsequently, the number of infrequent words is counted. It has been shown that the P_Lex yields stable results between two texts by the same learner if they are longer than approximately 120 words (Meara & Bell 2001: 12-13). However, it could be considered problematic that all words except Nation's first 1,000 word list (1984) are categorised as 'hard words' since the difficulty of a word cannot necessarily be determined by its frequency alone (Meara & Bell 2001: 15). Hence, whilst the P_Lex might be a more appropriate choice for low-level learners, the LFP might be the better choice for advanced learners.

Although the LFP and the P_Lex tend to be mentioned more frequently in the literature, numerous other frequency-based measures have also been developed, such as Beyond 2000 (Laufer 1995), Guiraud Advanced or the Advanced TTR (Daller, van Hout & Treffers-Daller 2003). However, they are similar to the P_Lex in that they merely distinguish between "basic versus advanced lexical items" (Kojima & Yamashita 2014: 27). Therefore, for this study the LFP will be used due to its more diverse and precise distinction of words into different frequency bands.

Lastly, despite all of the mentioned advantages of frequency-based measures of lexical richness in general, some problematic aspects should also be taken into consideration. As there are no standard frequency lists and the various measures that are used tend to differ from each other (Nation & Webb 2011: 201), the lists that are used to measure lexical richness can have a substantial effect on the result. Furthermore, Nation and Webb note that it should not be assumed that "words will be learned according to their frequency" (2011: 201), as there are numerous exceptions to this rather general rule. Hence, in the present study, the LFP will not be used as the sole indicator of the participants' lexical proficiency but combined with other measures of lexical richness, which is relatively common in the research literature (see, for instance, Agustín-Llach 2017).

2.6 Word lists

Word lists are essential in many different contexts, such as for vocabulary testing (Read 2000: 224), developing course material (Nation 2016: 3) or deciding on which vocabulary items students ought to learn in a course (Coxhead 2000: 214; Nation & Webb 2011: 132). However, another way of using them - which is also the most important one for this study - is for analysing the vocabulary used in a written production and, for instance, the lexical richness of a text (Nation & Webb 2011: 133). Due to these manifold areas of application, researchers have compiled various word lists for different purposes. For the present study, word lists focusing on academic vocabulary and high-frequency words are of most importance. Hence, the most popular and most frequently used ones will be discussed in the following in chronological order with regard to their advantages and disadvantages in order to determine which lists ought to be used for this study.

One of the lists focusing on high-frequency words is the *General Service List (GSL)* which was compiled by West in 1953. Although there, of course, existed word lists before this one (for an overview see, for instance, Nation 2016: 9-10) and numerous more recent lists have been compiled, the GSL can be considered one of the most influential word lists (Brezina & Gablasova 2015: 1) and "still remains the one that others try to improve on" (Nation 2016: 10). The list is based on a corpus consisting of 5 million words and West selected the 2,000 most useful word families to be in the GSL. In order to be chosen, the word families had to fulfil several criteria, all of which are based on the notion that the list was compiled for ESL (English as a second language) and EFL (English as a foreign language) learners. Not only the frequency of the word families was taken into account but also their cost (i.e. how easy or difficult it is to learn them), their necessity, their coverage (when multiple items express the same concept, "the item nearer to the root sense" (West 1953: ix) is chosen) and their stylistic level. Interestingly, West did not consider emotional and intensive words (which are "the equivalent of underlining, or an exclamation mark" (West 1953: x)) essential for ESL or EFL learners as their goal usually is to "express ideas rather than emotion" (West 1953: x). Although emotional and intensive words might be considered essential for authentic and native-like language use, it could be argued that they are not pivotal when learning a foreign language, particularly for learners with a lower proficiency.

Even though the GSL already exists for more than half a century and compiling word lists was tremendously more difficult decades ago due to the lack of computerised programmes than it is nowadays (Read 2000: 224), some researchers argue that no other list has been able to replace the GSL yet (Coxhead 2000: 214; Nation & Webb 2011: 134). However, due to many of the vocabulary items in the GSL being outdated (Brezina & Gablasova 2015: 2) and more current vocabulary missing in the list as it is based on a corpus from the 1900s (Gardner & Davies 2014: 308; Nation & Webb 2011: 152), numerous researchers have called for a more recent list. This criticism has led to various different lists being compiled in the last few years, such as the new-GSL and the NGSL, which will be discussed at a later point in this subchapter.

Another word list that has frequently been used is the *University Word List (UWL)* which was compiled by Xue and Nation in 1984 and consists of 836 word families (Coxhead 2000: 226). The UWL is a combination of four previously existing lists, namely the ones by Campion and

Elley (1971), Praninskas (1972), Lynn (1973) and Ghadessy (1979). Whilst the former two "based their lists on corpora and identified words that occurred across a range of texts", the latter two lists were created "tracking student annotations above words in textbooks" (Coxhead 2000: 214). Hence, as these four lists were used for the compilation of the UWL instead of a corpus, it has to be taken into account that not only the strengths of the lists were transferred to the UWL but also their weaknesses. Moreover, as the four lists had different criteria, it can also be considered problematic that the UWL does not have consistent criteria for selecting word families (Coxhead 2000: 214; Nation 2016: 11).

Due to these and other criticisms, Coxhead developed the *Academic Word List (AWL)* in 2000, which is one of the most frequently used lists and even "became the new standard" (Gardner & Davies 2014: 307). As Coxhead states that a new word list ought to be based on a "large, well-designed corpus of academic English" (Coxhead 2000: 214), she decided to create a new corpus consisting of written academic English, namely the Academic Corpus (Coxhead 2000: 214). It comprises 414 academic texts, which were taken from various sources, such as academic journals, edited academic journals that can be found online, university textbooks and texts from various different corpora (Coxhead 2000: 219-220). The Academic Corpus was subdivided into four subcorpora focusing on texts from the areas of arts, commerce, law and science. Each subcorpus was again separated into seven subject areas (for instance, the subcorpus of arts was divided into education, history, linguistics, philosophy, politics, psychology and sociology) (Coxhead 2000: 219). Subsequently, the programme Range was used to gather information on the frequency of the words in the Academic Corpus (Coxhead 2000: 221). In order for the words to be selected, they had to fulfil three criteria, namely specialised occurrence (i.e. they could not be in the GSL), range (they had to occur a specific number of times in both the corpus and the individual subject areas) and frequency, meaning that "[m]embers of a word family had to occur at least 100 times in the Academic Corpus" (Coxhead 2000: 221).

The AWL was compiled with the aim of creating a list useful for students of various disciplines who strive to enhance their academic vocabulary (Nation & Webb 2011: 150). Coxhead chose the word family as the unit of counting as she primarily aimed at enabling students of different disciplines to read academic texts (Nation & Webb 2011: 150). As Nation and Webb note, it would be interesting to have a comparable list for productive

academic vocabulary with lemmas as the unit of counting (Nation & Webb 2011: 150). Other researchers also criticize the choice of the unit of counting for multiple reasons, such as the enormous difference in meaning between different members of a word family (for instance, *react* and *reactionary* belong to the same word family but differ in meaning tremendously (Gardner & Davies 2014: 307)).

Concerning the scope of the AWL, 570 of the word families found in the Academic Corpus meet the criteria of not being in the GSL and the majority of the word families are spread over all 28 subject areas or at least over a high number of the subject areas (Coxhead 2000: 222). The word families occur in commerce, arts, law and science texts with varying frequencies (from 12.0% to 9.1%) and cover 10.0% of the Academic Corpus (Coxhead 2000: 222). In comparison to this, the first 2,000 words of the GSL and the word families of the AWL combined make up 86% of the Academic Corpus (Coxhead 2000: 222). The overlap between the UWL and the AWL is 51% as 435 word families can be found in both lists. Despite the UWL (with 836 word families) containing 266 word families more than the AWL, Coxhead argues that the UWL covers 0.2% less of the Academic Corpus. This would mean that it would suffice for learners to know the smaller number of word families of the AWL and still cover the same amount of academic texts (Coxhead 2000: 226). This example goes to show that combining already existing lists, as it was done in order to create the UWL, can be problematic as this might "compound the errors and weaknesses" (Nation & Webb 2011: 148) of all of the lists used to compile the new one. However, as a consequence, Coxhead's decision to include only words in the AWL that are not in the GSL could be considered problematic as well, as this inevitably means that any weaknesses of the GSL (including its age) transferred directly onto the AWL (Nation & Webb 2011: 151). Another problem arising due to the AWL being based on the GSL is that some words that have an academic meaning, such as *interest* or *account* (Gardner & Davies 2014: 309) are not included in the AWL because they are in the GSL.

For these reasons (particularly the choice of word families as the unit of counting and the AWL's connection to the GSL) some researchers have called for a new word list (Gardner & Davies 2014: 307-308). As a result of this discontent with already existing lists, the *Academic Vocabulary List (AVL)* was developed by Gardner and Davies in 2014. In order to ensure their new list would fulfil all criteria of a successful and relevant list, Gardner and Davies have

developed five criteria they deem crucial to consider when compiling a new word list. One of these aspects is that (in contrast to all of the aforementioned ones which made use of word families as the unit of counting), a new list using lemmas ought to be established (Gardner & Davies 2014: 312). Moreover, the list should be "based on a large and representative corpus of academic English" (Gardner & Davies 2014: 312). They also hold the opinion that this corpus has to be capable of distinguishing between the aforementioned *academic core words*, *general high-frequency words* and *academic technical words* and moreover consist of words considered contemporary English. Lastly, in order to ensure that the new list is valid and reliable it should be tested against various different corpora (Gardner & Davies 2014: 312).

As the AVL was derived from the academic subcorpus (containing 120 million words) of the Corpus of Contemporary American English (COCA), the corpus used for the AVL is more recent and 35 times larger than the one used for the compilation of the AWL (Gardner & Davies 2014: 312-313). All of the words in the corpus used for the AVL are written academic words, which were found in academic journals and "academically oriented magazines" (Gardner & Davies 2014: 313). Gardner and Davies also established various selection criteria, namely that the frequency of a word has to be considerably higher in the academic corpus when compared to the non-academic part of the COCA. Moreover, the words must occur a specific amount of times in seven out of the nine disciplines and have to be spread evenly across the corpus. Lastly, in order to ensure that discipline-specific and technical vocabulary is excluded from the list, the frequency of a word must not exceed a certain number in each of the disciplines (Gardner & Davies 2014: 315-316). In addition, the AVL uses lemmas as the unit of counting, which was possible due to the words having been taken from the COCA and already having been tagged for parts of speech (Gardner & Davies 2014: 313). However, since many other word lists selected the word family as the unit of counting, the AVL was also converted into a word family-based list to facilitate comparisons between the lists (Gardner & Davies 2014: 321).

Another word list selecting lemmas as the unit of counting was developed by Brezina and Gablasova in 2013, namely the *New General Service List* (*new-GSL*). In order to compile this list they not only compared four corpora (LOB, BNC, BE06 and EnTenTen12) to determine a "common lexical core" (Brezina & Gablasova 2015: 9) but also created a list of current words

that represent "present-day use of the English language" (Brezina & Gablasova 2015: 9). Subsequently, the common lexical core (consisting of 2,116 items) and the current words (consisting of 378 items) were combined to generate the new-GSL with a total of 2,494 items. Despite the GSL including almost twice as many items as the new-GSL (2,000 word families in the GSL equalling approximately 4,100 lemmas as opposed to 2,494 lemmas in the new-GSL), the coverage of the two lists is extremely similar (Brezina & Gablasova 2015: 17). However, some researchers offer criticism regarding the new-GSL. Stein, for example, criticises the new-GSL's "striking omissions and strange analyses" (2016: 761) as, for instance, for some items that can belong to more than one word class some of the word classes are omitted. Furthermore, she suggests a revision of the new-GSL in the course of which it should be shortened by merely including overlapping items with Stein's Common Core Vocabulary (2016: 763).

In the same year another word list carrying the name *New General Service List (NGSL)* was developed by Browne, Culligan and Phillips (2013) in order to create an update of the original GSL by West (1953). The NGSL was derived from a 273 million word sample of the Cambridge English Corpus (CEC) (Browne 2014: 1), including "texts drawn from fiction, journals, magazines, non-fiction, radio, documents, TV, spoken interactions, and learner output" (Stoeckel & Bennett 2015: 2) of which 2,800 high-frequency words were selected (Browne 2014: 1). In the NGSL a modified lexeme approach is used, where - in contrast to using lexemes - "all the inflected forms from the different parts of speech" (Browne 2014: 6) are included. Moreover, Browne, Culligan and Phillips have also compiled a *New Academic Word List (NAWL)*, which supplements the NGSL in the same way that West's orginal GSL worked with Coxhead's AWL. This is particularly beneficial for the present study as by measuring learners' LFPs with both a high-frequency and an academic word list more insight regarding their lexical proficiency can be gained. The NAWL can be considered an update of Coxhead's AWL and was derived from "a 288 million word academic corpus" (Browne 2021: 4). Moreover, when lemmatizing the original GSL in order to be able to compare the two lists, it could be seen that the NGSL covers 6.1% more texts with 800 fewer lemmas than the GSL (Browne 2014: 6). Furthermore, the NAWL combined with the NGSL covers 92% of most academic texts (Browne 2021: 51) and hence more than the AWL combined with the GSL with merely 86% coverage of the Academic Corpus (Coxhead 2000: 222).

Browne, Culligan and Phillips' decision to use the modified lexeme as the unit of counting in the NGSL was questioned by other researchers, such as Laufer, as derivatives with a lower learning effort are not included by the modified lexeme approach whilst words with a higher one are (Laufer 2014: 40). Nevertheless, the NGSL and NAWL have been used in studies investigating productive proficiency. Higginbotham and Reid (2019), for instance, have analysed longer essays of 2,000 words produced by university students using the LFP based on the GSL and AWL as well as on the NGSL and NAWL. Interestingly, they have found a correlation between learners receiving lower scores and using a higher number of high-frequency words, with the correlation being stronger when using the GSL and AWL than the NGSL and NAWL. The authors speculate that this might partly be due to marker reliability issues or the older lists influencing teaching materials (Higginbotham & Reid 2019: 136). They conclude that while LFP measures in general should not be the sole indicator for productive proficiency, they are certainly useful when used in combination with other measures (Higginbotham & Reid 2019: 136).

This section has shed light on the most important word lists existing as well as their respective strengths and weaknesses. As mentioned in chapter 2.3, selecting lemmas as the unit of counting is particularly advisable when measuring productive knowledge (Nation & Webb 2011: 201). This is also applicable for the present study as productive knowledge in the form of written compositions is investigated. Hence, a word list using lemmas was selected, which, as the discussion above has shown, leaves the AVL, the new-GSL, the NGSL and the NAWL (with the latter two selecting the modified lemma as the unit of counting). As will be described in more detail in chapter 4.3, the NGSL in combination with the NAWL was selected due to the possibility of using both lists in order to determine the LFPs of learners using VocabProfile. When analysing learners' LFPs with regard to their academic vocabulary, the classic version of the VocabProfile is often selected which uses the GSL and the AWL. However, due to the GSL being rather outdated (as discussed at length above) and the NGSL and the NAWL offering a higher coverage of academic texts, it was decided to select the newer version of the VocabProfile using the NGSL and the NAWL.

3. Different language learning contexts and vocabulary acquisition

Since numerous studies have found that the amount of exposure to a language affects a person's proficiency in that language, this chapter addresses five factors that can lead to an increased exposure to a second language. In the first subchapter a working definition of bilingualism will be established, followed by potential effects a bilingual upbringing can have on language and vocabulary learning. In the subsequent sections, other language learning contexts, such as bilingual education, spending time abroad or exposure to Extramural English will be introduced. Moreover, it will be discussed whether gender has an effect on language and lexical proficiency.

3.1 Factors connected to bilingualism

In this section, numerous definitions of bilingualism and different categories of bilinguals will be introduced and the potential effects of a bilingual upbringing on language and vocabulary learning will be discussed. Subsequently, the potential influence of bilingualism on third language acquisition will be explored as this is particularly interesting for the present study with regard to those participants that have been raised bilingually and learned English as a third language.

First of all, it is crucial for the present study to establish a working definition of what bilingualism is. As Baker illustrates in great detail, it is difficult to determine who is bilingual and who is not (Baker 2001: 4-14). For some researchers, people can only be considered bilingual if they are equally fluent in both languages (Steiner 1992, quoted in Edwards 2006: 7). Similarly, Bloomfield, as early as 1935, defined bilingualism as having "native-like control of two languages" (Bloomfield 1935: 56). Other linguists have a broader definition of bilingualism. Whilst for Haugen (1953: 7), bilingualism merely means being able to communicate effectively in another language without the demand for perfection, Weinreich even states that bilingualism is simply "the practice of alternately using two languages" (Weinreich 1968: 1).

In addition, a distinction between different types of bilinguals is often made with regard to not only the time and age of acquisition but also the level of presence of the first language. Beardsmore, for instance, differentiates between consecutive bilinguals, meaning that "two languages are present from the onset of speech" (Beardsmore 1986: 29) and successive

bilinguals, who developed their first language before acquiring a second one (Beardsmore 1986: 29). However, different terms can also be found in literature on this topic as McLaughlin, for example, refers to simultaneous bilinguals if they have started acquiring the second language before the age of three and successive bilingualism if the second language is learned after that age (McLaughlin 1984: 101, quoted in Beardsmore 1986: 29). Furthermore, a differentiation between additive and subtractive bilingualism can frequently be encountered in previous research. The former refers to the acquisition of an additional language without one's first language being replaced or neglected (Lambert 1974: 25). In contrast, the latter refers to "ethnolinguistic minority groups who (...) feel forced to put aside or subtract out their ethnic languages for a more necessary and prestigious national language" (Lambert 1974: 25).

In the context of defining up to which age of acquiring a second language a person can be considered bilingual, the so-called critical period is an important notion as well. This concept is "the idea that there is a biologically based critical period for second-language acquisition that prevents older learners from achieving native-like competence" (Hakuta, Bialystok & Wiley 2003: 31). The critical period is frequently used to distinguish between early and late learners of a language, with the line usually drawn at the age of six (Wattendorf & Festman 2008: 5). This distinction between early and late learners will also be used as a categorisation for the present study and hence, all participants that have started learning a second (or third) language before the age of six will be assigned to the bilingual group.

Furthermore, as mentioned in the introduction, there are also different contexts in which a second language can be acquired, such as being raised bilingually, moving to a different country or simply learning it in an educational context. For a long time a bilingual upbringing was perceived to be unfavourable for children due to the increased burden with regard to the acquisition of two languages (Bialystok et al. 2009: 97). However, as Bialystok et al. (2009: 97) mention, these notions of bilingualism having negative connotations regarding children were somewhat rebutted by Peal and Lambert in 1962, when they gave multiple intelligence tests to monolingual and bilingual children and found that the latter outperformed monolinguals in all of the tests (Peal and Lambert 1962, quoted in Bialystok et al. 2009: 97). Since then, countless studies have focused on the effect being raised bilingually can have on language development and proficiency.

One of these effects of bilingualism, which is also mentioned rather frequently in the literature, can be observed with regard to executive function (Carlson & Meltzoff 2008: 282; Bialystok & Viswanathan 2009: 494), which includes areas such as working memory, inhibition or switching attention (Miyake et al. 2000: 50). Malakoff and Hakuta, for instance, have observed advantages of bilinguals regarding cognitive development and add that these are particularly discernible in balanced bilinguals and in the case of additive bilingualism (1991: 141). Other studies have also detected that bilinguals outperform monolinguals in tasks testing attentional control (Kapa & Colombo 2013; Krizman et al. 2014). This could be explained by the fact that bilinguals are more frequently confronted with attentional control due to both languages being active and having to select the appropriate one (Bialystok 2009: 3). In spite of some evidence indicating that bilinguals may have advantages over monolinguals regarding inhibition (see, for example, Bialystok & Martin 2004), Bialystok notes that these results should be taken with a grain of salt due to the lack of entirely clear evidence of a bilingual advantage regarding inhibition. Moreover, it also has to be kept in mind that bilinguals outperform monolinguals on numerous different tasks, including some that do not require inhibition (2015: 118). Furthermore, despite numerous studies reporting a higher performance of bilinguals in tasks testing another area of executive control, namely working memory (see for example Soliman 2014 or Bialystok, Craik, Klein & Viswanathan 2004), studies not being able to replicate this advantage of bilinguals have cast a doubt on whether this benefit actually exists (Grundy & Timmer 2017: 326; Morales, Calvo & Bialystok 2013: 189).

In contrast to these observed advantages of bilinguals, some researchers have also reported that monolinguals outperform bilinguals in some respects. Bialystok, for instance, states that numerous studies have demonstrated that bilinguals have a smaller vocabulary in each language when compared to monolinguals (2009: 4). Amongst these is, for example, a study by Bialystok and Martin, which detected that monolinguals achieved higher scores, particularly regarding receptive vocabulary (2004: 332). This stands to reason, as the exposure of bilinguals to each language is limited when compared to that of monolinguals. Furthermore, it has been found that the vocabulary size regarding an educational context is relatively similar amongst monolinguals and bilinguals (Bialystok et al. 2010: 530). This might

mainly be attributed to the fact that despite speaking different languages at home, both groups are primarily exposed to formal vocabulary in school.

