Univerzita Karlova

Filozofická fakulta

Ústav anglického jazyka a didaktiky

Diplomová práce

Martin Sotona

Examining lexical complexity in the written production of L2 proficient learners of English

Lexikální komplexnost písemného projevu nerodilých mluvčích anglického jazyka

Praha, leden 2022

Vedoucí práce: PhDr. Tomáš Gráf, Ph.D.

Poděkování

Rád bych poděkoval PhDr. Tomáši Gráfovi, Ph.D. za ochotu a cenné rady, které mi při psaní diplomové práce poskytl.

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 2. ledna 2022 Martin Sotona

Abstrakt

Diplomová práce analyzuje lexikální komplexnost písemného jazyka nerodilých mluvčí (L2) anglického jazyka na úrovni C2 s následnou komparací výsledné lexikální komplexnosti s komplexností u rodilých mluvčí (L1). Lexikální komplexnost jazyka je sledována ve dvou klíčových parametrech, v dimenzi lexikální rozmanitosti (lexical diversity) a lexikální propracovanosti (lexical sophistication). Práce obsahuje kvantitativní analýzu dat s pomocí jednoduchých indikátorů, která je doprovázena analýzou s kompozitními metrikami VOCD-D a MTLD v oblasti lexikální rozmanitosti. Lexikální sofistikovanost je analyzována prostřednictvím nástroje English Vocabulary Profile (EVP), který měří výskyt slov v textu na základě předdefinovaného frekvenčního seznamu slov podle kategorií jazykové úrovně A1-C2. Základní stanovená hypotéza předpokládala nižší lexikální komplexnost nerodilých mluvčích jazyka ve srovnání s rodilými mluvčími, a to i přes nejvyšší možnou jazykovou kompetenci nerodilých mluvčích na úrovni C2, která je přirovnávána k „native-speaker competence". V rámci jednotlivých skupin uživatelů jazyka (L2 a L1) se očekávaly srovnatelné výstupy. Další pomocnou hypotézou byl nižší předpokládaný počet slov na úrovni C2 (low-frequency words) u L2 ve srovnání s L1.

Soubor dat, které byly podrobeny analýze, obsahuje 20 srovnatelných textů nerodilých mluvčích anglického jazyka na úrovni C2 co do jejich rozsahu (s průměrnou délkou 781 slov), žánru (ekonomie), cíle (napsání komentáře k publikovanému odbornému článku) a struktury (doporučená struktura komentáře); a 20 textů rodilých mluvčích anglického jazyka opět s podobnými parametrickými vlastnostmi. Data byla analyzována ve svých skupinách (L2 a L1) samostatně a také vzájemně mezi sebou. Pro analýzy byl využit automatizovaný webový nástroj Text Inspector, jehož bližší popis je součástí práce.

Lexikální rozmanitost byla srovnatelná v rámci jednotlivých skupin. Při komparaci byla prokazatelně vyšší u L1 uživatelů jazyka a to při použití obou nástrojů (VOCD-D a MTLD). Dalším poznatkem je, že L2 uživatelé mají tendenci více opakovat stejná slova bez většího využití synonymie. Výsledky srovnávání v dimenzi lexikální propracovanosti s využitím frekvenčních seznamů slov nepotvrdily hypotézu, neboť nebyly statisticky signifikantní. Rozdělení slov podle jednotlivých slovních kategorií se ukázalo jako přibližně stejné s klesajícím trendem od A1 do C2. Problémem je zde větší množství nekategorizovaných slov, což je důsledkem nástroje EVP, který mnohdy neobsahuje specifickou slovní zásobu z oblasti

ekonomie, a také specifičnosti ekonomických textů, které obsahují více čísel, vlastních jmen a zkratek (převážně akronymů) nežli texty obecné povahy. Pro větší spolehlivost výsledků by bylo vhodné výzkum dále rozšířit, ať už by se jednalo o větší délku textů nebo o jejich množství. Studie poukázala na citlivost žánru a proto by bylo žádoucí v dalším případném výzkumu stejný žánr dodržet. Dalším ovlivněním výzkumu je tzv. „priming effect“ (tendence L2 studentů opakovat stejná slova vyskytující se v článcích L1 autorů) , který automatizované nástroje nejsou schopny zachytit.