In addition to some researchers not having found beneficial effects of bilingualism, there are even different beliefs or hypotheses concerning a potential harming effect of a bilingual upbringing. One of these is that children that grow up learning two or more languages risk becoming 'semilinguals', i.e. "not having 'sufficient' competence in either language" (Baker 2001: 9). Although it is usually the case that bilingually raised children or bilingual people in general are dominant in one of the languages they know and thus usually never have the same competence in both languages (which can also easily change over time, e.g. by moving) (Baker 2001: 9), there are no reasons presented in research why bilinguals should be semilingual and thus not be able to achieve a high linguistic competence in both languages (MacSwan 2000: 11). Therefore, some researchers have suggested that semilingualism is merely a political label (Baker 2001: 9) and Hinnenkamp even considered it a political weapon (Hinnenkamp 2005: 62). Another hypothesis focusing on potential negative consequences of bilingualism was proposed by Cummins, namely that there is "a threshold level of linguistic competence which a bilingual child must attain in order to both avoid cognitive deficits and allow the potentially beneficial aspects of becoming bilingual to influence his cognitive functioning" (Cummins 1976: 3). Similarly, to the debate on semilingualism, Cummins' Threshold Hypothesis is rather controversial with opinions of researchers ranging from proposing adaptions to discarding the hypothesis altogether (see, for instance, Ardasheva, Tretter & Kinny 2012 or MacSwan 2000).

Third language acquisition

The previously discussed controversy about scientific evidence for potential positive and negative effects of bilingualism on language and vocabulary acquisition is also continued with regard to the relationship between bilingualism and third language acquisition. Numerous studies give reason to believe that bilinguals have advantages over monolinguals when learning an additional language since they are "more linguistically attuned to language learning" (Baker 2001: 271). As they have already made the experience of learning two languages, they might be able to use this linguistic repertoire when acquiring a third language (Cenoz 2013: 72). Not only might they be able to use certain strategies or skills they have already obtained (Cenoz 2013: 72) but there might also be positive transfer from

having previously learned another language (Baker 2001: 271). Further evidence for this was also found in a study with bilinguals conducted by Bartolotti and Marian as the participants transferred vocabulary items from both their L1 and L2 when acquiring a third language (2017: 131). This demonstrates that the native language(s) of a learner can be utilized for scaffolding purposes when learning an additional language (Bartolotti & Marian 2017: 133). Hence, bilinguals might have an advantage concerning language acquisition due to having more languages at their disposal to use for scaffolding. Another study reporting similar results was conducted by Sanz, who showed that bilingual students that were participating in a Catalan immersion programme outperformed students of the monolingual school in the vocabulary and structure sections of the Comprehensive English Language Test for Speakers of English as a Second Language (CELT) (Sanz 2000: 33-34). Similarly, Keshavarz and Astaneh reported that bilingual participants scored higher in a vocabulary test. Interestingly, those bilinguals that had learned their L1 and L2 both academically and orally performed better regarding vocabulary production and achievement than those who had merely learned it orally (Keshavarz & Astaneh 2004: 295). Hence, the results of these studies support the theory that bilinguals have an advantage over monolinguals when acquiring an additional language.

However, Cenoz, who gave a comprehensive overview of the studies conducted on various different areas of third language acquisition, states that some studies were not able to replicate these findings (Cenoz 2003: 77). Despite many of them indicating that bilinguals outperform monolinguals, this also seems to depend on the area of language learning that is being tested, as advantages of bilingualism are reported more frequently when testing general proficiency than more specific aspects (see Cenoz 2003: 82-83 for more details). It is also worth mentioning that the level of proficiency in the L1 and L2 seems to play a considerable role regarding third language acquisition (Thomas 1988: 240, quoted in Cenoz 2003: 78). This was also mentioned by Sagasta Errasti since those bilinguals who had a higher competence in both their L1 and their L2 achieved the highest scores when producing a written composition in English (Sagasta Errasti 2003: 39).

In conclusion, this subchapter reviewed the existing research on bilingualism and its effect on language and vocabulary learning. As could be shown, there are somewhat contradictory findings regarding the potential impact of bilingualism on language and vocabulary

acquisition. Whilst some researchers have reported bilingualism to be advantageous regarding language and vocabulary acquisition, others have either failed to find such advantages or have even suggested that bilingualism might have negative effects on it. This controversy indicates that further research is required to detect potential benefits or drawbacks of a bilingual upbringing. Nevertheless, it has to be mentioned that, as has been shown in numerous studies, there is reason to believe that being raised bilingually might be advantageous for some areas of executive control. Since vocabulary acquisition and certain aspects of executive control, particularly working memory, can be considered closely intertwined (Baddeley, Gathercole & Papagno 1998; Ellis 2001; Vulchanova et al. 2014), one could assume that bilinguals have benefits regarding the acquisition of vocabulary. The present study is an attempt to make a contribution to this research field, as it will investigate whether bilingually raised participants show a higher lexical proficiency in their academic texts than monolingually raised ones.

3.2 Bilingual education

Contrary to the widespread belief that bilingual education, i.e. two different languages being used as a medium of instruction, is a modern phenomenon, bilingual education has not been introduced only recently (Abello-Contesse, Chandler, López-Jiménez & Chacón-Beltrán 2013: xx) but has actually existed for several hundred (Dalton-Puffer 2007: 2) or even thousands of years (Baker 2001: 182). Nowadays, numerous types of bilingual education exist (see Baker 2001 for a more detailed categorisation of the different types of bilingual education), of which the most relevant ones for the topic of this paper will be introduced in the following. One form of bilingual education are immersion programmes, which aim for the children to become bilingual and bicultural (Baker 2001: 204). Other forms of bilingual education focus on majority languages (Baker 2001: 221), for instance international schools, which are often attended by children of geographically mobile parents and feature "private, selective, independent education" (Baker 2001: 223). In contrast to immersion programmes, which cannot be found in Austria, international schools do exist in Austria. The number of such institutions in Austria is, however, limited (Schmid & Gruber 2012: 3).

A programme that is increasingly common, especially in European countries (Abello-Contesse 2013: 12), is Content and Language Integrated Learning (CLIL), which refers to

"educational settings where a language other than the students' mother tongue is used as medium of instruction" (Dalton-Puffer 2007: 1). Hence, as opposed to the traditional EFL classroom, a foreign language is used to teach a content subject. Although, in theory, any language could be used for this purpose, usually English is the foreign language that is selected as a medium of instruction (Dalton-Puffer 2007: 1). CLIL is thus a form of bilingual education, which Abello-Contesse considers a "semi-partial-immersion [...] program"[1] (Abello-Contesse 2013: 10) due to the "level of intensity in the L2" (Abello-Contesse 2013: 12). However, it is special in the sense that the key concept is that language learning and subject learning ought to be combined, which can be said to lead to "tension and sometimes conflict between the two areas" (Dalton-Puffer 2007: 5). Although it might be argued that a language is best and easiest learned in the environment where it is naturally spoken, i.e. in the country in which it is spoken, this is usually not possible when learning a foreign language. CLIL seems to offer a solution to this problem as it enables learners to use a language naturally and thus, makes authentic communication possible (Dalton-Puffer 2007: 3).

As the desired results of CLIL lessons are "high levels of functional bilingualism and biliteracy" (Abello-Contesse 2013: 14), it is only natural that countless studies have been conducted on the relation between CLIL and its potential benefits. However, there is little consensus between researchers as some studies have found that CLIL has positive effects on language proficiency, whilst others could not observe that CLIL students experienced an increase in language proficiency when compared to non-CLIL students. Regarding an advantage concerning vocabulary, which is one of the most explicitly addressed aspects in the CLIL classroom (Matiasek 2005, quoted in Dalton-Puffer 2008: 144), Dalton-Puffer states that CLIL students usually "possess larger vocabularies of technical and semi-technical terms and possibly also of general academic language" (2008: 144). This was also affirmed by Meriki and Pietilä's study, which showed that CLIL students clearly outperformed non-CLIL students regarding both receptive and productive vocabulary (2014: 495). Similar positive

---

[1] Whilst total immersion programmes usually mean that every subject is entirely taught in the foreign language, in partial immersion programmes a certain percentage of subjects is taught in the foreign language (de Courcy 2002: 5). Abello-Contesse's reference to CLIL being a *semi-partial immersion programme* might hence refer to the lower amount of hours taught in the foreign language in CLIL programmes when compared to (partial) immersion programmes (although, as mentioned, CLIL programmes can differ substantially from each other in terms of the extent to which a foreign language is used for instruction).

effects of CLIL were reported by Ruiz de Zarobe, who found a significant difference regarding the type-token ratios of CLIL students as opposed to the non-CLIL group (2008: 70). Furthermore, Lasagabaster analysed written texts, which showed that CLIL students significantly outperformed non-CLIL students in several aspects, including vocabulary (2008: 38). Agustín-Llach also investigated written productions and observed slight tendencies showing an advantage of CLIL which were, however, not significant (2016: 92). In addition, Gierlinger and Wagner found a significant growth of receptive vocabulary of CLIL students within the 1,000 most frequent words (2016: 53-54) but no difference regarding the students' overall receptive vocabulary growth.

These studies already indicate that whilst some researchers detected significant advantages of CLIL students, others merely observed weak tendencies which cannot be generalized. Furthermore, numerous studies found no influence of CLIL on the students' linguistic competence at all, leading to doubts whether CLIL can be considered beneficial for language development. Amongst these was, for instance, a study conducted by Agustín-Llach, in which written productions of primary school children (9-10 years old) were analysed. Despite the CLIL students' input in English being almost double of that of traditional learners, no significant differences could be observed (Agustín-Llach 2017: 565). However, the relatively young age as well as the participants' low L2 proficiency were given as potential reasons for the lack of positive effects of CLIL (Agustín-Llach 2017: 567-568). Similar to Agustín-Llach, Arribas did not find significant differences in receptive vocabulary tests by CLIL students and non-CLIL students (Arribas 2016: 284). Moreover, Bruton (2011: 237) addresses issues with regard to some of the studies on CLIL, for instance, that without a pre-test of both the CLIL and the non-CLIL students, it is not possible to assess whether there has been an improvement which was caused by CLIL or not. Similarly, Paran is critical of CLIL and argues that in order for it to be beneficial it has to be implemented selectively (Paran 2013: 325). Moreover, he states that it is dependent on the language competence and educational level of CLIL teachers (Paran 2013: 327-328) and "works best with high achievers" (Paran 2013: 326) or in private education (Paran 2013: 330).

These very diverse studies clearly show that more research is needed to clarify the potential benefits of CLIL lessons and under which circumstances they can evolve. Furthermore, the literature on the effect of CLIL on academic vocabulary, especially with regard to tertiary

education, is extremely scarce. However, it also ought to be mentioned at this point that there is a "remarkable variety of practices" (Dalton-Puffer, Nikula & Smit 2010: 2) that can be considered CLIL, meaning that CLIL lessons can be implemented in various different ways regarding, for instance, the amount of lessons that are taught. This diversity of CLIL programmes has to be taken into consideration when reading studies such as the ones discussed above, as it renders comparisons and generalisations difficult.

3.3 Time spent abroad

It seems reasonable to assume that increased exposure to a language will lead to an improvement of a person's language proficiency. This exposure is, of course, tremendously higher when a person is in a foreign country and is inevitably confronted with the language for the majority of the day. Therefore, despite the lack of study abroad research until the early 1990s (Sanz & Morales-Front 2018: 1), a common belief is that by staying or studying abroad for a considerable amount of time, one's language competence increases almost automatically (Isabelli-García 2006: 231) which seems to be affirmed by an improvement of language skills of students returning from abroad (Brecht, Davidson & Ginsberg 1995: 37).

Although this belief can, of course, not be confirmed in its entirety, a high immersion in the target culture is nevertheless likely to result in an increase of a person's language proficiency. However, it also has to be mentioned that the acquisition of a language and its development depend on various aspects which are often connected to the individual, such as the attitude and motivation of the person staying abroad as well as their social network in the foreign country (Isabelli-García 2006: 254). Similarly, the required length of a stay abroad for it to cause significant changes in the language proficiency is not entirely clear to researchers. As Zaytseva, Pérez-Vidal and Miralpeix (2018: 213) as well as Briggs (2015: 132) mention, various studies have investigated whether the length of a stay abroad influences lexical development. Despite these studies reporting different findings, some of them give reason to believe that the effect the length of a stay abroad can have on the development of the participants' lexical proficiency should not be underestimated (Zaytseva, Pérez-Vidal & Miralpeix 2018: 213; Briggs 2015: 132). However, more research in this area is required as there is no indication to be found in literature concerning the minimum length of a stay abroad required to achieve an increase in lexical proficiency. Some researchers have already

found improvements after a rather short amount of time, such as Llanes and Muñoz, whose study has revealed that even a short stay of three to four weeks can result in higher scores regarding, for instance, oral fluency and accuracy (Llanes & Muñoz 2009: 362). However, other studies indicate that a longer stay abroad leads to a higher improvement. Ife et al., for instance, have found that when staying abroad for one semester the lexical knowledge improved only by a third of the improvement that students staying abroad for two semesters experienced (Ife et al. 2000, quoted in Zaytseva, Pérez-Vidal & Miralpeix 2018: 213). Moreover, the type of stay abroad might also have an impact on the development of language proficiency: whilst for oral fluency, a work-and-travel programme might be equally beneficial as studying at a university abroad, for other aspects of language proficiency, such as academic writing, being exposed to this specific type of vocabulary could be considered essential. However, Zaytseva, Pérez-Vidal and Miralpeix confirm that research on different types of stays abroad and their potential effects is relatively scarce, with the majority of studies focusing on studying abroad (2018: 220).

Despite these different factors that can affect the development of lexical proficiency during a stay abroad, such as motivation, social network or length and type of stay, it can be said that if a person spends a considerable amount of time in a foreign country and immerses themselves in the foreign culture (which also includes having enough contact to speakers of the target language and, in the best case, to native speakers (Lapkin, Hart & Swain 1995: 93)), improvements regarding various aspects of their language proficiency are very likely. Numerous studies have shown that spending time abroad can, for instance, be beneficial for a person's oral proficiency development (Wright 2018: 175; Juan-Garau 2018: 204) and their pronunciation in the foreign language (Solon & Long 2018: 81).

Whilst research on the influence of a stay abroad on language proficiency seems to be steadily increasing, relatively few studies have investigated the relation between studying abroad and lexical development according to Zaytseva, Pérez-Vidal and Miralpeix, who have made an attempt to review the existing literature on this topic (Zaytseva, Pérez-Vidal & Miralpeix 2018: 210). Studies that did investigate this relation indicated that "vocabulary is among the top areas that improve the most after an immersion experience, well above reading, writing, and grammar skills" (Zaytseva, Pérez-Vidal & Miralpeix 2018: 210). Barquin, for instance, has found a significant improvement of lexical diversity in students' writing

after having spent a semester abroad but no significant increase in lexical sophistication (Barquin 2012: 201). Whilst a number of studies indicate that receptive knowledge of vocabulary is affected positively during a longer stay abroad, there is little consensus among researchers as to whether productive knowledge increases more orally or in writing (Zaytseva, Pérez-Vidal & Miralpeix 2018: 211). Moreover, as already mentioned, the type of stay abroad might have a considerable impact on the aspect of language proficiency that is improved most. Hence, further studies in this area are required to investigate whether correlations exist between a higher lexical proficiency - especially regarding written academic vocabulary - and a stay abroad.

3.4 English used in leisure time

As numerous studies have shown, at least a part of the vocabulary people learn is acquired incidentally, i.e. it is not studied on purpose but rather acquired through other activities, such as reading books, watching films or even playing online games (see, for instance, Peters & Webb 2018; González Fernández & Schmitt 2015; Sylvén & Sundqvist 2012; de Wilde, Brysbaert & Eyckmans 2020). Therefore, this chapter will focus on cognitive activities outside formal education that influence vocabulary acquisition, namely reading books or magazines, watching videos or films, playing video games and talking in English.

There are multiple terms that are being used for subsuming activities in students' leisure time for which the English language is used. One of these is *Extramural English* (EE), a term that was coined by Sundqvist in 2009 which "refers to the English that learners come in contact with or are involved in outside the walls of the classroom" (Sundqvist 2009: 24). It is important to note that activities that can be considered EE are usually not engaged in with the aim of acquiring the English language. That being said, EE does not exclude intentional learning (e.g. when students frequently expose themselves to English films outside of the English classroom with the aim of improving their English) (Sundqvist 2009: 25). Since activities such as watching TV, reading books or magazines and playing video games in English can all be considered activities that are usually engaged in outside the walls of the English classroom, the term *Extramural English* will be used in the present paper. However, due to the only relatively recent introduction of the term, previous studies have investigated so-called incidental learning, which, albeit being somewhat similar to EE, excludes the

possibility of students exposing themselves to EE on purpose to improve their language proficiency.

Since EE activities lead to an increased amount of exposure to the English language, it stands to reason to assume that a person's language proficiency is positively affected by it. This assumption has also been affirmed by numerous studies, which have shown that EE seems to be not only beneficial regarding language proficiency in general but also vocabulary acquisition. Peters et al.'s study (2019), for instance, has shown that EE is beneficial for receptive vocabulary. The study compared Flemish and Dutch students' vocabulary knowledge of two languages and found that despite the considerably lower number of years of English instruction as opposed to French, students were familiar with a higher number of words in English, which can partly be explained by a higher out-of-school exposure to English (2019: 770). Moreover, Perez, who refers back to Webb, states that "watching audio-visual input can be considered an important activity for L2 lexical development." (Webb 2015, quoted in Perez 2020: 770). These benefits concerning lexical proficiency were also corroborated by Sundqvist, who found a statistically significant correlation between EE and lexical proficiency. Interestingly, this correlation could only be found for the male participants, whereas for females there was no such correlation (2009: 150). According to Sundqvist, this difference might be due to "boys and girls engag[ing] in totally different types of EE activities" (Sundqvist 2009: 203). She therefore concludes that the type of activity can considerably affect the influence EE has on language proficiency.

As Olsson and Sylvén state, EE can indeed be beneficial for the development of L2 vocabulary. They conclude that the benefits of watching films or reading books in English for vocabulary size and range might, in fact, even be comparable to those of CLIL (Sylvén 2004, quoted in Olsson & Sylvén 2015: 77). However, with regard to potential advantages for the development of academic vocabulary, little research has been done so far (2015: 80). The results of the study by Olsson and Sylvén suggest that the positive effect of EE on academic vocabulary might be restricted to lower proficiency levels, as a correlation between the students' exposure to EE and the academic vocabulary found in their written productions was only observed amongst students with a lower proficiency (Olsson & Sylvén 2015: 94).

3.5 Gender

This chapter focuses on another aspect that might have an influence on language and vocabulary acquisition despite not being directly related to language exposure or learning, namely the gender of a person. For a considerable number of people, it almost seems to be common knowledge that females outperform males regarding language proficiency. However, this notion is by no means confirmed by scientific research but can instead best be explained with the help of the following quote:

> "There is widespread belief in many western cultures that females tend to be better L2 learners than males, but this belief is probably primarily a social construct, based on outcomes which reflect cultural and sociopsychological constraints and influences." (Saville-Troike 2005: 90, quoted in van der Slik, van Hout & Schepens 2015: 1)

Hence, females outperforming males in linguistic tasks or regarding language proficiency in general might merely be a reflection of the underlying social constructs they are inevitably confronted with, such as the belief that female students often excel at languages whilst male students perform better in more scientific subjects.

As mentioned above, research on the effect of gender on language proficiency is relatively scarce. Moreover, a large part of the research on the effect of gender on language proficiency focuses on L1 proficiency. Hence, in this chapter findings on the effect on learners' L1 will be discussed, followed by research on L2 proficiency.

First of all, it ought to be noted that there is no consensus amongst researchers as to whether neurophysiological differences between males and females exist and, if they do, what kind of linguistic effects these dissimilarities could potentially have (Steele 1987: 264). Nevertheless, as Shucard, Shucard and Thomas state, numerous studies have shown a "more rapid rate of development for language-related skills in infancy and early childhood" (1987: 284) amongst females. Shucard, Shucard and Thomas have also reported that in their study investigating potential gender-related differences, the rates of linguistic development of their male and female participants (that were three and six months old) differed. Their results are consistent with findings from other studies that suggest "earlier development of language-related functions in females than in males" (1987: 294). In addition, having reviewed the literature on this topic, Ullman, Miranda and Travers conclude that women

tend to outperform men in most language-related tasks (Ullman, Miranda & Travers 2008: 301). However, a limitation of the previous research also has to be mentioned, as researchers have primarily focused on investigating differences in the language of female and male participants under the age of two years (Ullman, Miranda & Travers 2008: 292). Hence, although it might be possible that this advantage of females continues in their further lives (Ullman, Miranda & Travers 2008: 292), tangible results are required in order to substantiate this hypothesis.

Regarding studies focusing on learners' L2, a large proportion of research on this topic investigates the question of whether gender plays a role in L2 acquisition and proficiency. However, the results have been rather inconsistent (Agustín Llach & Gallego 2012: 47) as will be discussed in the following paragraphs.

Some researchers have observed a female advantage concerning L2 acquisition, for instance Lin, who reported females performing better than males in vocabulary acquisition and retention (2011: 295). Moreover, van der Slik, van Hout and Schepens have found that their female participants scored higher in speaking proficiency and writing proficiency, whereas no gender-related difference could be observed concerning reading and listening proficiency (2015: 8-15). In addition, other researchers have found differences between males and females regarding vocabulary learning strategies. Jiménez Catalán, for instance, reported that females made use of a higher number and a wider range of vocabulary learning strategies than male participants (Jiménez Catalán 2003: 62). Hence, it cannot be ruled out that this might influence vocabulary acquisition of males and females to an extent great enough to affect their lexical proficiency and, further, to contribute to a difference between these two groups.

In contrast to the findings showing a female advantage, Pahom, Farley and Ramonda reported that their male participants performed significantly better with regard to lexical recall (2015: 168). Furthermore, other researchers stated that gender is not a relevant factor in L2 acquisition, such as Lee and Pulido who indicated that there was no difference in L2 vocabulary learning through reading between males and females (2017: 129). Similarly, Agustín Llach and Gallego found no considerable differences between males and females regarding receptive vocabulary acquisition (2012: 45).

In conclusion, it can be stated that despite the variety of research on the differences of language proficiency of males and females, further research is needed in order to find more compelling evidence of a potential gender-related advantage. Although some of the previously conducted studies found that females performed better concerning language or lexical proficiency and the acquisition thereof, other studies reported a male advantage or found no gender-related difference at all. Moreover, there is a lack of literature on gender-specific performance regarding academic vocabulary in written productions. In addition, numerous factors that might have an influence on supposed gender-related differences ought to be taken into consideration, such as testing formats but also motivation and attitudes (for further information see van der Slik, van Hout and Schepens (2015: 3)). Hence, this subchapter showed that despite a considerable amount of research on gender and second language acquisition existing, more research is required to determine the potential effect gender has on L2 acquisition and proficiency.

4. Methodology

This study aims to discover the development of learners' lexical proficiency by analysing two written compositions produced by the same learners at different points in time. Moreover, it aims to investigate a potential correlation between learners' lexical proficiency and their age, gender and linguistic background. Thus, answers to the following research questions will be sought:

1. Which aspects of lexical proficiency improved the most and which the least from the first written production to the second (i.e. from the Proseminar Linguistics 1 paper to the BA or BEd thesis)?

2. Is the development of lexical proficiency dependent on its initial level (i.e. do students with a lower initial lexical proficiency improve less or more than those with a higher one)?

3. Can correlations be found between lexical proficiency and the participants' linguistic background (i.e. bilingual upbringing or education, stays abroad and extent of Extramural English activities)?

4. Can correlations be found between lexical proficiency and the participants' age (at the time of the submission of their BA or BEd thesis) or gender?

4.1 Data collection

In this chapter, the data will be discussed, which consists of a questionnaire the participants filled in, as well as their Proseminar Linguistics 1 paper (the participants' end-of-term paper for their first seminar in linguistics usually produced during the 3rd semester of studies) and their BA or BEd thesis in linguistics produced at the end of year 3 or 4. Therefore, first of all, the design of the questionnaire will be explained, followed by a discussion of the areas which were investigated. Secondly, the written productions will be focused on, including their selection and adaptation to facilitate the subsequent analysis.

4.1.1 Questionnaire

The first part of the study was a questionnaire, which the participants were asked to fill in. The questionnaire was designed carefully adhering to the guidelines for a successful questionnaire (Dörnyei 2010) and using the website *soscisurvey.de*. The questionnaire consisted primarily of closed-ended items apart from several clarification questions (Dörnyei 2010: 38) if the participants stated that none of the options above applied to them as well as a voluntary comment section at the end. Moreover, specific open questions were included when enquiring, for instance, specific details about the participants' time abroad or their experience with Content and Language Integrated Learning (CLIL). After the retrieval of demographic information, such as age (at the time of submitting the BA or BEd thesis), gender and studies as well as the semesters in which the two papers were submitted, the participants' linguistic background was investigated, which can be divided into several different categories. The first category was multilingualism, including questions about the participants' and their parents' first language(s), a potential change in their ability to speak their first language(s) and their knowledge regarding other languages, followed by a focus on the educational context, namely the attendance of a bilingual school and Content and Language Integrated Learning (CLIL) lessons. Another aspect was whether time had been spent abroad and, in case of an affirmation, detailed information about this time was asked for. Subsequently, the participants' use of Extramural English was investigated by asking them to estimate their exposure to the English language regarding talking, watching TV, playing video games and reading as well as a potential alteration in this exposure during the time between the two written compositions. Lastly, it was enquired whether the participants deliberately studied vocabulary or consulted dictionaries when writing a paper and whether

this behaviour had changed during the time between producing the two papers. In order to ensure that the data and all other answers of the participants could be treated anonymously, the two written compositions could be uploaded directly into the online questionnaire.

4.1.2 Written productions

The second part of the study was the analysis of two academic texts previously produced by students of the English Department in the course of their studies. In the following, it will be illustrated according to which criteria the texts were selected. Moreover, the adaption of the texts in order to facilitate the subsequent analysis will be discussed.

Selection of written productions

For the analysis of participants' lexical proficiency, two written texts per student were selected. The texts under analysis were the paper written for the Proseminar Linguistics 1 and the BA or BEd thesis, both of which were produced during their studies of the BA English and American Studies or the Bachelor of Education for the subject English. The reason for these two written productions being chosen was the fact that for the majority of students there was a longer time span between writing them as opposed to other papers which are partly produced in consecutive semesters. Interestingly, this time span varied tremendously with one to eight semesters between writing the two papers. Since it could be assumed that the longer the time span between the productions is, the more apparent the difference in the participants' lexical proficiency might be, this will be revisited when discussing the results. Moreover, no distortion of the results due to varying writing tasks was expected as both texts can be considered academic written productions in the field of linguistics.

Data treatment

The written productions had to be adapted prior to the analysis in order to ensure that the results were not distorted due to various other factors but that any differences in the lexical proficiency scores resulted from the improvement of the participants' lexical proficiency.

First of all, the PS 1 Linguistics papers (2,500 words) were considerably shorter than the BA theses (8,500 words) and the BEd theses (7,000 words) which could potentially lead to a difference in the results of the measured lexical proficiency, as some aspects of lexical

richness, such as the type-token ratio or the LFP, are sensitive to the length of the written productions under analysis (Nation & Webb 2011: 250; Laufer & Nation 1995: 314). Therefore, the texts had to be adapted in order for all of them to be of the same length. The starting point for this process was the shortest PS Linguistics paper, on the length of which the other texts were based and shortened accordingly. As Read reports, researchers have had various different approaches towards finding a solution for this problem, such as selecting the first 250 words (or tokens) of every text or randomly selecting the words with the help of a computer programme (Read 2000: 202). For this study, a similar approach to the first one was chosen. However, it was not possible to simply select, for instance, the first 1,000 words of every text without causing potential distortions since this would have resulted in some texts merely consisting of introduction and literature review whilst others would also have included sections with fewer references to literature, such as the results section. Therefore, a different solution had to be found for this study. As the main body of most of the PS papers merely consisted of a literature review, it was decided that the first 700 words of the literature reviews of all written productions would be analysed. Moreover, since these parts of the texts consisted of a relatively high amount of quotations that had to be deleted, parts of the introduction and conclusion were included as well. Since the majority of texts had an introduction longer than 170 words and a conclusion longer than 130 words, these numbers were selected as benchmarks, leading to each text compilation being 1,000 words long. For those texts not containing an introduction or conclusion long enough to reach the benchmark, words from either the introduction or conclusion were added in order to yield sections of 300 words[2]. In other cases the literature review was not long enough as participants included a review of the literature on the approach or programmes they considered using in the methodology rather than the literature review. For this reason, these sections strongly resemble the literature review stylistically and also contain a similar number of references to literature. Thus, in these cases words were added from the methodology section to the literature review in order to reach the benchmark of 700 words as it can be assumed that this will have no affect on the lexical proficiency scores due to the similarity to the literature review.

---

[2] Table 4 in the appendix displays the exact number of words used from each section for all papers.

Secondly, as already addressed, a large number of quotations could lead to higher lexical proficiency scores despite the words not being actively produced by the author of the written composition. Therefore, all direct quotations appearing between inverted commas were deleted and not included in the analysis. However, indirect quotations were more difficult to detect, and therefore not excluded from the analysis.

Thirdly, for the programme VP-Compleat[3] to operate successfully, all numbers were spelled out and proper nouns, including names of people and places, titles of books or programmes as well as abbreviations and references to specific years were substituted with different codes. Moreover, vocabulary from languages other than English was substituted as well. Acronyms, such as ELF or CLIL, were not substituted but marked as acronyms, resulting in them being categorized as lexical words by *analyzemywriting.com*[4] and as words from the Off-List (similar to proper nouns and abbreviations) by VP-Compleat. Lastly, some items were adapted, for instance the symbol % which was changed to *percent*. The list of codes and changes used for this process can be found in table 5 in the appendix.

Moreover, there are different approaches as to how errors should be treated: whilst some researchers correct minor spelling errors when the words are used correctly (e.g. Meara & Bell 2001: 11; Laufer & Nation 1995: 315), others omit words when they are "clearly used incorrectly" (Laufer & Nation 1995: 315). The justification given for this procedure is that in the former case it can be assumed that the word is familiar to the person using it despite the spelling error, whilst in the latter case the words cannot "be considered as part of the subject's productive lexicon" (Laufer & Nation 1995: 315). Regardless of the choice of error treatment, Engber points to the importance of counting and categorising errors as their number and type can be insightful regarding a text's lexical richness (Engber 1995: 146). However, in contrast to the present study in which two rather long written productions by 28 students were analysed, previous research has mostly focused on shorter written productions of 150 to 400 words. Therefore, correcting the errors would not have been feasible due to the scope of this study. Moreover, as the participants were students of the BA English and American Studies or the BEd Teacher Education for the subject English in addition to having already learned English for at least eight years before taking up their

---

[3] The programme VP-Compleat was used to calculate lexical variation and lexical sophistication (see chapter 4.3 for more details).

[4] The website *analyzemywriting.com* was used to calculate lexical density (see chapter 4.3 for more details).

university studies, one might be able to assume that their lexical proficiency in the English language is sufficient enough to produce academic papers without an enormous number of errors that would lead to severely distorted results. Furthermore, an infrequent occurrence of errors concerning lexical precision might impact the lexical density and lexical variation of a text less than one might expect, as neither the proportion of lexical words nor the TTR would necessarily be affected by it. In addition, the focus of the present study was to analyse different aspects of participants' lexical proficiency rather than lexical precision. Due to these reasons errors were not corrected or omitted in the present study.

4.2 Participants

In total, 28 people participated in the study, five of whom were male and 23 female (the option *none of the above* was never selected). All of the participants were students at the University of Vienna, 20 of them were either enrolled in the BA English and American Studies at the time of participating in the study or had already completed their BA and 8 were enrolled in the BEd Teacher Education for the subject English. Students that had completed their BA or BEd at other universities were excluded from this study in order to establish a sample group that was as homogenous as possible and exclude potential differences in the university programme as a confounding factor. Moreover, solely students who had already submitted both the seminar paper for PS Linguistics 1 and their BA or BEd thesis were selected for the study.

4.3 Data analysis

Whilst just a few decades ago, researchers used to measure lexical proficiency by calculating with pen and paper (Meara & Miralpeix 2016: x), nowadays there are numerous tools focusing on different aspects of lexical proficiency. Due to this great variety of measures and programmes, only the most widely used and most suitable ones for this study will be discussed in the following.

Lexical variation

As mentioned in chapter 2.5.1, there are various different measures that can be used to calculate the lexical variation of a written text. As the TTR is often criticised for being sensitive to text length, numerous alternative indices have been developed to resolve this

problem, such as Guiraud's Index (Guiraud 1954), Advanced TTR, Guiraud Advanced (Daller, van Hout & Treffers-Daller 2003) and D (Malvern et al. 2004).

Guiraud's Index (G) is a measure that calculates the relation between types and tokens by dividing the number of types by the square root of the number of tokens ($G = types/\sqrt{tokens}$) and by doing so, attempts to decrease the effect of text length (Daller, van Hout & Treffers-Daller 2003: 200). However, since Guiraud's Index is merely a "mathematical transformation of the TTR" (Treffers-Daller, Parslow & Williams 2018: 321), some researchers did not consider it an acceptable alternative. Daller, van Hout and Treffers-Daller, for instance, have criticised existing measures of lexical variation as being unreliable and, as a consequence, have developed two new measures, namely Advanced TTR and Guiraud Advanced. Their measures only differ from the traditional TTR and Guiraud's Index in that they do not use types but advanced types. Hence, they distinguish between basic and advanced lexical items and thus "combine characteristics of type/token related measures with the notion of different layers of frequency/productivity in the lexicon" (Daller, van Hout & Treffers-Daller 2003: 197). By using these measures they claim that more significant results can be achieved even with a smaller sample size (Daller, van Hout & Treffers-Daller 2003: 217-218). Lastly, the measure D (Malvern et al. 2004) takes the issue of text length into consideration and thus creates several samples of text in order to calculate 16 TTRs. These are then displayed as a curve, which "can be summarised by a single parameter, D" (Meara & Miralpeix 2016: 22). The more lexically rich the text under analysis is, the higher the value of D (Meara & Miralpeix 2016: 25).

As this brief introduction has shown, there are numerous different measures available to analyse lexical variation as it is one of the most discussed aspects of lexical richness. The above-mentioned more complex measures doubtlessly have numerous advantages over the more traditional TTR, particularly regarding the issue of text length. However, numerous researchers (for instance, Nation & Webb 2011: 250; Meara & Miralpeix 2018: 22 or Daller, van Hout & Treffers-Daller 2003: 199) agree that these benefits are primarily of importance when using texts of differing lengths. When analysing texts of equal length, the criticism of the TTR being unreliable due to its sensitivity to text length does not apply and more complex measures, such as D, have been found to yield very similar results to the traditional TTR (Miralpeix 2006: 99). Hence, since all written productions analysed in the present study

were adapted to create texts with the exact same number of words, the traditional TTR is a valid measure (Schmitt 2010: 213) and using more complex measures such as the ones introduced above is not necessary.

As the measure of TTR has been used for a considerable amount of time, numerous (online) programmes measuring it are offered. However, since calculating the TTR is a rather straightforward procedure and the majority of websites and programmes are highly similar, the selection of the programme used for it is of marginal importance. The programme chosen to calculate the TTR for the present study can be found on Cobb's website *lextutor.ca* (Cobb, n.d.). Not only does this website offer an analysis using LFPs and various other lexical statistics but it also grants users more detailed insights into different aspects connected to the TTR, such as tokens per type or tokens per word family or lemma depending on the selected unit of counting.

Lexical sophistication

Various different programmes exist that focus on analysing lexical sophistication, with most of them using measures based on word frequency, such as the Lexical Frequency Profile (LFP) (see, for instance, Lindqvist, Gudmundson & Bardel 2013: 110; Kalantari & Gholami 2017: 8; Higginbotham & Reid 2019: 127). The LFP of a text can be analysed online on various websites or with various programmes, for instance by using Nation's Range programme (Nation n.d.), AntWordProfiler (Anthony 2021) or VocabProfilers on Cobb's website *lextutor.ca* (Cobb, n.d.).

Range is a programme that was developed by Paul Nation in order to analyse the vocabulary used in written text. It can be downloaded with the GSL/ AWL and the BNC (British National Corpus) lists and can be used for different measures, such as calculating distribution figures, numerous different frequency figures or texts' coverage of a certain word list (Nation, n.d.).

AntWordProfiler is a vocabulary analysis programme similar to Nation's Range but offers additional features. It can be downloaded from the website *laurenceanthony.net* and can be used to calculate statistical information about texts. Moreover, it helps analyse the coverage of different vocabulary levels as well as the coverage of different lists (Anthony 2021). For this purpose, the website offers the already adapted versions of the GSL/ AWL, BNC/ COCA and NGSL/ NAWL.

The website *lextutor.ca* offers a web-based alternative to Range and AntWordProfiler, namely Vocabulary Profilers. There are various different versions for different age groups or purposes. One of them, VP-Compleat, allows for texts to be analysed using different lists, such as the GSL/ AWL, BNC/ COCA or NGSL/ NAWL. VP-Compleat not only calculates the LFP of the learner that has produced the text under analysis but also other relevant statistical measures, such as type-token ratio or lexical density (Cobb, n.d.).

For the present study, VP-Compleat has been selected due to its extremely user-friendly online interface as well as the possibility to use the NGSL in combination with the NAWL. As was already mentioned in chapter 2.6 on word lists, these two lists by Browne, Culligan and Phillips (2013) will be used as they have been compiled more recently and provide a higher coverage of academic texts when compared to the GSL and AWL. Moreover, VP-Compleat is appropriate for learners from grade 9 to university and hence suitable for the participants of this study. VP-Compleat divides words (or, in the case of the NGSL/ NAWL, modified lexemes) into the first 1,000 most frequent words of the NGSL (henceforth NGSL1), the second 1,000 (henceforth NGSL2) and the third 1,000 most frequent words (NGSL3), words from the NAWL and words not occurring in any of the lists (henceforth Off-List).

Lexical density

In contrast to the previously mentioned measures of lexical richness, relatively few programmes exist for determining a learner text's lexical density. However, the calculation of lexical density is relatively uncontroversial, as the number of lexical words simply has to be divided by the total number of words. Moreover, deciding which words are lexical words (namely nouns, verbs, adjectives and adverbs) and which grammatical words (e.g. articles, prepositions, conjunctions) is a lot clearer than determining, for instance, which words can be considered high-frequency or low-frequency as is the case with lexical sophistication. Hence, a programme dividing written productions into lexical and grammatical words and subsequently calculating the lexical density was required. Despite the website *lextutor.ca* also offering an analysis of lexical density, the website *analyzemywriting.com* was chosen for the present study, due to the possibility to quickly check the results after the analysis in order to ensure the correct categorisation of words.

5. Results

The aims of the current study are to determine whether the lexical proficiency of the participants changed during the time between the first and second written production and whether the linguistic background of the participants affected this development. In this chapter, it will be discussed whether a development in lexical density, lexical variation and lexical sophistication between the two written productions (henceforth WP1 and WP2) could be detected and whether the initial proficiency level influenced the degree of development. Subsequently, differences in lexical proficiency due to individual learner differences, namely bilingual upbringing or education, time abroad, Extramural English as well as gender and age, will be addressed.

5.1 Development of different aspects of lexical proficiency

Lexical density

As table 1 below illustrates, the mean of the lexical density scores of the first written production was 56.44 whilst for the second written production it was 56.56. Hence, despite there being an increase in lexical density, it is only a minor difference, namely 0.12.

Table 1. Lexical density 1 and 2 scores

	N	Minimum	Maximum	Mean	Std. Deviation
Lexical density 1	28	50.10	62.20	56.44	2.79
Lexical density 2	28	51.50	62.73	56.56	2.86

In order to test for the significance of this difference in means, first a Kolmogorov-Smirnov test had to be conducted which revealed that both lexical density 1 and lexical density 2 were normally distributed (p=.200 respectively). Subsequently, a paired samples t-test was administered showing that the development of lexical density was not significant (p=.822). It is also interesting to mention that merely sixteen of the participants achieved higher scores for lexical density in the second written production compared to the first one, whilst twelve had a lower lexical density in WP2.

Lexical variation

For the type-token ratio the mean score of WP1 was 0.3643[5] and increased to 0.3696 in WP2 (see table 2 below).

Table 2. TTR 1 and 2 scores

	N	Minimum	Maximum	Mean	Std. Deviation
TTR1	28	0.31	0.43	0.3643	0.0333
TTR2	28	0.33	0.42	0.3696	0.0281

As the Kolmogorov-Smirnov test showed that both the TTR1 (p=.200) and the TTR2 (p=.103) were normally distributed, a paired samples t-test was administered, which showed no significance of the difference in means (p=.387). However, this is not surprising when looking at figure 1 below, which shows the TTR scores for both written productions:
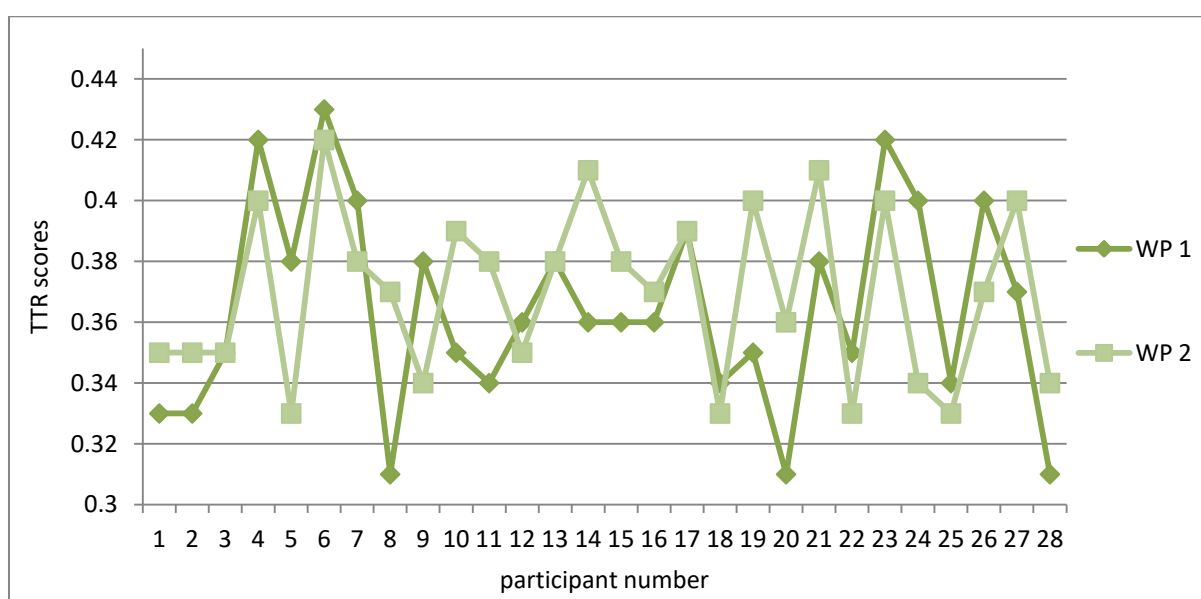
Figure 1. TTR1 and TTR2 scores for each participant

Similar to the results for lexical density, only thirteen of the participants had a higher TTR in the second written production than in the first one, whilst twelve had a lower TTR and for the remaining three participants the TTR stayed exactly the same.

---

[5] The scores for the TTR are presented in four-digit numbers in order to display the development more clearly as the range of the TTR scores is relatively narrow.

Lexical sophistication

Lexical sophistication was measured by using an LFP. Hence, there is not only one value as is the case with both lexical density and TTR but multiple scores for different values. For the sake of brevity, the different categories of the LFP will henceforth be abbreviated: for instance, the NGSL2 scores of the first written production will be labelled 1NGSL2, with the first digit referring to the written production and the second one to the NGSL category.