Klíčová slova

Lexikální komplexnost, jazyk nerodilých mluvčí, písemný jazyk, frekvenční seznam slov, English Vocabulary Profile (EVP), Text Inspector

Abstract

The thesis analyses lexical complexity in the written production of L2 proficient learners of English (the highest C2 level). Additionally, it compares L2 lexical complexity with L1 lexical complexity of English native speakers. This lexical complexity is investigated in two key parameters: lexical diversity and lexical sophistication. A quantitative analysis is made by the means of single indicators and is followed by an analysis where composite indicators VOCD-D and MTLD are employed to measure lexical diversity. Lexical sophistication is explored through the English Vocabulary Profile tool (EVP), which categorises words in a text according to predefined frequency word lists (A1-C2 types). The main hypothesis presumed that the lexical complexity of L2 English speakers is inferior to L1 English speakers, despite the fact that their L2 English language competence is at the highest level possible (C2 level), often compared to "native-speaker competence". It was expected that the results in respective groups (L2 and L1 speakers) would be similar. Another working hypothesis is that low-frequency words at the C2 level will be smaller for L2 English speakers than that for L1 English speakers.

The data comprises 20 comparable texts of L2 proficient English speakers in the dimension of their length (the mean is 781 words), genre (economy), aim (to write a commentary based on a published article) and structure (recommended structure of the commentary); and 20 texts of L1 English speakers, again with very similar parameters. Firstly, the data was analysed independently in their respective groups (L2 and L1) and afterwards, between these two distinctive groups. For these analyses, an automated website software Text Inspector was applied. Its detailed description can be found in the thesis.

The results showed that lexical diversity inside each group (L2 and L1) is similar. On the other hand, if compared, lexical diversity was substantially higher for L1 English speakers measured by VOCD-D and MTLD indicators. Another interesting result is that L2 English speakers tend to repeat the same words more than L1 English speakers, thus not using synonymy to a greater extent. In the case of lexical sophistication, the hypothesis has not been proven as the results were statistically insignificant. The word distribution into different types (A1-C2) reached similar results with the decreasing number of words from A1 to C2 in both groups. A large number of unlisted types was the result of the fact that the EVP database does not include specific economic lexis and that economic texts incorporate more numerals, proper

names and abbreviations (mainly acronyms) than other general texts. It would be advisable to increase the data sample in future research to achieve better reliability of results. It could be achieved via the greater number of texts or the longer text length. The research showed that there is a genre sensitivity and for this reason, it would be recommended to keep the same genre in follow-up research. Another strong influence on the results is a so-called “priming effect” (the tendency of students to repeat words from the chosen article in their commentaries), which automated tools like the Text Inspector can not measure.

Keywords

Lexical complexity, L2 language, written language, lexical frequency wordlist, English Vocabulary Profile (EVP), Text Inspector

Table of Contents

1 Introduction 12

2 Theoretical Part 15

2.1 Thesis Aim and Hypotheses 15

2.2 CEFR and Language Proficiency 15

2.3 Business English 19

2.4 Lexical Complexity 20

2.4.1 The CAF Model 22

2.4.2 Lexical Diversity – Concept and Studies 22

2.4.2.1 VOCD-D 23

2.4.2.2 MTLD 25

2.4.3 Lexical Sophistication – Concept and Studies 25

3 Empirical Part 27

3.1 Data 27

3.1.1 Data Selection 27

3.1.2 Data Processing 28

3.2 Methodology 29

3.2.1 The Text Inspector 29

3.2.1.1 Statistics and Readability 30

3.2.1.2 Lexical Diversity 31

3.2.1.3 English Vocabulary Profile 31

3.2.1.4 Other Text Inspector tools 32

3.2.2 Method 32

3.3 Results 33

3.3.1 Quantitative analysis 33

3.3.1.1 Number of Sentences 36
3.3.1.2 Average sentence length 36
3.3.1.3 Number of Tokens 37
3.3.1.4 Number of Types 38
3.3.1.5 Type/Token Ratio 39
3.3.1.6 Number Count 40
3.3.2 Lexical complexity analysis 40
3.3.2.1 Lexical diversity 40
3.3.2.2 Lexical sophistication 46
3.3.3 Discussion 51
3.3.3.1 Limitations of the study 53
4 Conclusion 55
5 Bibliography 57
6 Czech Résumé 62
7 Appendix 1 – L2 English Sample Text (1C) 67
8 Appendix 2 – L1 English Sample Text (1A) 71
9 Appendix 3 – L2 English Sample Text (20C) 73
10 Appendix 4 – L1 English Sample Text (20A) 76
11 Appendix 5 – Text Inspector´s Print-Screens for EVP (10A) 78