The following figure illustrates all of the mean scores of lexical sophistication in both written productions.
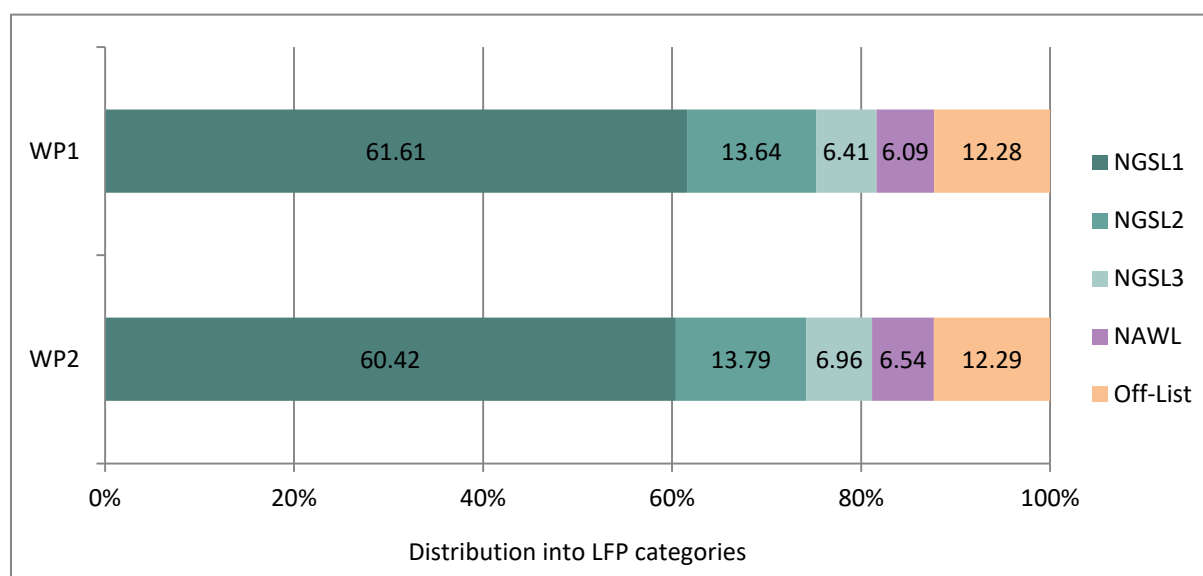
Figure 2. Distribution of words into different categories of the LFP in WP1 and WP2

As can be observed, both written productions mostly consisted of words assigned to the NGSL1. It also has to be noted that the percentage of words from the first 1,000 words was slightly lower in the second written productions (from 61.61% to 60.42%). In addition, a slightly higher number of words from the NGSL2 were used and the percentage of words from the NGSL3 increased as well. Moreover, words from the NAWL could be found more frequently in the second production. Interestingly, the percentage of words assigned to the Off-List stayed almost exactly the same.

In order to analyse whether the differences between WP1 and WP2 were significant, paired samples t-tests for the normally distributed variables and a Wilcoxon signed-rank test for the not normally distributed variable (i.e. the 1NGSL2) were conducted. As table 3 illustrates, no significant differences could be detected.

Table 3. T-tests and Wilcoxon signed-rank test for development of lexical sophistication

Pairs	Mean	Std. deviation	Std. Error Mean	Two-Sided p
1NGSL1 - 2NGSL1	1.18607	4.50239	.85087	.175
1NGSL3 - 2NGSL3	-.54821	1.71318	.32376	.102
1NAWL - 2NAWL	-.45679	2.02580	.38284	.243
1Off-List - 2Off-List	.00000	4.39124	.82987	1.000
Pairs	Z	Asymp. Sig. (2-tailed)		
1NGSL2 - 2NGSL2	-148	.882		

Hence, despite the participants indeed having used a lower percentage of words from the NGSL1 and a higher percentage of vocabulary assigned to the NGSL2, NGSL3 and NAWL in WP2, the difference was only a minor one.

5.2 Differences regarding the initial proficiency level

Apart from discussing different aspects of lexical proficiency it is also interesting to consider the participants' initial level of lexical proficiency and its potential effect on their development. For this purpose, the three lowest scoring participants (group 1) and the three highest scoring ones (group 2) were selected for lexical density and TTR. As the LFP yielded multiple scores, the NAWL was chosen as the most insightful aspect for this purpose.

Lexical density

The following figures illustrate the results of the lowest scoring students for lexical density on the left (figure 3) and the highest scoring students on the right (figure 4).
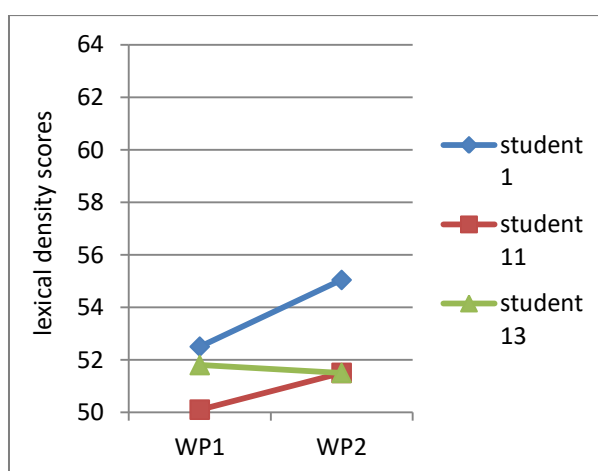

Figure 3. Lexical density scores of group 1 in WP1 and WP2; difference in means: 1.31
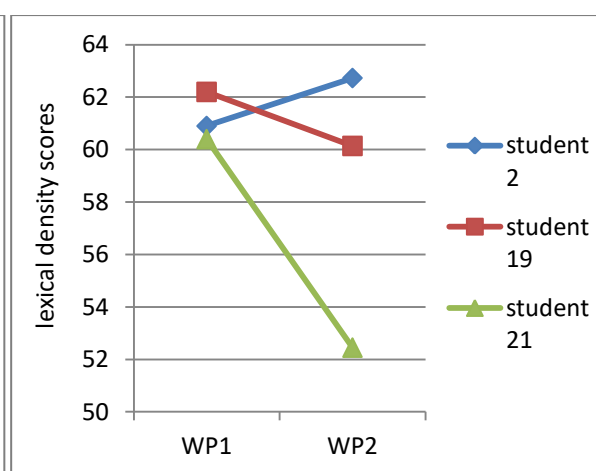
Figure 4. Lexical density scores of group 2 in WP1 and WP2; difference in means: -2.73

It can be observed that two of the lowest scoring students achieved higher scores in WP2 than in WP1. In contrast to that, the lexical density of two of the highest scoring students was lower in WP2 when compared to WP1 (see figure 4). The difference in means between WP1 and WP2 for group 1 was 1.31, whilst for group 2 it was -2.73. This is particularly interesting when compared to the development of the mean of all participants, which was merely 0.12. Not only does this mean that the lowest scoring students had a higher development of lexical density when compared to the mean of all participants but they, in fact, considerably outperformed the highest scoring students regarding their development of lexical density. However, as the two figures above show very clearly, despite the higher development the lowest scoring students only partly outperformed the highest scoring ones concerning their scores in WP2.

TTR

Regarding the TTR of the written productions, the development according to the initial level of the participants is striking as well. As can be seen very clearly in figure 5 below, the performance of the three lowest scoring students improved considerably, with 0.0467 being the difference in means. In contrast, all of the three highest scoring students achieved a lower TTR score in the second written production, with a difference in means of -0.0167.
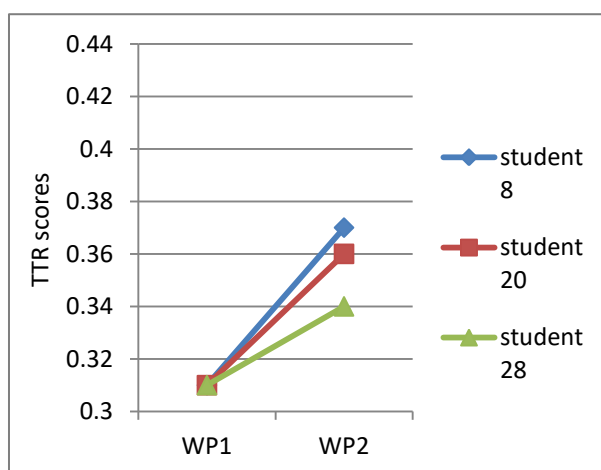
Figure 5. TTR scores of group 1 in WP1 and WP2; difference in means: 0.0467
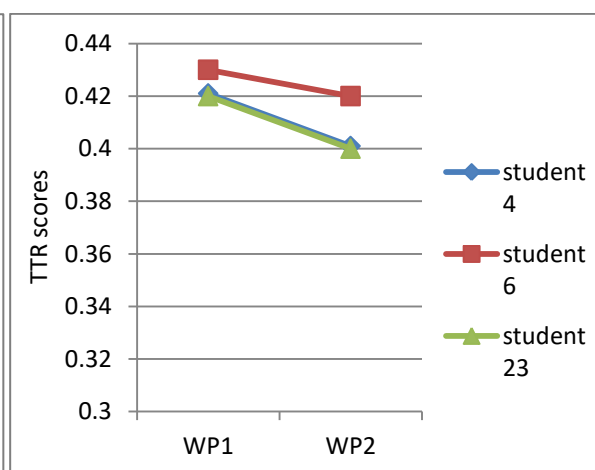
Figure 6. TTR scores of group 2 in WP1 and WP2; difference in means: -0.0167

This is particularly noteworthy when considering the difference in means of all participants of merely 0.0053. Nevertheless, it ought to be mentioned that despite the increase in the TTR of the lowest scoring students and the decrease in the scores of the highest scoring students, the latter group still achieved a higher TTR score in WP2.

52

NAWL

When comparing the two groups' percentage of the words used from the NAWL, the figures look rather similar to the ones for the TTR scores. Whilst the performance of the lowest scoring students (see figure 7) increased considerably in WP2 with a difference in means of 3.44, the highest scoring students (figure 8) achieved lower scores with the difference in means being -1.90. These numbers are particularly striking when considering the difference in means of all participants, which was 0.45. Hence, the lowest scoring students not only outperformed the average participant but also achieved equally high (or partly even higher) scores than the highest scoring students in WP2.
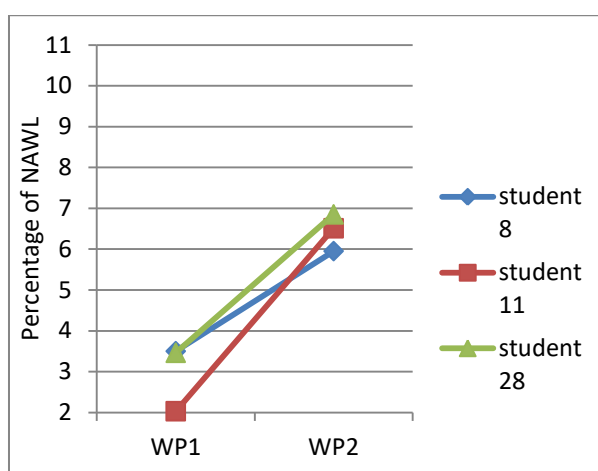
Figure 7. Percentage of words from the NAWL in WP1 and WP2, group 1; difference in means: 3.44
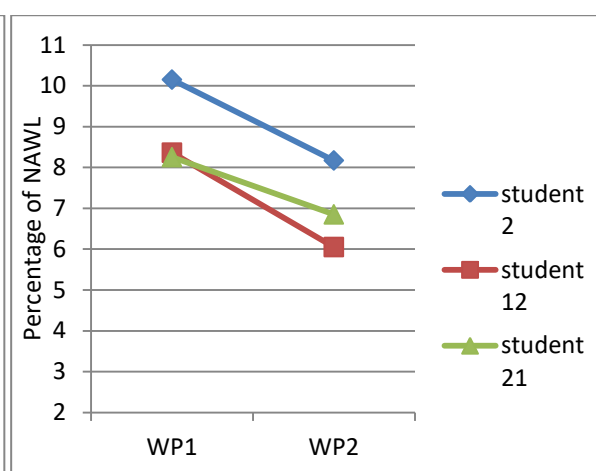
Figure 8. Percentage of words from the NAWL in WP1 and WP2, group 2; difference in means: -1.90

In conclusion, it could be observed that the three lowest scoring students had a considerably higher development of lexical density, lexical variation and the percentage of academic vocabulary used. Surprisingly, the highest scoring students experienced not only a lower development than the three lowest scoring students but also than the average participant, since the overall difference in means was higher.

5.3 Differences in lexical proficiency due to individual learner differences

As was thoroughly discussed in chapter 3, numerous aspects can affect the lexical proficiency of a person. In the present study, the factors that were of particular importance were gender and age, bilingual upbringing and education, time spent abroad and exposure to Extramural English. Hence, in this section potential correlations between aspects of lexical proficiency and individual learner differences will be presented. However, it has to be

mentioned at this point that it was not possible to recruit an even number of participants for all of these variables, i.e. the participants could not always be distributed equally into two groups, such as female and male or bilingually raised and monolingually raised. Hence, this limitation has to be taken into consideration with regard to the following results of the study.

Bilingual upbringing

The first factor that will be discussed is the number of languages learners had grown up speaking, i.e. whether they had been raised bilingually or not. Of the 28 participants of the present study merely three (10.71%) stated that they had been raised bilingually, whilst the remaining 25 had been raised monolingually. Due to this extremely low number of bilingual participants, the results presented below definitely have to be taken with a grain of salt as a generalisation of the findings is simply impossible. In order to investigate whether the number of languages learned when growing up had an impact on participants' lexical proficiency, the correlation between these two variables was measured. For this purpose, the Spearman Rank Order Correlation was used since this test neither requires a sample size larger than 30 participants nor data to be normally distributed.

Interestingly, the monolingual participants achieved higher scores for the TTR in both written productions. They also used a lower number of words from the NGSL1 and a higher number of words assigned to the NGSL2 and the NAWL in WP1. Furthermore, the monolinguals exhibited a higher development from WP1 to WP2 in both lexical density and TTR. In contrast, the bilingual participants achieved a higher lexical density in both written productions and used fewer words ascribed to the NGSL1 and more from the NGSL2 in WP2. However, as figure 9 shows, there were merely minor differences between the two groups.
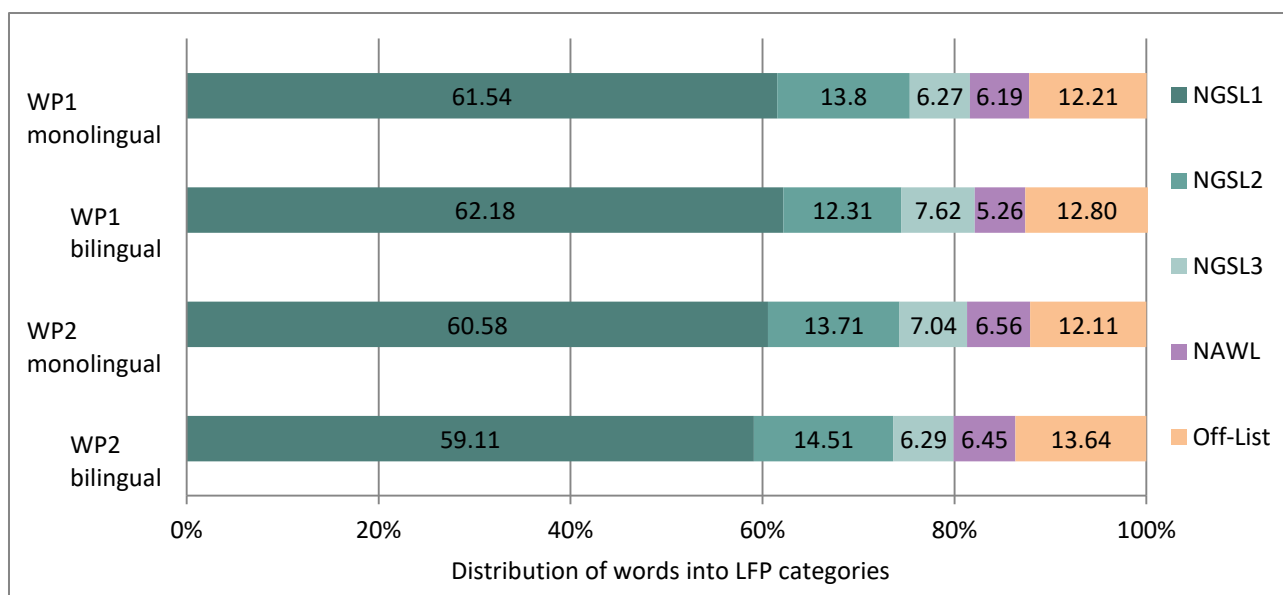
Figure 9. Distribution of words into different LFP categories based on monolingual or bilingual upbringing

Interestingly, the only significant difference between the two groups was the development of the use of words belonging to the NGSL3: whilst the monolingual group used 0.77% more words from the NGSL3 in WP2 than in WP1, the bilinguals used -1.33% less, with Spearman's rho being -0.393 (p=.039).

Bilingual education

Another factor resulting in a higher exposure to a foreign language is bilingual education, which five students (17.86%) had experienced. Whilst four of these participants stated that up to 25% of their courses had been held in a foreign language, one learner indicated that as much as 75-100% of the courses had been taught in English. Apart from one student whose bilingual course had been in Polish, all other students had had English as the language of instruction. Moreover, it ought to be mentioned that the student having received bilingual education in Polish had not attended a bilingual school but an extracurricular Polish school once a week for approximately ten years. The low number of participants that had attended a bilingual school has to be kept in mind concerning the results presented in the following. Overall, it can be observed that both groups achieved relatively similar results in both written productions, with the only significant difference being that those participants that had attended a bilingual school attained lower scores of lexical density in WP1 than those not having received bilingual education, namely 54.28 and 56.91 respectively (see figure 10), resulting in a p-value of 0.021.
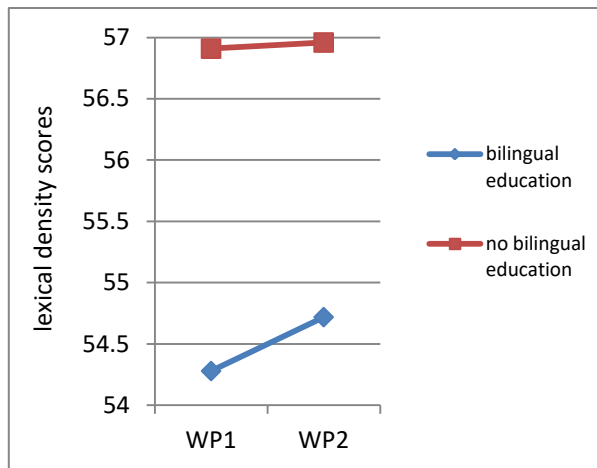
Figure 10. Lexical density scores of participants with and without bilingual education
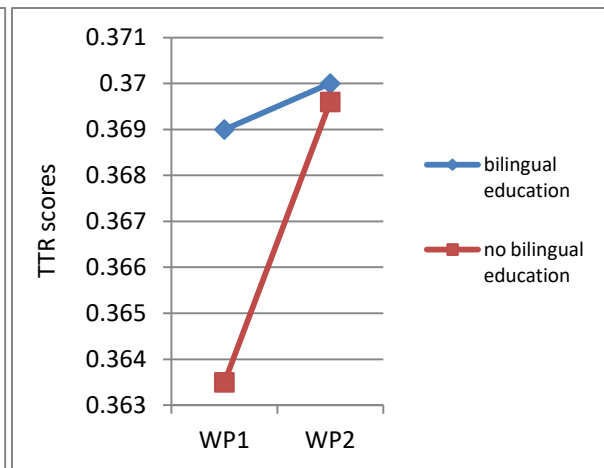
Figure 11. TTR scores of participants with and without bilingual education

Regarding the development of the numerous aspects of lexical proficiency from WP1 to WP2, no significant differences between the two groups could be detected albeit the participants that had not received bilingual education having scored slightly higher. Even though - as figure 11 above illustrates - those participants that had not attended a bilingual school having scored lower regarding the TTR1, they had a higher development of their TTR scores than the participants that had received bilingual education. Nevertheless, the difference between the two TTR1 scores was merely 0.0045 and hence only a very minor one.

Time spent abroad

Eleven participants (39.29%) had spent time abroad (which was defined as being in a foreign country for longer than one month) before completing their BA or BEd thesis. The majority of them (eight students) had been abroad once, one participant had spent time in a foreign country twice and two participants even three times. All of them had had to produce written assignments during their stay abroad. Whilst two participants stated that 50-75% of the communication during their stay abroad had been in the official language of the country, the remaining nine indicated that the official language had been used 75-100% of the time. Furthermore, 39.29% stated that they had frequently spent time abroad due to their family living there.