List of Figures

Figure 1: English Exams on the CEFR 16

Figure 2: The user/learner´s competences 18

Figure 4: VOCD means and sub-ranges (10th – 90th percentiles) of D for various cohorts ... 24

Figure 5: A comparison of sentence count between L1 and L2 texts 36

Figure 6: A comparison of average sentence length between L1 and L2 texts 37

Figure 7: A comparison of token count between L1 and L2 texts 37

Figure 8: A comparison of type count between L1 and L2 texts 38

Figure 9: A comparison of TTR between L1 and L2 texts 39

Figure 10: A comparison of number count between L1 and L2 texts 40

Figure 11: A comparison of VOCD-D between L1 and L2 texts 43

Figure 12: A comparison of MTLD between L1 and L2 texts 44

Figure 14: L2 text lexical sophistication 50

Figure 15: Comparison of lexical sophistication 50

List of Tables

Table 1: Common Reference Levels: global scale 17

Table 2: Number of tokens in L1 and L2 texts 28

Table 3: L2 text quantitative analysis 34

Table 4: L1 text quantitative analysis 35

Table 5: D values for three independent measures 42

Table 6: VOCD-D average values 42

Table 7: MTLD values 44

Table 8: VOCD-D and MTLD values 45

Table 9: L1 text lexical sophistication 47

Table 10: L2 text lexical sophistication 49

List of Abbreviations

AWL	Academic Word List
BEC	Business English Certificate
BNC	British National Corpus
BULATS	Business Language Testing Service
CAF	complexity, accuracy, fluency
CEFR	Common European Framework of Reference
CLC	Cambridge Learner Corpus
CLIL	Content and Language Integrated Language
COCA	Corpus of Contemporary American English
CPE	The Certificate of Proficiency in English
DP	Diploma Programme
EFL	English as a foreign language
EMI	English as a medium of instruction
ESP	English for specific purposes
EVP	English Vocabulary Profile
IB	International Baccalaureate
L1	first language
L2	second language
LCCI	London Chamber of Commerce and Industry
SD	standard deviation
SERR	Společný Evropský Referenční Rámec
SLA	Second Language Acquisition
TTR	Type Token Ratio

1 Introduction

My personal motivation to commence this research in the field of lexical complexity in business/economics written English started more than a year ago when I had succeeded in obtaining a teaching post at a Czech grammar school. During my economics classes, which are taught in English as a part of the International Baccalaureate (IB) Diploma Programme (DP) for students aged 18–19, I wondered many times how high the lexical written competence of these students is. The students are required to write 3 essay-type assignments during their 2-year course in economics: the first one about Microeconomics, the second one about Macroeconomics and the last one about The Global Economy. These assignments are all based on English newspaper extracts from reliable economics sources like The Economist or The Wall Street Journal. Having two texts with the same economic topic – an essay "commentary" by L1 Czech student (hereinafter "L2 text") and a newspaper article written by a native English speaker (hereinafter "L1 text") [for concrete samples please see appendices 1-4] – I also started to ask myself whether student written sophistication with the highest C2 level is still lower than that of a native speaker.

Apart from my internal motives, there are some external aims of this thesis. Firstly, the thesis could help to understand the differences between L2 proficient English learners and native speakers in the area of lexical complexity. Secondly, the thesis will compare and discuss the results with the current research in this field.

As the title of this thesis reads *Examining lexical complexity in the written production of L2 proficient learners of English*, it is important to answer two important questions in the beginning:

- Who is a proficient learner of English?
- What is exactly lexical complexity?

In this thesis, the proficient learner of English is understood as someone who is not a native speaker nor a simultaneous bilingual person, who started learning English as a second L1 in babyhood, and will be referred to mostly as an L2 English learner or student or user or speaker. The level of proficiency of L2 English users will be defined according to the Common European Framework of Reference (Council of Europe 2001) as this is the most used

proficiency scale in the European context. The CEFR scale recognises seven language levels (pre-A1, A1, A2, B1, B2, C1 and C2) and the highest level C2 will be considered as our proficient learner of English. The supreme level of English proficiency in CEFR is described in two possible ways: by means of "can-do statements", for instance, "I can write summaries and reviews of professional or literary works" (Council of Europe 2001: 27), and by an illustrative set of descriptors. The first approach is more intended towards prospective applicants for Cambridge courses and examinations, the second helps interlocutors to evaluate the ability of examinees, for example, to "backtrack and restructure around a difficulty so smoothly the interlocutor is hardly aware of it" (Council of Europe 2001: 37).

L2 proficient English speakers are sometimes compared to native speakers as having near native-speaker language competence: "the degree of precision, appropriateness and ease with the language which typifies the speech of those who have been highly successful learners" (Council of Europe 2001: 37). Interestingly, this mentioned "degree of precision, appropriateness and ease" is approximately the same as accuracy (~ precision), complexity (~ appropriateness) and fluency (~ ease) in the CAF model, which will help to answer the second question about lexical complexity.

The CAF model, which stands for complexity, accuracy and fluency, has been a reliable framework for language assessment since the 1970s. It is used to measure learners´ competence of both L2 and L1 speakers. Complexity has been defined as "the ability to use a wide and varied range of sophisticated structures and vocabulary", accuracy as "the ability to produce target-like and error-free language" and fluency as "the ability to speak with native-like rapidity, without over-use of pause, hesitation, or reformulation" (Housen et al. 2012). Studies of L2 ability are typically directed at one or more aspects of CAF. The research in this thesis focuses only on the complexity aspect, specifically on lexical complexity (syntactic, morphological and phonological complexity being other subcategories of complexity [Boulté and Housen 2012: 23]).

Lexical complexity will be understood in this work as consisting of two subsystems which are lexical diversity and lexical sophistication. Bulté and Housen (2012: 28) add other subsystems: lexical density, measuring the ratio of lexical words and function words, and lexical compositionality, measuring syllables and words in texts (see figure 3). In this thesis, lexical diversity will be initially measured by single indicators like the number of types, the number of tokens, TTR, and by two more precise composite indicators VOCD-D and MTLD.

Lexical diversity represents the word variation inside a text (parameter of “breadth”) whereas lexical sophistication explores low-frequent words in a text (parameter of “depth”). Lexical sophistication will be analysed by comparing the words in sample texts to lexical frequency lists.

The thesis structure is as follows: chapter 2 states hypotheses and the rest of the chapter is devoted to theoretical aspects concerning the definition of language proficiency according to CEFR and lexical complexity in their two dimensions of lexical diversity and sophistication. This chapter includes an overview of current research on lexical complexity and a detailed description of VOCD-D and MTLD composite indicators. Chapter 3.1 focuses on the research data (sample L2 texts and L1 texts can be seen in appendices), chapter 3.2 on the description of applied tools for analyses. Chapter 3.3 introduces all results with the following discussion on research limitations and prospective follow-up research. Conclusions, bibliography, Czech resumé and appendices can be found in chapters 4 to 11, respectively.

2 Theoretical Part

2.1 Thesis Aim and Hypotheses

The thesis focuses on lexical complexity in the written production of L2 proficient learners of English, specifically in two dimensions: lexical diversity and lexical sophistication (by the means of lexical frequency wordlists). The ambition is to increase the knowledge of written L2 complexity and to complement existing research in this area. The results of previous studies are presented in respective chapters dealing with lexical diversity (chapter 2.4.2) and lexical sophistication (chapter 2.4.3).

Several hypotheses will be investigated, supposing that lexical complexity is measurable by automated online tools. For the purposes of this thesis, the online website tool Text Inspector will be utilised. The Text Inspector is described in detail in chapter 3.2.1.