Regarding the scores of the two groups', merely minor differences could be detected: whilst the participants that had spent time abroad scored slightly higher in both TTR1 and TTR2, the other group achieved a higher lexical density in WP1 and WP2. Concerning lexical

56

sophistication, figure 12 below shows that the group that had been abroad used fewer words assigned to the NGSL1 and slightly more from the NGSL2 and NAWL in WP1 as well as a higher percentage of words from the NGSL3 and NAWL in WP2.
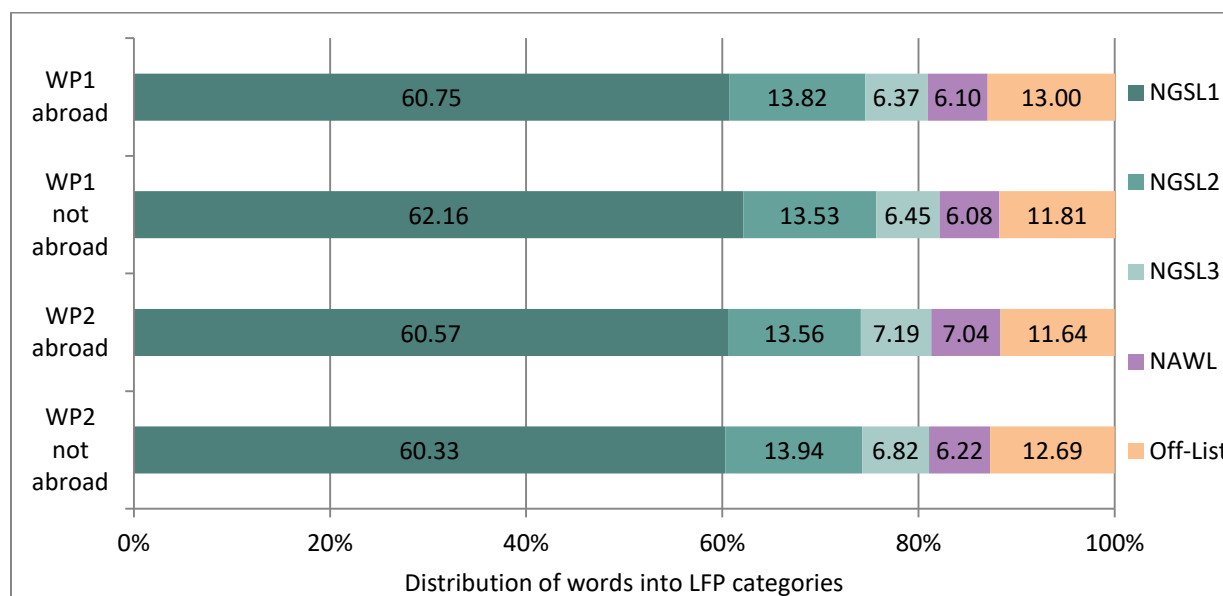
Figure 12. Distribution of words into different LFP categories based on a potential stay abroad

Not only differences regarding their LFP scores could be found but also differences in the development of the participants' scores: the development of lexical density as well as the use of words from both the NGSL3 and NAWL of the group that had been abroad was more prominent. In contrast, the group that had not spent time abroad had a higher development of the TTR and used less words from the NGSL1 and more from the NGSL2. However, despite these findings being interesting, they are only minor differences since no statistical significance between the group that had spent time abroad and the one that had not could be detected.

Extramural English

Regarding the use of Extramural English not only questions about the time devoted to EE activities (talking in English, watching films or videos in English, playing video games and reading English texts) prior to writing the BA or BEd thesis were included but also a potential change in this behaviour during the time between writing the PS1 Linguistics paper and the BA or BEd thesis. The participants could select seven answers based on the time spent on the activity with the lowest option on the scale being *less than 1 hour* and the highest being *more than 30 hours a week*. Subsequently, a score (henceforth EE score) was calculated

based on the overall hours spent on all activities involving Extramural English in order to investigate whether a correlation between exposure to Extramural English and lexical proficiency could be found. Significant correlations between this score and lexical density 1 as well as 1NGSL2 scores could be detected. Surprisingly, participants with a higher EE score had not only a significantly lower lexical density 1 (p=.009) but also used fewer words from the NGSL2 (p=.023) in WP1 than those with a lower EE score. In the following figure it can be observed that with an increasing EE score lexical density seemed to decrease.
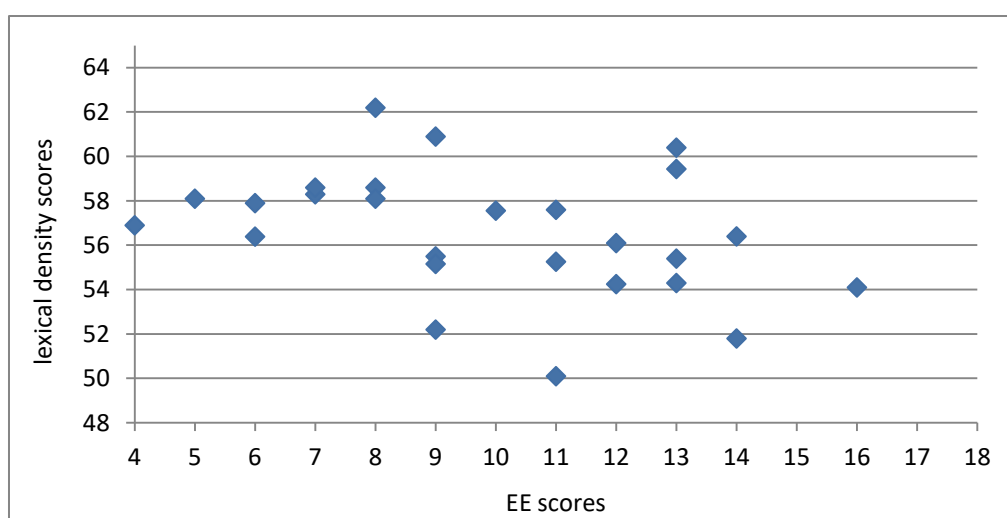
Figure 13. Comparison of lexical density 1 scores and EE scores

However, the figure also illustrates that the results are relatively heterogeneous. For instance, participants with an EE score of 9 achieved scores between 52.2 and 60.9 for lexical density 1, participants with an EE score of 11 between 50.1 and 57.6.

In order to facilitate the comparison of EE scores with all other scores, the participants were divided into two groups: as the median of the EE score was 10.50, group 1 consisted of those participants that achieved an EE score lower than the median and group 2 of those with a higher score. However, apart from group 2 scoring lower in lexical density 1 (p=.014) and 2 (p=.058) as well as 1NGSL2 (p=.068), the scores of both groups were relatively similar.

Another interesting aspect regarding the influence of EE activities on lexical proficiency is analysing potential correlations between individual EE scores (e.g. for talking or reading in English) and different lexical proficiency scores. Surprisingly, participants that stated spending more time talking in English had a significantly lower lexical density 1 but also a significantly higher development of their lexical density scores (p=.013 and .016

respectively). Furthermore, with increasing hours spent talking in English, the difference between 1NGSL3 and 2NGSL3 was also significantly higher (p=.042), meaning that participants that stated having spent more time talking in English underwent a higher development regarding the use of words from the NGSL3.

Another aspect of Extramural English for which significant results could be found is watching films or videos in English. Interestingly, the participants with the highest amount of time spent watching audiovisual material in English had a significantly lower lexical density 1 (p=.035), used considerably fewer words from the NGSL2 in WP1 (p=.045) but also had a significantly higher development of their TTR scores (p=.036).

Concerning video games, it can be said that those participants that had played more than five hours per week had not only a considerably higher TTR1 (p=.036) and TTR2 (p=.027) but also used significantly fewer words from the NGSL1 in WP2 (p=.032). However, it also has to be taken into consideration that 24 participants (85.71%) stated that they had spent less than five hours per week playing video games and merely four participants (14.29%) had played between five and twenty hours per week.

Lastly, the weakest effects could be found for the activity of reading English texts. Overall, all participants achieved relatively similar scores regardless of how much time they had spent reading English texts. The only finding was that those who had spent more time reading English texts used significantly fewer words from the NGSL2 (p=.012) in the first written production.

Vocabulary studied actively

Another factor that might potentially affect lexical proficiency is whether the participants had actively studied vocabulary with the purpose of improving their lexical proficiency, particularly with regard to their academic writing. More than half of the participants (53.57%) had studied vocabulary during their time at university with the aim of improving their performance in written productions. Most of them had collected unknown words they had encountered whilst watching films or reading in order to study them afterwards. Interestingly, merely two participants stated that they had compiled lists of phrases and words they thought might be convenient for academic writing.

Not surprisingly, the participants that stated that they had actively studied vocabulary achieved higher scores for numerous aspects of lexical proficiency in both written productions. As graph 14 below demonstrates, the participants that stated they had actively studied vocabulary did not only use a significantly lower percentage of words from the 1NGSL1 (p=.001) but also significantly more words from the 1NAWL (p=.004).
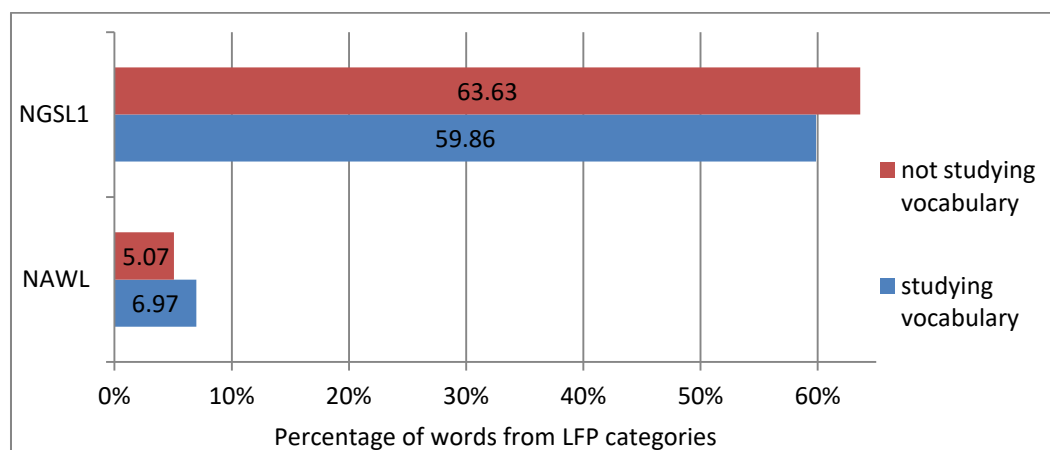
Figure 14. Percentage of words used from NGSL1 and NAWL in WP1 based on whether participants actively studied vocabulary

Moreover, they also achieved higher scores for lexical density and lexical variation in both written productions with the p-values for lexical density 2 (p=.029) and TTR 1 (p=.012) even being statistically significant. However, it also ought to be mentioned that they had a less noticeable development of TTR (p=.047), NGSL1, NGSL2 and NAWL than those participants that had not actively studied vocabulary.

Gender

Regarding the gender of the participants, five of them were male and 23 female. When analysing the results of both groups, it can be observed that the lexical density of females and males was almost identical in both written productions with the males having undergone a slightly higher development. However, the TTR1 of the males was higher (0.3860) than that of the females (0.3596). Despite the scores of both groups having increased in WP2, the male participants still outperformed the female ones with a TTR of 0.3900 and 0.3652 respectively. Although the difference in TTR2 scores not being significant, the p-value of .085 shows that there was a tendency of males having outperformed females.

Concerning lexical sophistication (see figure 15 below), the percentage of words belonging to the NGSL1 is particularly interesting: whilst in WP1 males used slightly more words from the NGSL1 than females (62.08% and 61.51% respectively), in WP2 females used significantly (p=.036) more words from the NGSL1, namely 61.00% compared to 57.79% used by males. Hence, the development between the two groups differed considerably, with males having used 4.29% fewer words from the NGSL1 in the second written production compared to merely 0.51% for the female participants.
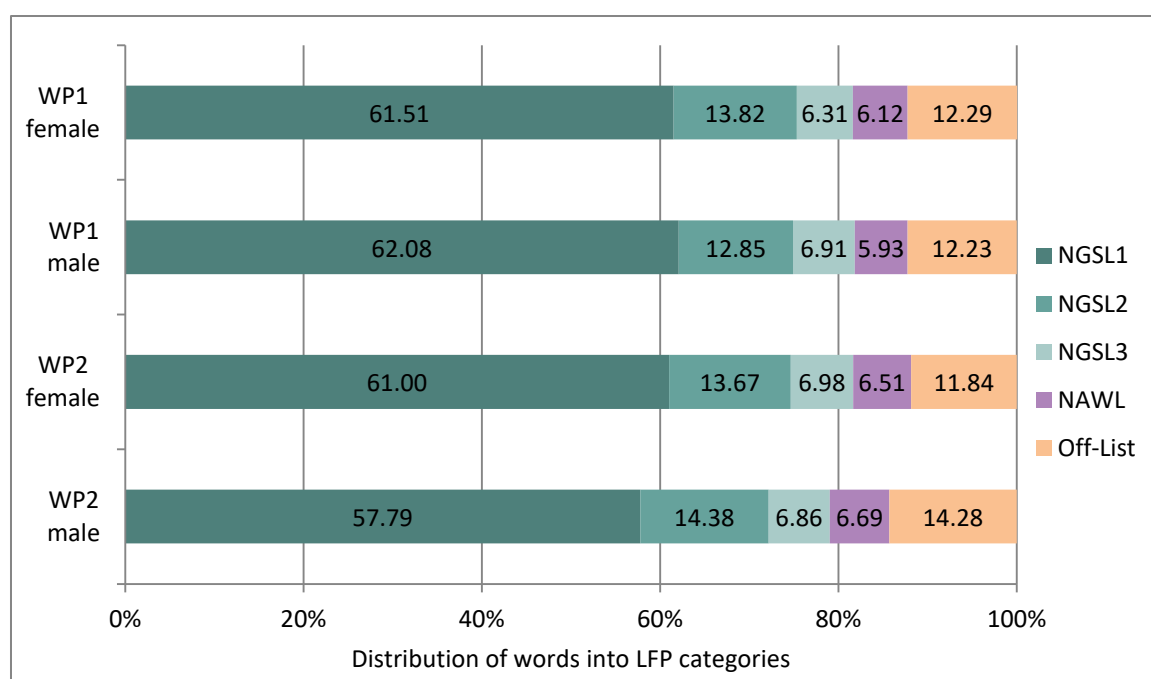
Figure 15. Distribution of words into different LFP categories based on the participants' gender

Regarding the remaining categories for lexical sophistication, in WP1 females used a higher percentage of words ascribed to the NGSL2 and NAWL than males. Surprisingly, this changed in WP2, as the male participants used more words from the NGSL2 and NAWL.

Age

Out of the 28 participants, 20 stated that they were between 20 and 23 years old when submitting the BA or BEd thesis in linguistics, whilst seven were between 24 and 27 and only one participant was between 28 and 30. Hence, the participants were divided into two groups: those that were between 20 and 23 years old when submitting the BA or BEd thesis and those that were between 24 and 30. Interestingly, the older group used significantly (p=.034) more academic vocabulary (7.51%) in their second written production when compared to the younger group (6.16%), as can be seen in figure 16.
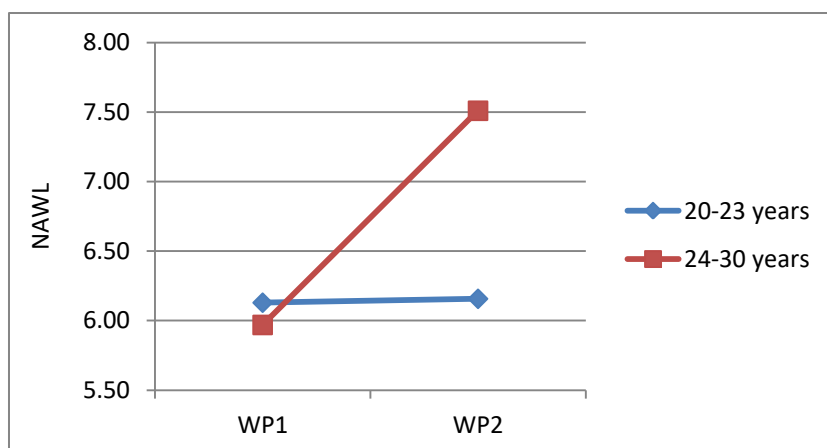
Figure 16. Percentage of words from the NAWL in percent in WP1 and WP2 based on participants' age

Furthermore, despite having scored slightly lower in WP1, a tendency (p=.07) of the older group having exhibited a considerably higher development of the use of academic vocabulary can be observed.

It is also noteworthy to mention that whilst the older group slightly outperformed the younger group regarding TTR1, TTR2 and lexical density 1, the younger group had higher scores for lexical density 2. Moreover, in WP1 the younger group used fewer words from the NGSL1 but more from the NGSL3 and NAWL. In WP2, however, the older group used not only a lower percentage of words assigned to the NGSL1 but also more words from the NGSL2, NGSL3 and NAWL. Hence, it is not surprising that the younger group also had a higher development of TTR and lexical density, whilst the older participants experienced a higher development regarding all aspects of lexical sophistication.

6. Discussion

In this section the results presented above will be discussed in more detail with regard to previous research on the topic and an attempt to explain the findings (or the lack thereof) will be made. Subsequently, limitations of the present study will be mentioned as well as aspects that might be interesting for further research.

6.1 Different aspects of lexical proficiency

Research question one focuses on which aspects of lexical proficiency improve most. As the previous chapter has shown, participants showed improvements in lexical density, lexical variation and lexical sophistication in WP2. However, no significant differences between the scores for the two written productions could be found. Hence, despite there being an

improvement in all aspects of lexical proficiency, the differences in scores were relatively minor. Due to the rather heterogeneous group of participants, the lack of significant results might be attributable to several factors.

One of these is that, as already mentioned, the time span between WP1 and WP2 differed tremendously with one to eight semesters. As it can be assumed that a longer exposure to the English language in the course of their studies had an effect on participants' language proficiency, potential correlations between the time span and different aspects of lexical proficiency were investigated. And indeed, those participants with a longer time span between the two written productions also achieved significantly higher scores for TTR2 (p=.003), used fewer words from the NGSL1 (p=.022) and more words from the NAWL (p=.029)[6] in WP2. Hence, it might be assumed that with an increasing amount of time between the written productions the lexical proficiency improves simultaneously. However, it also ought to be mentioned that those participants with a higher number of semesters between WP1 and WP2 already performed better regarding TTR1, NGSL1 and NAWL in WP1 (with the latter two even showing significant differences). Concerning lexical density, those participants with more time between WP1 and WP2 also scored higher. However, the results were less substantial with p=.086 and p=.222 for lexical density 1 and 2 respectively. These findings illustrate that the amount of time between two written productions might indeed have had an impact on the extent of development of participants' lexical proficiency although the initial level of proficiency might also have been of importance, which will be discussed in more detail in section 6.2.

Furthermore, numerous other factors ought to be considered as the participants of the present study were relatively heterogeneous regarding their age and linguistic background (which will be discussed at length in section 6.3.) but also the research area which their academic papers can be ascribed to. Although there is a lack of research on the percentage of technical words used in different research areas of linguistics, studies have shown that when compared to other fields, such as anatomy, applied linguistics contains considerably fewer technical words (Chung & Nation 2003: 108). Hence, it might very well be possible for different areas of linguistics to consist of a varying amount of technical vocabulary, i.e. that

---

[6] Since the time span between the two written productions was not normally distributed, Spearman Rank Order Correlation was used to calculate a potential significant difference.

papers on a topic of, for instance, cognitive or historical linguistics, require more technical vocabulary than applied linguistics. In addition, words that can be considered technical vocabulary for the discipline of applied linguistics might not be recognised as such due to them also being found in general use, such as *instruct* or *interaction* (Chung & Nation 2003: 111).

Hence, to answer research question one, it can be said that despite all aspects of lexical proficiency that were analysed (i.e. lexical density, lexical variation and lexical sophistication) improving in the period from the first to the second written production, the difference in scores was merely a subtle one. Moreover, the development is relatively similar concerning all aspects of lexical proficiency, and hence, it cannot be stated that a particular aspect was improved more than another. However, the lack of significant differences might be due to other factors, which will be discussed in more detail in the following sections.

6.2 Initial level of proficiency

Research question two is concerned with whether the development of lexical proficiency is dependent on its initial level. As shown in chapter 5.2, a correlation between the initial level and lexical proficiency could indeed be detected: students with lower scores in WP1 had a more noticeable development of lexical proficiency than those with higher scores in WP1. Similar findings have been described by numerous researchers, who have reported that participants with a lower initial level of proficiency experienced a higher development than those with a higher initial level (e.g. Llanes & Muñoz 2009: 361; Barquin 2012: 215-216; Pérez-Vidal 2014: 26; Aguilar & Muñoz 2014: 12). However, it ought to be mentioned that with the exception of Aguilar and Muñoz (who investigated the effect of the initial level of proficiency in a CLIL context), most studies focused on whether participants with a lower initial level benefited more from studying abroad than those with a higher initial level. Nevertheless, the results from the present study as well as previous research illustrate that the initial level indeed seems to have an effect on the extent to which lexical proficiency improves.

Interestingly, it was also discovered that the higher the initial proficiency, the higher the scores measured in WP2. For both lexical density and TTR, the participants with a higher initial level outperformed the ones with a lower initial level, despite the latter experiencing a

more noticeable development. Only for the percentage of words used from the NAWL did the lower scoring group achieve higher scores in WP2 than the higher scoring one. These findings seem to confirm those of previous research which report that despite a more noticeable development of proficiency experienced by participants with a lower proficiency, they did not succeed in outperforming the participants with a higher initial level in WP2 (Aguilar & Muñoz 2014: 9; Barquin 2012: 168).

Hence, to answer research question two it could indeed be observed that the initial proficiency level seems to be pivotal with regard to not only the extent of the development of participants' proficiency but also their proficiency level demonstrated in WP2.

6.3 Linguistic background

Research question three focuses on whether the linguistic background of the participants has an influence on their lexical proficiency. Hence, in the following the results will be discussed with a focus on potential correlations between different factors.

Bilingual upbringing

As the previous chapter has shown there were hardly any noticeable differences between monolingually and bilingually raised participants. Although this might be surprising due to some researchers reporting that bilinguals outperformed monolinguals with regard to executive control, which is strongly related to vocabulary acquisition (Carlson & Meltzoff 2008: 282; Bialystok & Viswanathan 2009: 494), this is not necessarily the case with academic vocabulary. Since both monolingually and bilingually raised students are primarily exposed to formal vocabulary in school rather than at home, the number of languages they grow up speaking seems to have relatively little effect on their lexical proficiency concerning low-frequency or academic vocabulary. This observation is also affirmed by studies showing a relatively similar vocabulary size of monolingually and bilingually raised people (Bialystok et al. 2010: 530). Moreover, the low number of participants that had grown up bilingually (10.71%; N=3) has to be taken into consideration as well.

In contrast to this lack of a significant difference based on the number of languages the participants had learned whilst growing up, the overall number of languages known as well as the depth of knowledge indeed seemed to have an effect on their lexical proficiency. In order to investigate potential correlations, the participants' self-assessment regarding their

knowledge of other languages, particularly their presumed CEFR (Common European Framework of Reference for Languages) level, was used to ascribe a language score (1 point for A1, 2 points for A2, 3 points for B1, etc.) for each language apart from their L1 to the students and subsequently adding them up. Although no correlations between this score and lexical density or lexical variation were found, participants with a higher score did indeed use considerably fewer words from the NGSL1 in WP1 (p=.052) and significantly more academic vocabulary in WP2 (p=.048). As numerous researchers have reported that bilinguals might benefit from positive transfer of their first languages when acquiring an additional language (Cenoz 2013: 72; Baker 2001: 271; Bartolotti & Marian 2017: 131), it stands to reason to assume that monolingually raised students might also benefit from other foreign languages they have learned before. However, it might be of importance which languages the students have acquired as Romance languages, for instance, bear more similarities to the English language, particularly regarding academic vocabulary, than other languages. Concerning the present study the majority of the participants (71.43 %) knew at least one Romance language, namely French, Spanish, Italian and Romanian, but they also stated knowing other languages, such as Austrian Sign Language, Russian, Japanese, Swedish, Albanian, Modern Greek, Estonian and Dutch.

Bilingual education

Concerning a potential effect of bilingual education on lexical proficiency, only one significant difference in the scores of the participants that had received bilingual education and those that had not could be detected, namely that the latter performed significantly better regarding lexical density 1. This might seem surprising as an increased exposure to English might be assumed to be beneficial for lexical proficiency. However, there are multiple potential explanations for this lack of an advantage of the students that had experienced bilingual education: firstly, merely 17.86% of the participants received bilingual education, which automatically renders a comparison between the groups problematic. Secondly, their experience differed considerably, from less than 25% up to more than 75% of courses being held in the foreign language. Similarly, the extent to which the CLIL lessons required the students to use academic and low-frequency vocabulary productively, particularly in written productions, might have varied substantially. Thirdly, as the detailed discussion of the research literature in chapter 3.2 has demonstrated, even in the case of all

participants of a study being part of the same bilingual programme, there is hardly any consensus amongst researchers as to whether an impact on lexical proficiency can be detected. In fact, Olsson and Sylvén have described findings similar to those of the present study as they have not found an advantage of CLIL students concerning the use of academic vocabulary (Olsson & Sylvén 2017: 14). However, as mentioned above, it ought to be kept in mind that the students might not have been required to produce written compositions containing academic and low-frequency vocabulary, which might be a potential explanation for these findings.

Other researchers have reported that a positive impact of CLIL on academic vocabulary acquisition was strongly connected to the use of students' L1 for instruction in addition to English (Olsson 2021: 17). By partly using the L1, particularly when introducing more challenging topics or concepts, the students were assisted in comprehending not only the subject-specific topic but also the academic vocabulary in both English and their L1 (Olsson 2021: 19). Hence, given all of the above-mentioned potentially influencing factors, the lack of a higher performance of the group that had received bilingual education is hardly surprising although more data would be necessary to generalise these findings.

Time spent abroad

As already discussed, increased exposure to a foreign language by spending time abroad is likely to have a positive effect on language proficiency. Surprisingly, no significant differences between participants that had spent time abroad and those that had not could be detected. However, stays abroad can vary tremendously, from six-months university exchange programmes and work-and-travel programmes to visiting family abroad for a shorter period of time. Hence, the length and type of stay is crucial to consider when dividing participants into two groups according to whether or not they have spent time in a foreign country. As mentioned in chapter 3.3, a semester at a university abroad might be more beneficial for making progress concerning academic vocabulary and writing proficiency than a work-and-travel programme. However, in the present study no significant correlations between these aspects and lexical proficiency could be found despite the participants' stays differing in terms of length (from 3 months to multiple years) and type (e.g. exchange semester at a university or work-and-travel programme). Interestingly, participants that stated regularly having spent time with family abroad used significantly less

academic vocabulary in WP1 but significantly more in WP2. However, with merely eleven out of 28 participants having spent time abroad (seven for studying, one for working and three for both), the small sample size renders a comparison problematic.

In addition to the length and type of the stay, the number of visits abroad as well as the motivation to interact in the foreign language (Isabelli-García 2006: 233) (which automatically affects the amount of communication in the foreign language) are important aspects to take into consideration. Although merely one participant had been abroad twice and two participants three times, a comparison of their scores can be insightful: not only did those students that had been abroad more than once score significantly higher in TTR1 (p=.022) but they also had a significantly more noticeable development towards the use of a lower percentage of words from the NGSL1 (p=.047). Moreover, there were tendencies of them using fewer words ascribed to the NGSL1 (p=.087) and a higher percentage of words from the NGSL3 (p=.065) in WP2. Hence, these findings confirm those of previous studies which report that participants that had stayed abroad longer had more noticeable improvements of their lexical proficiency (Milton and Meara 1995, quoted in Zaytseva, Pérez-Vidal & Miralpeix 2018: 213). It also seems essential to note whether the participants had been abroad before or after WP1 (seven and four participants respectively) and what effect this might have. Not unexpectedly, those abroad before WP1 had a higher lexical density 1 and TTR1 but a less noticeable development of these scores, which are even significant in the case of TTR scores (p=.027). Regarding the amount of communication in the official language of the country the participants had stayed in, those that selected "75-100%" used significantly fewer words from the NGSL1 in WP1 than the group that chose "50-75%" as their answer. Moreover, although those participants that had been exposed to the foreign language in 50-75% of the time scored lower regarding 1NAWL, they had a significantly more noticeable development (p=.047).

This discussion has shown that not only the length and type of stay but also several other factors are essential and ought to be considered when investigating the effect of stays abroad. However, despite these results being intriguing, it also has to be mentioned that the number of participants would have to be considerably higher in order to allow for a generalisation of these findings.

Extramural English

As presented in chapter 5.3, numerous significant effects of Extramural English on the participants' lexical proficiency could be detected. However, despite some of the lexical proficiency scores showing positive correlations with time spent on EE activities, other scores were higher when participants engaged less in EE activities. Although this might be surprising as such activities automatically lead to a higher exposure to English, Olsson and Sylvén have reported a lack of impact of EE on academic vocabulary, particularly for students with a higher level of proficiency (Olsson & Sylvén 2015: 93-94). Moreover, in the present study another factor that has to be taken into consideration is that the participants were asked to conduct a self-assessment of the amount of time spent on EE activities, which might also be a confounding factor. Furthermore, the type of EE activity and material should not be neglected, as playing a video game might possibly involve a higher percentage of high-frequency vocabulary and thus have less of an impact on certain aspects of lexical proficiency than watching documentaries or reading articles containing more low-frequency or technical vocabulary. In addition, correlations between individual EE scores could be detected, as those students that stated spending more time reading English texts not only watched significantly more videos or films (p=.004) but also spent considerably more time talking in English (p=.011). Hence, the observed effects might not be attributable to individual EE activities but potentially to the overall amount of time invested in EE activities.

Vocabulary studied actively

As mentioned in chapter 5.3, those participants that stated having actively studied vocabulary with the purpose of improving their academic writing achieved higher scores regarding multiple aspects of lexical proficiency. However, this is hardly surprising as it stands to reason that studying vocabulary can have a positive effect on lexical proficiency. Nevertheless, it can be insightful to investigate potential correlations between studying vocabulary and other factors. For instance, those who had studied vocabulary had a longer time span between writing the PS1 paper and their BA or BEd thesis (p=.008), with a mean of 3.87 semesters compared to merely 2.15 semesters for those who did not actively study vocabulary. The higher scores of the former group might hence partly be explained by the potentially higher exposure to English due to the longer time span of studying English.

Moreover, those who stated having actively studied vocabulary were more likely to have spent time in a foreign country working or studying (p=.015) as 60.00% (N=9) were abroad compared to merely 15.39% (N=2) of those who did not actively study vocabulary. Particularly those students that had spent time in a foreign country in order to study there (N=10) stated almost unanimously (90.00%) that they had studied vocabulary with the purpose of improving their lexical proficiency concerning written productions, for instance by making lists of unknown words. In contrast, merely a third (33.33%) of the participants that had not spent time abroad before their BA or BEd thesis (N=17) or had spent time abroad for working purposes (N=1) stated having studied vocabulary in order to improve their lexical proficiency.

Lastly, correlations between actively studying vocabulary and EE activities were also found, as participants that had spent a considerable amount of time watching films or videos in English were less likely to study vocabulary in order to improve their lexical proficiency in written productions (p=.022). This is particularly interesting with regard to the finding that those students that had spent more time watching films in English scored significantly lower in lexical density 1 and 1NGSL2. Hence, despite studying vocabulary likely contributing to an increase of lexical proficiency, in the present study strong correlations with other factors could be detected. Potential interrelations between different factors should thus be taken into consideration.

In conclusion, it could be shown that various aspects of students' linguistic background ought to be considered when investigating their lexical proficiency. Whilst for bilingual upbringing and education only few positive effects on lexical proficiency could be found, spending time abroad, engaging in EE activities and actively studying vocabulary can indeed have an impact on certain aspects of lexical proficiency. Hence, the answer to research question three is that correlations between the students' linguistic background and their scores could be found to some extent. However, at this point the relatively low number and uneven distribution of participants into different groups according to their linguistic background ought to be noted again, as this might also be an explanation for the sometimes surprising findings. Moreover, most aspects of a person's linguistic background are interrelated and it might not be possible to analyse them as being entirely separate and, in further consequence, to ascribe differences in lexical proficiency to specific factors.

6.4 Gender and age

Research question four focuses on whether the participants' gender and age impacts their lexical proficiency. Concerning a potential effect of gender, chapter 5.3 has shown that despite males generally scoring higher in almost all aspects of lexical proficiency, hardly any statistically significant differences between females and males could be found. Similar results of males outperforming females in various aspects of lexical proficiency were reported by multiple researchers, such as Pahom, Farley and Ramonda (2015) or Olsson and Sylvén (2015). Nevertheless, as mentioned in chapter 3.5 there is little consensus between researchers as to whether there is an effect of gender on language proficiency and literature focusing on a potential impact on academic vocabulary is particularly scarce. However, the higher scores of male participants in the present study could be explained when inspecting other factors: a significantly (p=.041) higher number of males stated having spent time abroad with family frequently (80.00%, with the majority of them using English in this context) compared to merely 30.43% of female participants. Thus, they had had a higher exposure to English, possibly leading to an increased lexical proficiency. Moreover, male participants had spent a significantly higher amount of time playing video games in English (p<.001). Despite a comparison being difficult due to the rather uneven distribution of participants into groups according to their gender, other researchers have also reported a positive effect of video games on vocabulary acquisition (Ebrahimzadeh 2017: 10; Peterson 2010: 436). Hence, the increased exposure to English due to these factors might indeed have had a positive impact on the male participants' lexical proficiency and explain them outperforming the females.

Concerning the effect of age on lexical proficiency, older participants indeed outperformed younger ones in several aspects of lexical proficiency, although only the difference in the use of academic words in WP2 was statistically significant (p=.401). Although these findings might partly be explained by the fact that older participants might have had slightly more exposure to English, correlations between age and other factors could be detected: a significantly higher percentage of the older participants had grown up in countries other than Austria (p=.004), which might in turn explain the fact that they were also more likely to have spent time abroad (p=.038). Interestingly, despite the age at which the participants had started learning English varying considerably (between being five and eleven years old), no

effect on lexical proficiency could be detected. However, it ought to be kept in mind that while having started to learn English at an earlier age might have given the students an advantage at a lower proficiency level, it might not have a noticeable effect on their lexical proficiency when they reach a higher proficiency level. In the first years of acquiring a new language, students are primarily exposed to high-frequency vocabulary. Moreover, most learners are introduced to low-frequency and academic vocabulary at approximately the same time regardless of the age at which they have started learning English. Hence, their lexical proficiency regarding academic or low-frequency vocabulary might remain relatively unaffected by the number of years they have already been learning English.

To answer research question four, it can be stated that both gender and age did indeed have an effect on some aspects of lexical proficiency. However, as mentioned above, other factors (such as spending time abroad or engaging in EE activities) might have influenced the results and hence, ascribing the difference in scores solely to the participants' gender or age is not possible.

6.5 Limitations and further research

This study has a number of limitations, with the first one being a considerable difference in the topics of the written productions: although both the PS 1 papers and the BA and BEd theses can be considered academic texts and have relatively similar writing tasks due to not only the standardised course designs and requirements but also all of them being written on topics of linguistics, the research areas and topics of the papers varied tremendously from applied linguistics to cognitive linguistics and historical linguistics. Therefore, differences in the lexical proficiency scores might possibly also be due to differences in the papers' topics.

Moreover, the call for contribution was mostly distributed via social media and the sample therefore generated on a voluntary basis. Since analysing written productions by all students of a certain cohort was not possible due to not only the scope of this study but also bureaucratic reasons, it has to be kept in mind that only students who participated voluntarily constituted the sample. Although there are no indications for this assumption, it cannot be ruled out that students with a higher language proficiency might have been more willing to participate and have their written compositions analysed. Therefore, it would be

highly interesting for future research to analyse the texts produced by an entire cohort of students to gain more insight into the topic.

Lastly, as already mentioned in the discussion of the results, it was not possible to get groups of equal size for all variables, such as monolingual and bilingual upbringing or female and male. This ought to be taken into consideration, as it renders a generalisation of the findings impossible. Hence, it might be extremely insightful to analyse written productions of a considerably larger sample in order to be able to form groups of equal size.

7. Conclusion

To conclude, this study aimed to shed light on the development of different aspects of lexical proficiency and the potential effect that the participants' initial proficiency level, linguistic background, age or gender might have on their lexical proficiency.

Although slight improvements in the participants' performance between the two written productions could be observed for all aspects of lexical proficiency, none of them were statistically significant. Moreover, the development of scores was relatively similar for all aspects of lexical proficiency. Hence, it cannot be confirmed that certain aspects improved more discernibly than others.

In contrast, the initial level of the participants' lexical proficiency was indeed found to be pivotal for the further development of their lexical proficiency as students with a higher performance in the first written production achieved higher scores in the second written production as well. However, it also has to be added that the development of lexical proficiency of the students with a lower initial level was more discernible.

Regarding the potential effect of the participants' linguistic background on their lexical proficiency, numerous significant correlations could be found. Whilst bilingual upbringing and education merely had slight positive effects, those students that spent time abroad, engaged in EE activities and actively studied vocabulary indeed achieved significantly higher scores.

Lastly, despite males performing slightly better than females in most aspects, no significant difference based on the gender of the participants could be found. Moreover, the higher scores of males in this study might also be explained by their higher exposure to English due

to spending significantly more time abroad and playing video games in English. Regarding the effect of age on lexical proficiency a better performance of older participants was detected. However, the differences were not significant for most of the aspects and, similar to the effect of gender, might be explained by correlations with more time spent abroad.

Hence, this study has shown that certain factors, such as spending time abroad or engaging in EE activities, can indeed contribute to a higher lexical proficiency. However, it ought to be taken into consideration that numerous significant correlations between these factors could be found. As this indicates strong interrelations, ascribing positive effects on lexical proficiency to individual factors is problematic, particularly regarding the sample size. Hence, conducting a similar study on a larger scale or with an entire cohort of students would be highly insightful as it might facilitate an ascription of potential differences in scores to specific factors.

8. Bibliography

Abello-Contesse, Christian; Chandler, Paul M.; López-Jiménez, María Dolores; Chacón-Beltrán, Rubén (eds.). 2013. *Bilingual and multilingual education in the 21st century: building on experience*. Bristol: Multilingual Matters.

Abello-Contesse, Christian. 2013. "Bilingual and multilingual education: an overview of the field". In Abello-Contesse, Christian; Chandler, Paul M.; López-Jiménez, Mariá Dolores; Chacón-Beltrán, Rubén (eds.). *Bilingual and multilingual education in the 21st century: building on experience*. Bristol: Multilingual Matters, 3-23.

Aguilar, Marta; Muñoz, Carmen. 2014. "The effect of proficiency on CLIL benefits in engineering students in Spain". *International Journal of Applied Linguistics* 24(1), 1-18.

Agustín Llach, María P.; Gallego, Melania Terrazas. 2012. "Vocabulary knowledge development and gender differences in a second language". *ELIA* 12, 45-75.

Agustín-Llach, María P. 2016. "Age and type of instruction (CLIC vs. traditional EFL) in lexical development". *International Journal of English Studies* 16(1), 75-96.

Agustín Llach, María P. 2017. "The effects of the CLIL approach in young foreign language learners' lexical profiles". *International Journal of Bilingual Education and Bilingualism* 20(5), 557-573.

Analyze My Writing. n.d. www.analyzemywriting.com (5 Sept. 2021).

Anthony, Laurence. 2021. *AntWordProfiler.* (Version 1.5.1). [Computer Program]. Tokyo: Waseda University. https://www.laurenceanthony.net/software (5 Sept. 2021).

Ardasheva, Yuliya; Tretter, Thomas R.; Kinny, Marti. 2012. "English language learners and academic achievement: revisiting the threshold hypothesis". *Language Learning* 62(3), 769-812.

Arribas, Mario. 2016. "Analysing a whole CLIL school: students' attitudes, motivation, and receptive vocabulary outcomes". *Latin American Journal of Content & Language Integrated Learning* 9(2), 267-292.

Baddeley, Alan; Gathercole, Susan; Papagno, Costanza. 1998. "The phonological loop as a language learning device". *Psychological Review* 105(1), 158-173.

Baker, Colin. 2001. *Foundations of bilingual education and bilingualism*. (3rd edition). Clevedon: Multilingual Matters.

Bartolotti, James; Marian, Viorica. 2017. "Bilinguals' existing languages benefit vocabulary learning in a third language: bilingual third language learning". *Language Learning* 67(1), 110-140.

Baumann, James F.; Graves, Michael F. 2010. "What is academic vocabulary?". *Journal of Adolescent & Adult Literacy* 54(1), 4-12.

Barquin, Elisa. 2012. "Writing development in a study abroad context". Doctoral dissertation, University Pompeu Fabra.

Beardsmore, Hugo Baetens. 1986. *Bilingualism: basic principles*. (2nd edition). Clevedon: Multilingual Matters.

Bialystok, Ellen. 2009. "Bilingualism: the good, the bad, and the indifferent". *Bilingualism: Language and Cognition* 12(1), 3-11.

Bialystok, Ellen. 2015. "Bilingualism and the development of executive function: the role of attention". *Child Development Perspectives* 9(2), 117-121.

Bialystok, Ellen; Craik, Fergus I. M.; Klein, Raymond; Viswanathan, Mythili. 2004. "Bilingualism, aging, and cognitive control: evidence from the Simon task". *Psychology and Aging* 19(2), 290-303.

Bialystok, Ellen; Martin, Michelle M. 2004. "Attention and inhibition in bilingual children: evidence from the dimensional change card sort task". *Developmental Science* 7(3), 325-339.

Bialystok, Ellen; Craik, Fergus I.M.; Green, David W.; Gollan, Tamar H. 2009. "Bilingual minds". *Psychological Science in the Public Interest* 10(3), 89-129.

Bialystok, Ellen; Viswanathan, Mythili. 2009. "Components of executive control with advantages for bilingual children in two cultures". *Cognition* 112(3), 494-500.

Bialystok, Ellen; Luk, Gigi; Peets, Kathleen F.; Yang, Sujin. 2010. "Receptive vocabulary differences in monolingual and bilingual children". *Bilingualism: Language and Cognition* 13(4), 525-531.

Bloomfield, Leonard. 1935. *Language*. London: Allen & Unwin.

Brecht, Richard D.; Davidson, Dan E.; Ginsberg, Ralph B. 1995. "Predictors of foreign language gain during study abroad". In Freed, Barbara (ed.). *Second language acquisition in a study abroad context*. Amsterdam: John Benjamins, 37-66.

Brezina, Vaclav; Gablasova, Dana. 2015. "Is there a core vocabulary? Introducing the New General Service List". *Applied Linguistics* 36(1), 1-22.

Briggs, Jessica. 2015. "Out-of-class language contact and vocabulary gain in a study abroad context". *System* 53, 129-140.

Bruton, Anthony. 2011. "Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal and Moore (2010)". *Applied Linguistics* 32(2), 236-241.

Browne, Charles; Culligan, Brent; Phillips, Joseph. 2013. *New Academic Word List*. http://www.newgeneralservicelist.org/nawl-new-academic-word-list/ (8 Jan. 2022).

Browne, Charles. 2014. "A New General Service List: the better mousetrap we've been looking for?". *Vocabulary Learning and Instruction* 3(1), 1-10.

Browne, Charles. 2021. "The NGSL Project: building wordlists and resources to help EFL learners (and teachers) to succeed". Paper presented at *JALTCALL 2020 Conference*, Tokyo, 1-18.

Bulté, Bram; Housen, Alex; Pierrard, Michel; can Daele, Siska. 2008. "Investigating lexical proficiency development over time: the case of Dutch-speaking learners of French in Brussels". *French Language Studies* 18, 277-298.

Campion, Mary; Elley, Warwick. 1971. *An Academic Vocabulary List*. Wellington: New Zealand Council for Educational Research.

Carlson, Stephanie M.; Meltzoff, Andrew N. 2008. "Bilingual experience and executive functioning in young children". *Developmental Science* 11(2), 282-298.

Cenoz, Jasone. 2003. "The additive effect of bilingualism on third language acquisition: a review". *The International Journal of Bilingualism* 7(1), 71-87.

Cenoz, Jasone. 2013. "The influence of bilingualism on third language acquisition: focus on multilingualism". *Language Teaching* 46, 71-86.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. (2nd print). Cambridge, Mass.: MIT Press.