Foster and Tavakoli (2009) claim that there is no difference in lexical diversity between native and non-native speakers (chapter 2.4.2). In this thesis, it is presumed that L2 proficient learners of English have lower lexical complexity than native speakers. Another hypothesis is that the results in two distinctive groups (L2 and L1) will be similar (Foster and Tavakoli 2019, Palfreyman 2019).

2.2 CEFR and Language Proficiency

Modern proficiency scales for L2 learners, such as CEFR (Common European Framework of Reference), ILR (Interagency Language Roundtable scale) or ACTFL (American Council on the Teaching of Foreign Languages), usually take into account the use of language and communicative competences. Many labels for different proficiency groups are used across these proficiency scales which can lead to general confusion (the names of the highest levels: CEFR – “proficiency”, ILR – “native or bilingual”, ACTFL – “superior”). In this thesis, the CEFR scale is used throughout as it is widely used in the European context. Its scale and the comparison with IELTS (International English Language Testing System) and Business Cambridge Exams can be seen in Figure 1.

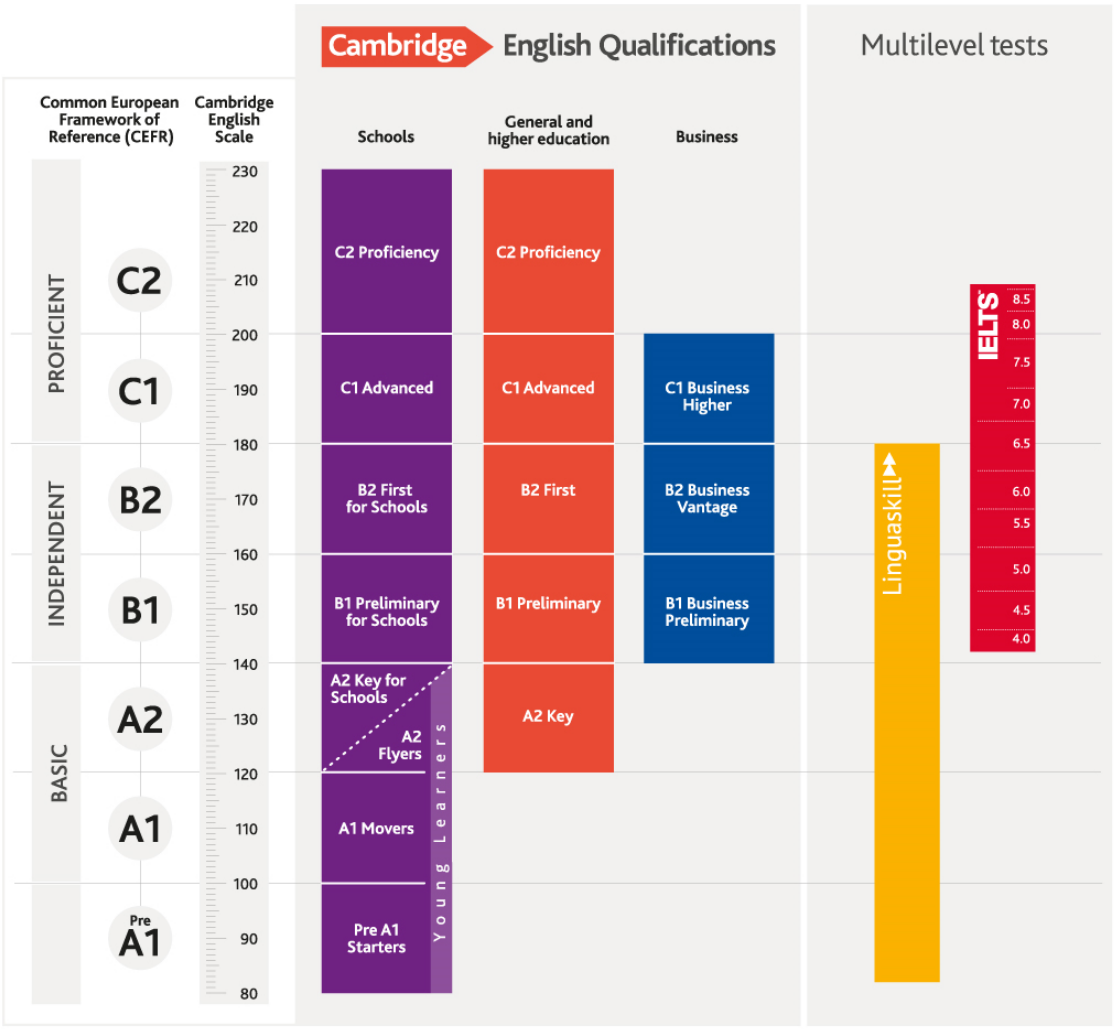


Figure 1: English Exams on the CEFR
(adapted from https://www.cambridgeenglish.org/exams-and-tests/cefr/)

The CEFR scale is comprised of seven language levels (pre-A1, A1, A2, B1, B2, C1 and C2) where the last two are considered as proficient user competence levels. The language skills competences are differentiated for the purpose of self-evaluation. This self-assessment grid implements the so-called "can-do statement approach". A general overview of the self-assessment grid for all CEFR levels is presented in table 1.

Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Table 1: Common Reference Levels: global scale (Council of Europe 2001: 24)

The writing competence (from the skills-set of reading, listening, speaking and writing) at the level of C2 is described as:

"I can write clear, smoothly flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure that helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works." (Council of Europe 2001: 27).

Apart from the self-assessment grid, the Council of Europe also employs concrete descriptors for measuring different levels of CEFR. The concrete illustrative set of descriptors used for the highest C2 level of CEFR is described as follows (Council of Europe 2001: 37):

- convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices,
- has a good command of idiomatic expressions and colloquialisms with awareness of the connotative level of meaning,
- backtrack and restructure around a difficulty so smoothly the interlocutor is hardly aware of it.

Nevertheless, it is highlighted in the document that these descriptors are only illustrative as can be seen in figure 2 below.