Chung, Teresa Mihwa; Nation, Paul. 2003. "Technical vocabulary in specialised texts". *Reading in a Foreign Language* 15(2), 103-116.

Cobb, Tom. n.d. *Compleat Web VP*. https://www.lextutor.ca/vp/comp/ (5 Sept. 2021).

Cowan, J. Ronayne. 1974. "Lexical and syntactic research for the design of reading materials". *TESOL Quarterly* 8(4), 389-399.

Coxhead, Averil. 2000. "A new Academic Word List". *TESOL Quarterly* 34(2), 213-238.

Coxhead, Averil. 2020. "Academic vocabulary". In Webb, Stuart (ed.). *The Routledge handbook of vocabulary studies*. London: Routledge, 97-110.

Crossley, Scott A.; Cobb, Tom; McNamara, Danielle S. 2013. "Comparing count-based and band-based indices of word frequency: implications for active vocabulary research and pedagogical applications". *System* 41, 965-981.

Cummins, Jim. 1976. "The influence of bilingualism on cognitive growth: a synthesis of research findings and explanatory hypotheses". *Working Papers on Bilingualism* 9, 1-43.

Daller, Helmut; van Hout, Roeland; Treffers-Daller Jeanine. 2003. "Lexical richness in the spontaneous speech of bilinguals". *Applied Linguistics* 24(2), 197-222.

Dalton-Puffer, Christiane. 2007. *Discourse in content and language integrated learning (CLIL) classrooms*. Amsterdam: John Benjamins.

Dalton-Puffer, Christiane. 2008. "Outcomes and processes in content and language integrated learning (CLIL): current research from Europe". In Delanoy, Werner; Volkmann, Laurenz (eds.). *Future perspectives for English language teaching*. Heidelberg: Carl Winter, 139-157.

Dalton-Puffer, Christiane; Nikula, Tarja; Smit, Ute. 2010. "Charting policies, premises and research on content and language integrated learning". In Dalton-Puffer, Christiane; Nikula, Tarja; Smit, Ute (eds.). *Language use and language learning in CLIL classrooms*. Amsterdam: John Benjamins, 1-19.

Davies, Alan. 1989. "Communicative competence as language use". *Applied Linguistics* 10(2), 157-170.

de Courcy, Michèle. 2002. *Learners' experience of immersion education: case studies of French and Chinese*. Clevedon: Multilingual Matters.

De Wilde, Vanessa; Brysbaert, Marc; Eyckmans, June. 2020. "Learning English throughout-of-school exposure: which levels of language proficiency are attained and which types of input are important?". *Bilingualism: Language and Cognition* 23, 171-185.

Dörnyei, Zoltán. 2010. *Questionnaires in second language research: construction, administration, and processing*. New York: Routledge.

Ebrahimzadeh, Mohsen. 2017. "Readers, players, and watchers: EFL students' vocabulary acquisition through digital video games". *English Language Teaching* 10(2), 1-18.

Edwards, John. 2006. "Foundations of bilingualism". In Bhatia, Tej K.; Ritchie, William C. (eds.). *The handbook of bilingualism*. Malden, MA: Blackwell Publishing, 7-31.

Ellis, Nick C. 2001. "Memory for language". In Robinson, Peter (ed.). *Cognition and Second Language Instruction.* (1st edition). Cambridge: Cambridge University Press, 33-68.

Engber, Cheryl A. 1995. "The relationship of lexical proficiency to the quality of ESL compositions". *Journal of Second Language Writing* 4(2), 139-155.

Gardner, Dee; Davies, Mark. 2014. "A new Academic Word List". *Applied Linguistics* 35(3), 305-327.

Ghadessy, Mohsen. 1979. "Frequency counts, words lists, and materials preparation: a new approach". *English Teaching Forum* 17, 24-27.

Gierlinger, Erwin M.; Wagner, Thomas. A. 2016. "The more the merrier – revisiting CLIL-based vocabulary growth in secondary education". *Latin American Journal of Content & Language Integrated Learning* 9(1), 37-63.

González Fernández, Beatriz; Schmitt, Norbert. 2015. "How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure". *International Journal of Applied Linguistics* 166(1), 94-126.

Grundy, John. G.; Timmer, Kalinka. 2017. "Bilingualism and working memory capacity: a comprehensive meta-analysis". *Second Language Research* 33(3), 325-340.

Guiraud, Pierre. 1954. *Les caracteres statistiques du vocabulaire: essai de méthodologie*. Paris: Presses universitaires de France.

Hakuta, Kenji; Bialystok, Ellen; Wiley, Edward. 2003. "Critical evidence: a test of the critical-period hypothesis for second-language acquisition". *Psychological Evidence* 14(1), 31-38.

Haugen, Einar. 1953. *The Norwegian language in America: a study in bilingual behavior*. Philadelphia: University of Pennsylvania Press.

Higginbotham, George; Reid, Jacqui. 2019. "The lexical sophistication of second language learners' academic essays". *Journal of English for Academic Purposes* 37, 127-140.

Hinnenkamp, Volker. 2005. "Semilingualism, double monolingualism and blurred genres - on (not) speaking a legitimate language". *Journal of Social Science Education* 4(1), 57-90.

Housen, Alex; Kuiken, Folkert; Vedder, Ineke (eds.). 2012. *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA.* Amsterdam: John Benjamins.

Hulstijn, Jan H. 2011. "Language proficiency in native and nonnative speakers: an agenda for research and suggestions for second-language assessment". *Language Assessment Quarterly* 8(3), 229-249.

Ife, Anne; Vives Boix, G.; Meara, Paul M. 2000. "The impact of study abroad on the vocabulary development of different proficiency groups". *Spanish Applied Linguistics* 4(1), 55-84.

Isabelli-García, Christina. 2006. "Study abroad social networks, motivation and attitudes: implications for second language acquisition". In DuFon, Margaret A.; Churchill, Eton (eds.). *Language learners in study abroad contexts*. Clevedon: Multilingual Matters, 231-258.

Jiménez Catalán, Rosa María. 2003. "Sex differences in L2 vocabulary learning strategies". *International Journal of Applied Linguistics* 13(1), 54-77.

Juan-Garau, Maria. 2018. "Exploring oral L2 fluency development during a three-month stay abroad through a dialogic task". In Sanz, Cristina; Morales-Front, Alfonso (eds.). *The Routledge handbook of study abroad research and practice*. Milton Park: Routledge, 193-208.

Kalantari, Reza; Gholami, Javad. 2017. "Lexical complexity development from dynamic systems theory perspective: lexical density, diversity, and sophistication". *International Journal of Instruction* 10(4), 1-18.

Kapa, Leah L.; Colombo, John. 2013. "Attentional control in early and later bilingual children". *Cognitive Development* 28(3), 233-246.

Keshavarz, Mohammad H.; Astaneh, Hamideh. 2004. "The impact of bilinguality on the learning of English vocabulary as a foreign language (L3)". *International Journal of Bilingual Education and Bilingualism* 7(4), 295-302.

Kojima, Masumi; Yamashita, Junko. 2014. "Reliability of lexical richness measures based on word lists in short second language productions". *System* 42, 23-33.

Krizman, Jennifer; Skoe, Erika; Marian, Viorica; Kraus, Nina. 2014. "Bilingualism increases neural response consistency and attentional control: evidence for sensory and cognitive coupling". *Brain and Language* 128(1), 34-40.

Lambert, Wallace E. 1974. "Culture and language as factors in learning and education". Paper presented at the *Fifth Annual Learning Symposium,* Western Washington State College, Bellingham, WA, November 1973, 1-54.

Lapkin, Sharon; Hart, Doug; Swain, Merrill. 1995. "A Canadian interprovincial exchange: evaluating the linguistic impact of a three-month stay in Quebec". In Freed, Barbara (ed.). *Second language acquisition in a study abroad context*. Amsterdam: John Benjamins, 67-94.

Lasagabaster, David. 2008. "Foreign language competence in content and language integrated courses". *The Open Applied Linguistics Journal* 1(1), 30-41.

Laufer, Batia. 1995. "Beyond 2000 - a measure of productive lexicon in a second language". In Eubank, Lynn; Sharwood-Smith, Michael; Selinker, Larry (eds.). *The current state of interlanguage*. Amsterdam: John Benjamins, 265-272.

Laufer, Batia; Nation, Paul. 1995. "Vocabulary size and use: lexical richness in L2 written production". *Applied Linguistics* 16(3), 307-322.

Laufer, Batia; Paribakht, T. Sima. 1998. "The relationship between passive and active vocabularies: effects of language learning context". *Language Learning* 48(3), 365-391.

Laufer, Batia. 2014. "Vocabulary in a second language: selection, acquisition, and testing: a commentary on four studies for JALT vocabulary SIG". *Vocabulary Learning and Instruction* 3(2), 38-46.

Lee, Sunjung; Pulido, Diana. 2017. "The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading". *Language Teaching Research* 21(1), 118-135.

Leiner, Dominik J. 2019. *SoSci Survey*. (Version 3.1.06). [Computer Program]. https://www.soscisurvey.de (12 Jan. 2022).

Leńko-Szymańska, Agnieszka. 2020. *Defining and assessing lexical proficiency*. New York, NY: Routledge.

Lin, Lu-Fang. 2011. "Gender differences in L2 comprehension and vocabulary learning in the video-based CALL program". *Journal of Language Teaching and Research* 2(2), 295-301.

Lindqvist, Christina; Gudmundson, Anna; Bardel, Camilla. 2013. "A new approach to measuring lexical sophistication in L2 oral production". In Bardel, Camilla; Lindqvist, Christina; Laufer, Batia (eds.). *L2 vocabulary acquisition, knowledge and use: new perspectives on assessment and corpus analysis*. Amsterdam: EuroSLA, 109-126.

Llanes, Àngels; Muñoz, Carmen. 2009. "A short stay abroad: does it make a difference?". *System* 37(3), 353-365.

Lynn, Robert W. 1973. "Preparing word lists: a suggested method". *RELC Journal* 4, 25-32.

MacSwan, Jeff. 2000. "The threshold hypothesis, semilingualism, and other contributions to a deficit view of linguistic minorities". *Hispanic Journal of Behavioural Sciences* 22(1), 3-45.

Malakoff, Marguerite; Hakuta, Kenji. 1991. "Translation skill and metalinguistic awareness in bilinguals". In Bialystok, Ellen (ed.). *Language processing in bilingual children*. (1st edition). Cambridge: Cambridge University Press, 141-166.

Malvern, David; Richards, Brian; Chipere, Ngoni; Durán, Pilar. 2004. *Lexical diversity and language development: quantification and assessment*. Basingstoke: Palgrave Macmillan.

Matiasek, Stefanie. 2005. "English as the language of instruction in Austrian chemistry lessons: instance of explicit language teaching". MA thesis, University of Vienna.

Martin, Anne. 1976. "Teaching academic vocabulary to foreign graduate students". *TESOL Quarterly* 10(1), 91-97.

McLaughlin, Barry. 1984. *Second language acquisition in childhood. Vol. 1: preschool children*. (2nd edition). Hillsdale, NJ: Lawrence Erlbaum.

Meara, Paul. 1996. "The dimensions of lexical competence". In Brown, Gillian; Malmkjaer, Kirsten; Williams, John (eds.). *Performance and competence in second language acquisition*. Cambridge: Cambridge University Press, 35-53.

Meara, Paul; Bell, Huw. 2001. "P_Lex: a simple and effective way of describing the lexical characteristics of short L2 texts". *Prospect* 16(3), 5-19.

Meara, Paul; Miralpeix, Imma. 2016. *Tools for researching vocabulary*. Bristol: Multilingual Matters.

Meara, Paul; Miralpeix, Imma. 2018. *D_Tools*. [Computer Program]. https://lognostics.co.uk/tools/D_Tools/D_Tools.htm (5 Sept. 2021).

Merikivi, Riika; Pietilä, Päivi. 2014. "Vocabulary in CLIL and in mainstream education". *Journal of Language Teaching and Research* 5(3), 487-497.

Milton, James; Meara, Paul. 1995. "How periods abroad affect vocabulary growth in a foreign language". *ITL Review of Applied Linguistics* 107-108, 17-34.

Milton, James. 2009. *Measuring second language vocabulary acquisition*. Bristol: Channel View Publications.

Miralpeix, Imma. 2006. "Age and vocabulary acquisition in English as a foreign language". In Muñoz, Carmen (ed.). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters, 89-106.

Miyake, Akira; Friedman, Naomi P.; Emerson, Michael J.; Witzki, Alexander H.; Howerter, Amy; Wager, Tor D. 2000. "The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: a latent variable analysis". *Cognitive Psychology* 41(1), 49-100.

Morales, Julia; Calvo, Alejandra; Bialystok, Ellen. 2013. "Working memory development in monolingual and bilingual children". *Journal of Experimental Child Psychology* 114(2), 187-202.

Nation, Paul. 1984. *Vocabulary lists: words, affixes and stems*. Occasional Publication No 12. Victoria University of Wellington.

Nation, Paul; Waring, Robert. 1997. "Vocabulary size, text coverage and word lists". In Schmitt, Norbert; McCarthy; Michael (eds.). *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press, 6-19.

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

Nation, Paul; Meara, Paul. 2010. "Vocabulary". In Schmitt, Norbert (ed.). *An introduction to applied linguistics*. New York, NY: Hodder, 34-52.

Nation, Paul; Webb, Stuart. 2011. *Researching and analyzing vocabulary*. Boston, MA : Heinle, Cengage Learning.

Nation, Paul; Anthony, Laurence. 2013. "Mid-frequency readers". *Journal of Extensive Reading* 1(1), 5-16.

Nation, Paul. 2013. *Learning vocabulary in another language*. (2nd edition). Cambridge: Cambridge University Press.

Nation, Paul. 2016. *Making and using word lists for language learning and testing*. Amsterdam: John Benjamins.

Nation, Paul. n.d. *Range*. [Computer Program]. www.victoria.ac.nz/lals/about/staff/paul-nation (5 Sept. 2021).

Olsson, Eva; Sylvén, Liss. 2015. "Extramural English and academic vocabulary: a longitudinal study of CLIL and non-CLIL students in Sweden". *Journal of Applied Language Studies* 9(2), 77-103.

Olsson, Eva; Sylvén, Liss. 2017. "Validity in high- and low-stakes tests: a comparison of academic vocabulary and some lexical features in CLIL and non-CLIL students' written texts". *Oslo Studies in Language* 9(3), 127-146.

Olsson, Eva. 2021. "A comparative study of CLIL implementation in upper secondary school in Sweden and students' development of L2 English academic vocabulary". *Language Teaching Research*, 1-26.

Pahom, Olga; Farley, Andrew; Ramonda, Kris. 2015. "Are the best language learners from Mars or from Venus? Gender and vocabulary acquisition in the L2 Spanish classroom". *The Reading Matrix: An International Online Journal* 15, 158-172.

Paquot, Magali. 2010. *Academic vocabulary in learner writing: from extraction to analysis*. London: Continuum.

Paran, Amos. 2013. "Content and language integrated learning: panacea or policy borrowing myth?". *Applied Linguistics Review* 4(2), 317-342.

Peal, Elizabeth; Lambert, Wallace E. 1962. "The relation of bilingualism to intelligence". *Psychological Monographs: General and Applied* 76(27), 1-23.

Pérez-Vidal, Carmen. 2014. "Language acquisition in study abroad and formal instruction contexts". In Pérez-Vidal, Carmen (ed.). *Language acquisition in study abroad and formal instruction contexts*. Amsterdam: John Benjamins, 17-58.

Perez, Montero M. 2020. "Incidental vocabulary learning through viewing video: the role of vocabulary knowledge and working memory". *Studies in Second Language Acquisition* 42(4), 749-773.

Peters, Elke; Webb, Stuart. 2018. "Incidental vocabulary acquisition through viewing L2 television and factors that affect learning". *Studies in Second Language Acquisition* 40, 551-577.

Peters, Elke; Noreillie, Ann-Sophie; Heylen, Kris; Bulté, Bram; Desmet, Piet. 2019. "The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages". *Language Learning* 69(3), 747-782.

Peterson, Mark. 2010. "Massively multiplayer online role-playing games as arenas for second language learning". *Computer Assisted Language Learning* 23(5), 429-439.

Praninskas, Jean. 1972. *American University Word List*. London: Longman.

Read, John. 2000. *Assessing vocabulary*. Cambridge: Cambridge University Press.

Richards, Jack. 1976. "The role of vocabulary teaching". *TESOL Quarterly* 10(1), 77-89.

Ruiz de Zarobe, Yolanda. 2008. "CLIL and foreign language learning: a longitudinal study in the Basque country". *International CLIL Research Journal* 1(1), 60-73.

Sagasta Errasti, María P. 2003. "Acquiring writing skills in a third language: the positive effects of bilingualism". *International Journal of Bilingualism* 7(1), 27-42.

Sanz, Cristina. 2000. "Bilingual education enhances third language acquisition: evidence from Catalonia". *Applied Psycholinguistics* 21, 23-44.

Sanz, Cristina; Morales-Front, Alfonso. 2018. "Issues in study abroad research and practice". In Sanz, Cristina; Morales-Front, Alfonso (eds.). *The Routledge handbook of study abroad research and practice*. Milton Park: Routledge, 1-16.

Saville-Troike, Muriel. 2005. *Introducing second language acquisition*. (1st edition). Cambridge: Cambridge University Press.

Schmid, Kurt; Gruber, Benjamin. 2012. "Internationale Schulen in Österreich: Aktuelles Angebot und Potential als Standortfaktor". *ibw research brief* 73, 1-4.

Schmitt, Norbert. 2010. *Researching vocabulary: a vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Schmitt, Norbert; Schmitt, Diane. 2014. "A reassessment of frequency and vocabulary size in L2 vocabulary teaching". *Language teaching* 47(4), 484-503.

Shucard, David W.; Shucard, Janet L.; Thomas, David G. 1987. "Sex differences in the patterns of scalp-recorded electrophysiological activity in infancy: possible implications for language development". In Philips, Susan U.; Steele, Susan; Tanz, Christine (eds.). *Language, gender, and sex in comparative perspective.* (1st edition). Cambridge: Cambridge University Press, 278-296.

Soliman, Abdrabo Moghazy M. 2014. "Bilingual advantages of working memory revisited: a latent variable examination". *Learning and Individual Differences* 32, 168-177.

Solon, Megan; Long, Avizia Yim. 2018. "Acquisition of phonetics and phonology abroad: what we know and how". In Sanz, Cristina; Morales-Front, Alfonso (eds.). *The Routledge handbook of study abroad research and practice*. Milton Park: Routledge, 71-85.

Steele, Susan. 1987. "Introduction". In Philips, Susan U.; Steele, Susan; Tanz, Christine (eds.). *Language, gender, and sex in comparative perspective.* (1st edition). Cambridge: Cambridge University Press, 263-267.

Stein, Gabriele. 2016. "Some thoughts on the issue of core vocabularies". *Applied Linguistics* 35(5), 759-763.

Steiner, George. 1992. *After Babel*. Oxford: Oxford University Press.

Stoeckel, Tim; Bennett, Phil. 2015. "A test of the New General Service List". *Vocabulary Learning and Instruction* 4(1), 1-8.

Sundqvist, Pia. 2009. "Extramural English matters: out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary". Dissertation, Karlstad University Studies.

Sylvén, Liss K. 2004. "Teaching in English or English teaching? On the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition". Doctoral dissertation, University of Gothenburg.

Sylvén, Liss K.; Sundqvist, Pia. 2012. "Gaming as extramural English L2 learning and L2 proficiency among young learners". *ReCALL* 24(3), 302-321.

Taylor, David S. 1988. "The meaning and use of the term 'competence' in linguistics and applied linguistics". *Applied Linguistics* 9(2), 148-168.

Taylor, Linda. 1990. *Teaching and learning vocabulary*. Hemel Hempstead: Prentice Hall International.

Thomas, Jacqueline. 1988. "The role played by metalinguistic awareness in second and third language learning". *Journal of Multilingual and Multicultural Development* 9(3), 235-246.

Thomas, Margaret. 1994. "Assessment of L2 proficiency in second language acquisition research". *Language Learning* 44(2), 307-336.

Treffers-Daller, Jeanine; Parslow, Patrick; Williams, Shirley. 2018. "Back to basics: how measures of lexical diversity can help discriminate between CEFR levels". *Applied Linguistics* 39(3), 302-327.

Ullman, Michael T.; Miranda, Robbin A.; Travers, Michelle L. 2008. "Sex differences in the neurocognition of language". In Becker, Jill B.; Berkley, Karen J.; Geary, Nori;

Hampson, Elizabeth; Herman, James P.; Young, Elizabeth (eds.). *Sex differences in the brain: from genes to behavior*. Oxford: Oxford University Press, 291-309.

Van der Slik, Frans W. P.; van Hout, Roeland W. N. M.; Schepens, Job J. 2015. "The gender gap in second language acquisition: gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues". *PLOS ONE* 10(11), 1-22.

Vulchanova, Mila; Foyn, Camilla H.; Nilsen, Randi A.; Sigmundsson, Hermundur H. 2014. "Links between phonological memory, first language competence and second language competence in 10-year-old children". *Learning and Individual Differences* 35, 87-95.

Wattendorf, Elise; Festman, Julia. 2008. "Images of the multilingual brain: the effect of age of second language acquisition". *Annual Review of Applied Linguistics* 28, 3-24.

Webb, Stuart. 2015. "Extensive viewing: language learning through watching television". In Nunan, David; Richards, Jack C. (eds.). *Language learning beyond the classroom*. New York: Routledge, 159-168.

Weinreich, Uriel. 1968. *Languages in contact: findings and problems*. The Hague: Mouton.

West, Michael. 1953. *A General Service List of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. London: Longman.

Wright, Clare. 2018. "Effects of time and task on L2 Mandarin Chinese language development during stay abroad". In Sanz, Cristina; Morales-Front, Alfonso (eds.). *The Routledge handbook of study abroad research and practice*. Milton Park: Routledge, 166-180.

Xue, Guoyi; Nation, Paul. 1984. "A University Word List". *Language Learning and Communication* 3(2), 215-229.

Zaytseva, Victoria; Pérez-Vidal, Carmen; Miralpeix, Imma. 2018. "Vocabulary acquisition during study abroad: a comprehensive review of the research". In Sanz, Cristina; Morales-Front, Alfonso (eds.). *The Routledge handbook of study abroad research and practice*. Milton Park: Routledge, 210-224.

9. Appendix

9.1 Abstract

This study is concerned with the development of students' lexical proficiency and whether certain aspects of lexical proficiency improve more noticeably than others. Furthermore, a potential effect of the participants' initial level of proficiency, linguistic background, age or gender on lexical proficiency is investigated.

For this purpose, 28 students of the English department of the University of Vienna filled in an online questionnaire and provided two written productions, namely their first seminar paper and their BA or BEd thesis. These texts were subsequently analysed regarding lexical variation, lexical density and lexical sophistication using the programme VocabProfile and the website *analyzemywriting.com*.

The results showed that all aspects of lexical proficiency improved slightly, with no differences in the extent of the development being detected. Although participants with a higher initial level also achieved better scores in the second written production, those with a lower initial level had a more noticeable development. Furthermore, spending time abroad, engaging in Extramural English activities and actively studying vocabulary also had a positive effect. Despite gender and age influencing lexical proficiency, these effects could partly be due to strong correlations with other factors.

The merely weak improvements regarding the development of different aspects of lexical proficiency might be attributable to the sample group's heterogeneity. Moreover, whilst strong correlations between lexical proficiency and various factors could be observed, there is little consensus in the research literature on these effects. Hence, further research, for instance by conducting a similar study on a larger scale, would be highly insightful.

9.2 Zusammenfassung

Die vorliegende Studie beschäftigt sich mit der Entwicklung lexikalischer Kompetenz von Studierenden und einem potentiellen Zusammenhang mit verschiedenen Faktoren. Insbesondere soll herausgefunden werden, ob bestimmte Aspekte lexikalischer Kompetenz sich stärker entwickeln und ob das Ausgangsniveau der lexikalischen Kompetenz der Studierenden sowie ihr sprachlicher Hintergrund, Alter oder Geschlecht einen Einfluss auf das Ausmaß dieser Entwicklung haben könnten.

Zu diesem Zweck füllten 28 Studierende des Instituts für Anglistik und Amerikanistik der Universität Wien einen Online-Fragebogen aus und stellten eine Proseminar-Arbeit sowie ihre BA oder BEd Arbeit, zur Verfügung. Diese wurden bezüglich lexikalischer Varianz, lexikalischer Dichte und der Verwendung fortgeschrittenen Vokabulars analysiert, wofür das Programm VocabProfile und die Website *analyzemywriting.com* verwendet wurden.

Obwohl sich alle Aspekte lexikalischer Kompetenz verbesserten, konnte kein Unterschied im Ausmaß der Entwicklung der verschiedenen Aspekte festgestellt werden. Außerdem schnitten Studierende mit einem höheren Ausgangsniveau in der zweiten schriftlichen Arbeit besser ab, wenngleich jene mit einem niedrigeren Ausgangsniveau eine stärkere Entwicklung durchliefen. Auslandsaufenthalte, außerschulische Englischaktivitäten und aktives Vokabellernen hatten ebenfalls positive Auswirkungen. Obwohl auch für Geschlecht und Alter Zusammenhänge mit einer höheren lexikalischen Kompetenz gefunden werden konnten, könnte dies teilweise auf Korrelationen mit anderen Faktoren zurückzuführen sein.

Die lediglich schwache Entwicklung verschiedener Aspekte lexikalischer Kompetenz könnte sich teilweise durch die Heterogenität der Stichprobe erklären lassen. Starke Korrelationen zwischen verschiedenen Faktoren und der Entwicklung lexikalischer Kompetenz konnten zwar festgestellt werden, dieser Zusammenhang ist in der Forschungsliteratur jedoch teilweise umstritten. Daher braucht es weitere Forschung in diesem Bereich, wie beispielsweise eine ähnliche aber größer angelegte Studie.

9.3 Questionnaire

Dear Colleagues,

thank you very much for participating in my study, which consists of two parts:
- this questionnaire
- the upload of your PS 1 Linguistics paper and your BA thesis in Linguistics

The purpose of this research project is to investigate academic vocabulary. Further analysis of data will be strictly anonymised and the results of this study will be used solely for research purposes. You can withdraw from the study at any time without any consequences by sending an e-mail to a01405462@unet.univie.ac.at.
The questionnaire will take approximately 5-10 minutes. Please do not leave out any questions and answer all questions honestly as there are no right or wrong answers.

Do you want to participate in this study?

○ Yes

○ No (questionnaire will be stopped)

What is your gender?

○ Male

○ Female

○ Other/ Prefer not to answer

At which school did you take your Matura exam (= A-levels)?

○ AHS in Austria

○ BHS in Austria

○ Other school in Austria

○ Not in Austria

Age when submitting BA thesis in linguistics

○ 20-23

○ 24-27

○ 28-30

○ older than 30

Which undergraduate degree are you currently enrolled in?

- ☐ English and American Studies

- ☐ Other studies (please indicate)

Have you completed any other undergraduate degree in the last 5 years?

- ☐ No

- ☐ Yes – English and American Studies

- ☐ Yes – other studies (please indicate which ones)

In which semester did you take the PS1 Linguistics seminar (e.g. SS17)?

In which semester did you take the BA paper seminar for linguistics (e.g. WS19)?

In which country did you grow up and attend school?

- ○ Austria

- ○ Other (please specify)

Which language(s) is/ are your first language(s)? (more than one choice possible)

- ☐ German

- ☐ English

- ☐ Other language(s)

- ☐ Other language(s)

Does any of your parents have a different first language than the one you indicated above? If yes, which language?

- ☐ No

- ☐ Yes, my mother (which language?):

- ☐ Yes, my father (which language?):

Have you been raised speaking one language, two languages (if yes, which ones?) or more than two languages (if yes which ones?)?

- ○ one language

two languages (which ones)
- ○ [   ]

more than two languages (please indicate which ones)
- ○ [   ]

Is your ability to speak all of these languages the same now as it was when you were a child?

- ○ It is the same.
- ○ It has improved.
- ○ It has deteriorated.

At which age did you start learning additional languages to your first language(s) and how good are you at each language now? (e.g.: English 8: C1, French 14: B1, Spanish 21: A1)

[   ]
[   ]

For what purpose did you study the languages (e.g. French: at school, Hungarian: because I moved, etc.)

[   ]
[   ]

How good is your reading ability in your first language? Please choose one of the following CEFR levels.

- ○ A1
- ○ A2
- ○ B1
- ○ B2
- ○ C1
- ○ C2

Did you go to a bilingual school or did you have CLIL lessons (e.g. History lessons in English instead of German) in school?

○ Yes

○ No

How many of your classes were taught in a foreign language?

○ 0-25%

○ 25-50%

○ 50-75 %

○ 75-100%

Which foreign language was used to teach (CLIL lessons or other lessons apart from language classes)?

☐ English

☐ Other languages (please specify)

[ ]

Have you ever spent time abroad (longer than 1 month) before finishing your BA thesis in Linguistics?

○ Yes

○ No

In which countries did you spend time abroad,

for how long,

when,

for which purpose

and did you visit a language school?

(e.g. London, 3 months, 2017, semester at university abroad, no language school)

[ ]

[ ]

While you were there, how much of your communication was in the official language of the country of your stay?

○ 0-25%

○ 25-50%

○ 50-75%

○ 75-100%

While you were there, did you have to hand in any writing assignments (e.g. papers, essays, reflections etc.) in the official language of the country of your stay?

○ Yes

○ No

Do or did you frequently spend time abroad (e.g. spend weekends with family members in another country speaking a language other than German) before finishing your BA studies?

○ Yes (please specify)
  [   ]

○ No

Think back to the time between starting your BA studies and writing the PS 1 paper. How often would you have talked to people in English in a typical week in your free time (i.e. not in seminars)?

○ less than 1 hour a week
○ 1-5 hours a week
○ 5-10 hours a week
○ 10-15 hours a week
○ 15- 20 hours a week
○ 20-30 hours a week
○ more than 30 hours a week

Did this change in the time between writing the PS1 Linguistics paper and the BA thesis?

○ it has stopped
○ it has become a bit less
○ it has stayed the same
○ it has become a bit more
○ it has become a lot more

Think back to the time between starting your BA studies and writing the PS 1 paper. How often would you have watched films or videos in English in a typical week in your free time (i.e. not in seminars)?

○ less than 1 hour a week
○ 1-5 hours a week
○ 5-10 hours a week
○ 10-15hours a week
○ 15-20 hours a week
○ 20-30 hours a week
○ more than 30 hours a week

Did this change in the time between writing the PS1 Linguistics paper and the BA thesis?

○ it has stopped

○ it has become a bit less

○ it has stayed the same

○ it has become a bit more

○ it has become a lot more

Think back to the time between starting your BA studies and writing the PS 1 paper. How often would you have played video games in English in a typical week in your free time (i.e. not in seminars)?

○ less than 1 hour a week

○ 1-5 hours a week

○ 5-10 hours a week

○ 10-15hours a week

○ 15-20 hours a week

○ 20-30 hours a week

○ more than 30 hours a week

Did this change in the time between writing the PS1 Linguistics paper and the BA thesis?

○ it has stopped

○ it has become a bit less

○ it has stayed the same

○ it has become a bit more

○ it has become a lot more

Think back to the time between starting your BA studies and writing the PS 1 paper. How often would you have read English books, articles or texts in general (apart from texts for university) in a typical week in your free time (i.e. not in seminars)?

○ less than 1 hour a week

○ 1-5 hours a week

○ 5-10 hours a week

○ 10-15hours a week

○ 15-20 hours a week

○ 20-30 hours a week

○ more than 30 hours a week

Did this change in the time between writing the PS1 Linguistics paper and the BA thesis?

- ○ it has stopped
- ○ it has become a bit less
- ○ it has stayed the same
- ○ it has become a bit more
- ○ it has become a lot more

Have you ever studied vocabulary items on purpose to improve your writing at university (for example by studying academic word lists)?

If yes what did you do? (e.g. "I made a list of words I did not know when reading a text and studied them afterwards.")

- ○ Yes (please specify)

- ○ No

Do you frequently look for more academic words when writing a paper? (e.g. looking for more academic synonyms for 'maybe')?

If yes, where do you look for them?

- ☐ No
- ☐ Yes – dictionaries
- ☐ Yes – online dictionaries
- ☐ Yes – search engines (e.g. Google)
- ☐ Yes – others (please indicate)

Have these habits regarding studying vocabulary changed in the time between writing the PS1 Linguistics paper and the BA thesis?

- ○ it has stopped
- ○ it has become a bit less
- ○ it has stayed the same
- ○ it has become a bit more
- ○ it has become a lot more

Please upload your uncorrected PS 1 Linguistics paper as a WORD document (if possible) here.

Durchsuchen... Keine Datei ausgewählt.

Please upload your uncorrected BA thesis in Linguistics as a WORD document (if possible) here.

Durchsuchen... Keine Datei ausgewählt.

If you had any problems with the questionnaire, e.g. with uploading your papers, please leave a comment with your matriculation number and contact me via e-mail: a01405462@unet.univie.ac.at.

If you have any questions, requests or other comments, feel free to contact me as well.

Thank you very much for participating!

Your answers were transmitted, you may close the browser window or tab now.

9.4 Number of words from each section

Table 4. Number of words used from each section of the written productions

Participant	Introduction	Conclusion	Lit. Rev.	Methodology	Total	Participant	Introduction	Conclusion	Lit. Rev.	Methodology	Total
1_WP1	170	130	700	0	1000	1_WP2	170	130	700	0	1000
2_WP1	134	95	771	0	1000	2_WP2	170	130	700	0	1000
3_WP1	170	130	700	0	1000	3_WP2	170	130	300	400	1000
4_WP1	191	109	700	0	1000	4_WP2	170	130	700	0	1000
5_WP1	170	130	700	0	1000	5_WP2	170	130	700	0	1000
6_WP1	170	130	700	0	1000	6_WP2	170	130	700	0	1000
7_WP1	114	72	814	0	1000	7_WP2	170	130	700	0	1000
8_WP1	170	130	700	0	1000	8_WP2	170	130	700	0	1000
9_WP1	170	130	700	0	1000	9_WP2	170	130	700	0	1000
10_WP1	170	130	700	0	1000	10_WP2	170	130	700	0	1000
11_WP1	170	130	700	0	1000	11_WP2	170	130	700	0	1000
12_WP1	170	130	700	0	1000	12_WP2	276	24	700	0	1000
13_WP1	170	130	700	0	1000	13_WP2	115	185	700	0	1000
14_WP1	170	130	700	0	1000	14_WP2	170	130	700	0	1000
15_WP1	170	130	700	0	1000	15_WP2	170	130	700	0	1000
16_WP1	170	130	700	0	1000	16_WP2	170	130	700	0	1000
17_WP1	182	118	700	0	1000	17_WP2	170	130	700	0	1000
18_WP1	170	130	700	0	1000	18_WP2	170	130	700	0	1000
19_WP1	170	130	700	0	1000	19_WP2	170	130	700	0	1000
20_WP1	170	130	700	0	1000	20_WP2	170	130	700	0	1000
21_WP1	170	130	700	0	1000	21_WP2	170	130	700	0	1000
22_WP1	170	130	700	0	1000	22_WP2	170	130	700	0	1000
23_WP1	170	130	700	0	1000	23_WP2	170	130	700	0	1000
24_WP1	170	130	700	0	1000	24_WP2	170	130	700	0	1000
25_WP1	170	130	700	0	1000	25_WP2	170	130	700	0	1000
26_WP1	170	130	587	113	1000	26_WP2	170	130	700	0	1000
27_WP1	132	168	358	342	1000	27_WP2	170	130	700	0	1000
28_WP1	170	130	676	24	1000	28_WP2	170	130	700	0	1000

9.5 Codes for substituted and adapted items

Table 5. Codes for substituted and adapted items; underlined sections indicate that only parts of the example were substituted

Code	Category	Examples
SUBSTIN	Names of people	Names of programmes, websites, etc. Bialystok
	Names of programmes or concepts named after people	McGurk effect, Labovian, Likert scales, Spearmans correlation, Collins Dictionary, Meringer corpus, Heschl's gyrus
	Names of programmes or websites	NOW corpus, MAXQDA, Excel, website *Word Spy*, Facebook
SUBSTIY	Specific references to years	1953
SUBSTIT	Titles of books or articles	Chomsky's _Aspects of the Theory of Syntax_
SUBSTIA	abbreviations	i.e., e.g., etc., c.f.
SUBSTIP	Phonetic transcription or affixes in angle brackets	<r>, the _<–s>_ plural
Diff_lang	Words from languages other than English	the modal auxiliary _werden_
Original item	**Adapted item**	
%	percent	
Ca.	circa	
L1	First language	
L2	Second language	
Bi- and multilingual	Bilingual and multilingual	

9.6 Information on participants

Table 6. Participants' sociodemographic and linguistic background

participant	gender	age	studies	semesters betw. papers	biling. up-bringing	language score	bilingual education	percent biling. ed.	number of times abroad	purpose of time abroad	family time abroad	EE score	vocabulary studied
1	female	20-23	BA	2	no	8,00	no		0		no	9	no
2	female	20-23	BA	4	no	10,00	no		0		no	9	yes
3	female	20-23	BA	4	no	9,00	no		0		no	11	no
4	male	24-27	BA	8	no	6,00	no		0		yes	4	yes
5	female	24-27	BA	2	no	7,00	no		3	study, work	yes	6	yes
6	male	24-27	BA	5	no	16,00	no		3	study, work	yes	14	yes
7	female	20-23	BA	3	yes	8,00	yes	0-25%	0		yes	6	yes
8	female	20-23	BA	2	no	8,00	no		0		no	10	no
9	female	20-23	BA	2	no	16,00	no		1	study	yes	8	yes
10	female	20-23	BA	3	no	9,00	no		0		no	7	no
11	female	27-30	BA	3	no	15,00	no		1	study	yes	11	no
12	female	20-23	BA	4	no	12,00	yes	0-25%	1	study	no	16	yes
13	female	24-27	BA	2	no	9,00	yes	75-100%	0		no	14	no
14	female	20-23	BA	3	no	10,00	yes	0-25%	1	study	yes	16	yes
15	female	20-23	BA	3	no	8,00	no		1	study	no	13	yes
16	female	20-23	BA	3	no	11,00	no		1	study	no	13	yes
17	female	20-23	BA	4	no	10,00	no		1	study	no	9	yes
18	female	20-23	Bed	1	no	6,00	yes	0-25%	0		no	11	no
19	female	24-27	BA	6	no	11,00	no		0		no	8	yes
20	male	20-23	BEd	4	yes	7,00	no		0		yes	13	no
21	female	20-23	BA	3	no	8,00	no		0		no	13	no
22	female	24-27	BEd	1	no	17,00	no		1	work	yes	5	no
23	female	24-27	BA	7	no	10,00	no		2	study	no	8	yes
24	female	20-23	BEd	2	no	18,00	no		0		no	9	yes

25	female	20-23	BEd	2	no	5,00	no		0		no	7	yes
26	male	20-23	BEd	1	no	6,00	no		0		yes	18	no
27	male	20-23	BEd	1	no	5,00	no		0		no	12	no
28	female	20-23	BEd	1	yes	9,00	no		0		yes	12	no

9.7 Individual lexical proficiency scores

Table 7. Scores of all participants for all lexical proficiency measures

participant	lexical density 1	lexical density 2	TTR 1	TTR 2	1NGSL1	1NGSL2	1NGSL3	1NAWL	1OFFLIST	2NGSL1	2NGSL2	2NGSL3	2NAWL	2OFFLIST
1	52,20	55,04	0,33	0,35	63,11	15,24	6,10	5,18	10,37	66,10	13,11	8,55	3,70	8,55
2	60,90	62,73	0,33	0,35	56,12	13,43	3,88	10,15	16,42	58,87	12,96	6,48	8,17	13,52
3	57,60	56,10	0,35	0,35	60,17	16,67	6,78	5,65	10,73	60,62	13,60	5,38	8,22	12,18
4	56,90	55,84	0,42	0,40	61,43	14,29	5,95	7,14	11,19	59,75	15,25	5,00	6,50	13,50
5	57,90	56,24	0,38	0,33	61,88	16,97	6,79	6,01	8,36	61,49	14,33	7,76	6,57	9,85
6	56,40	59,54	0,43	0,42	58,96	14,29	4,99	6,35	15,42	55,19	12,97	9,43	9,43	12,97
7	56,39	53,85	0,40	0,38	62,03	14,39	6,95	5,21	11,91	61,74	12,93	5,28	4,49	15,57
8	57,56	58,50	0,31	0,37	73,57	9,24	5,10	3,50	8,60	62,70	11,62	7,30	5,95	12,43
9	58,10	58,40	0,38	0,34	60,99	13,87	6,81	4,97	13,35	66,18	13,01	6,07	6,36	8,38
10	58,30	55,71	0,35	0,39	66,95	14,81	5,70	3,70	8,83	63,33	14,62	7,69	4,62	9,74
11	50,10	51,50	0,34	0,38	65,22	8,41	7,54	2,03	16,81	67,97	11,98	7,81	6,51	5,73
12	54,10	54,30	0,36	0,35	59,61	11,70	5,29	8,36	15,04	63,11	15,27	7,20	6,05	8,36
13	51,80	51,50	0,38	0,38	63,19	13,05	5,22	7,57	10,97	60,16	12,76	8,59	8,07	10,42
14	53,83	55,63	0,36	0,41	60,06	14,05	4,68	6,61	14,88	57,60	12,99	6,62	7,11	15,69
15	55,40	54,79	0,36	0,38	61,64	14,25	5,21	7,95	10,96	63,42	16,84	6,05	4,21	9,47
16	54,30	56,90	0,36	0,37	58,36	13,70	7,67	7,95	12,33	58,67	8,53	8,53	6,67	17,60
17	55,50	56,50	0,39	0,39	53,44	15,78	8,40	5,85	16,54	58,72	14,87	6,15	7,95	12,31
18	55,26	58,30	0,34	0,33	65,79	11,40	7,60	5,26	9,94	67,55	11,21	7,96	6,49	6,78
19	62,20	60,14	0,35	0,40	58,76	14,69	6,50	7,63	12,43	51,23	17,00	6,90	6,40	18,47

20	59,44	58,54	0,31	0,36	61,61	9,03	7,10	7,10	15,16	56,08	16,02	5,25	8,01	14,64
21	60,40	52,45	0,38	0,41	52,06	15,98	6,70	8,25	17,01	56,48	14,91	6,36	6,85	15,40
22	58,10	53,10	0,35	0,33	64,23	14,37	6,20	5,07	10,14	61,75	13,86	5,42	7,83	11,14
23	58,60	59,50	0,42	0,40	63,88	14,59	6,46	5,98	9,09	52,13	14,54	8,02	8,77	16,54
24	55,16	61,20	0,40	0,34	57,80	13,41	5,61	7,32	15,85	60,62	15,86	7,37	6,23	9,92
25	58,60	59,70	0,34	0,33	62,91	14,24	5,04	7,12	10,68	62,95	12,95	4,82	5,72	13,55
26	55,00	54,50	0,40	0,37	63,68	12,19	7,71	3,73	12,69	58,40	13,33	7,73	5,07	15,47
27	54,25	55,46	0,37	0,40	64,71	14,44	8,82	5,35	6,68	59,51	14,32	6,91	4,44	14,81
28	56,10	57,70	0,31	0,34	62,89	13,52	8,81	3,46	11,32	59,52	14,58	8,33	6,85	10,71