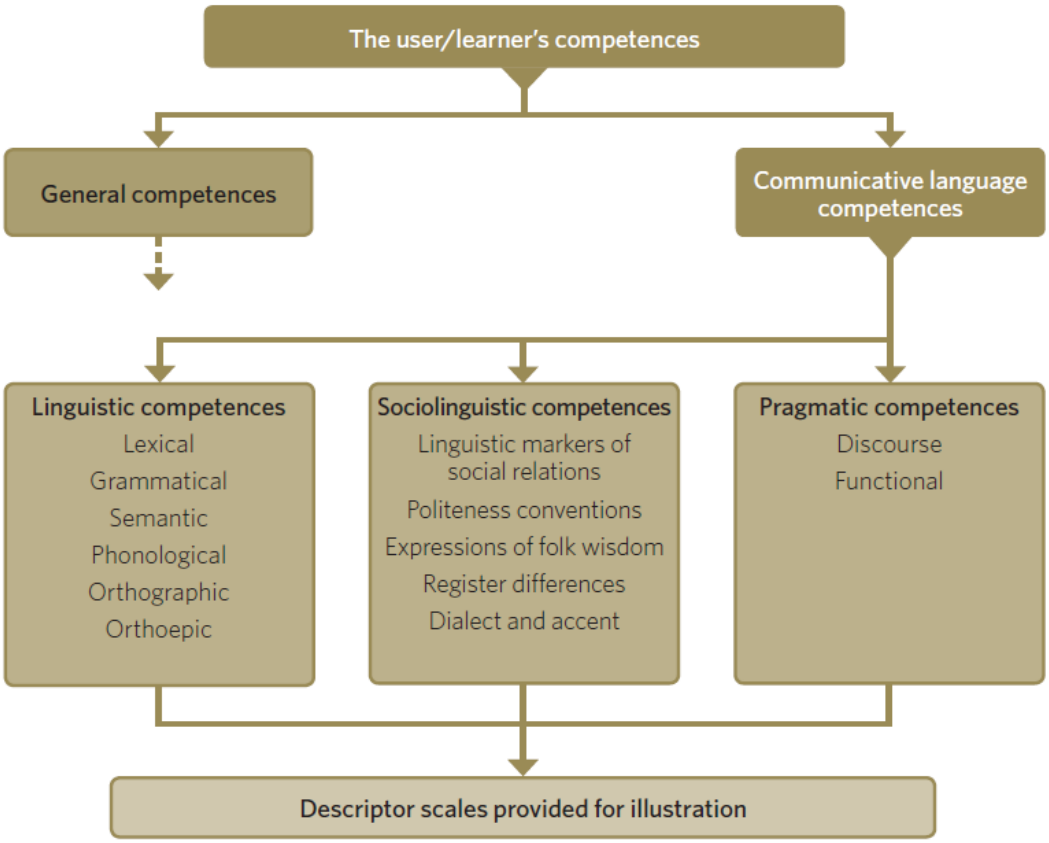


Figure 2: The user/learner´s competences
(adapted from https://www.cambridgeenglish.org/exams-and-tests/cefr/)

Lexical competence, which is a subsystem of linguistic competences (see figure 2), is described in the Council of Europe´s document *Common European Framework of Reference for Languages. Learning, Teaching, Assessment* (2001: 110 – 112) as the knowledge of, and ability to use, the vocabulary of a language; consists of lexical elements and grammatical elements. Lexical elements are divided into fixed expressions (sentential formulae – *How do*

you do?, phrasal idioms – *He kicked the bucket.*, fixed collocations – *make a mistake*, fixed frames – *Please may I have ...?*, and other fixed phrases) and single word forms (open and closed classes). Grammatical elements are connected with their word classes.

Illustrative scales are available for the range of vocabulary knowledge, and the ability to control that knowledge. Vocabulary range at the C2 level of CEFR is described as follows: "Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning" (Council of Europe 2001: 112). Vocabulary control assumes that a C2 user/learner consistently uses correct and appropriate vocabulary (ibid.: 112), which is close to the dimension of accuracy and complexity in the CAF model (chapter 2.4.1).

The Council of Europe´s document *Common European Framework of Reference for Languages. Learning, Teaching, Assessment* tries to cope with the comparison of native speakers and C2 English speakers too. "Level C2, whilst it has been termed 'Mastery', is not intended to imply native-speaker or near native-speaker competence. What is intended is to characterise the degree of precision, appropriateness and ease with the language which typifies the speech of those who have been highly successful learners." (Council of Europe 2001: 37).

According to Huang et al. (2018: 2), "the research on the application of descriptors in CEFR in English speaking assessment has been lacking". It is not the aim of the thesis to increase the research in this field but for understanding the differences between L2 proficient English users/learners and native speakers the knowledge of CEFR levels is crucial.

2.3 Business English

Business English is a type of English for Specific Purposes (ESP) and is intended to enhance the English learner competence in the realm of business communication.

The language competence in business English can be tested by internationally recognised tests at different CEFR levels, for instance by the Cambridge Business English Certificates (BEC) or the Pearson LCCI International Qualifications in English for Business. It seems that the Cambridge Business English Certificates are going to be the standard in future for business language proficiency testing as the Pearson LCCI International Qualifications in English for

Business are discontinued as of 31 December 2021. Similarly, the Business Language Testing Service, known as BULATS, was ended in the year 2019.

The so-called BECs are offered at three different CEFR levels – B1 (Business Preliminary Qualification), B2 (Business Vantage Qualification) and C1 (Business Higher Qualification). The highest C2-level examination is restricted only for general English, business English certificates are attainable only up to level C1 (see figure 1).

The BEC examination focuses on all language competencies – reading, listening, writing and speaking. The main difference between general English examinations and BEC examinations is that BEC exams are primarily based on everyday work and business tasks (https://www.cambridgeenglish.org). Examination tasks are thus from a business context and there is a strong focus on testing specific lexis.

2.4 Lexical Complexity

Bulté and Housen (2012: 28) argued that "complexity has rarely been adequately defined in the [CAF] literature". Nevertheless, they perceive the notion of complexity in two dimensions: grammatical complexity and lexical complexity. According to them, two major sources of grammatical complexity can be distinguished: syntactic and morphological complexity (ibid.: 27). Lexical complexity can be measured through lexical density, diversity, compositionality and sophistication (ibid.: 28). Bulté and Housen´s notion (2012) of lexical complexity can be seen below.

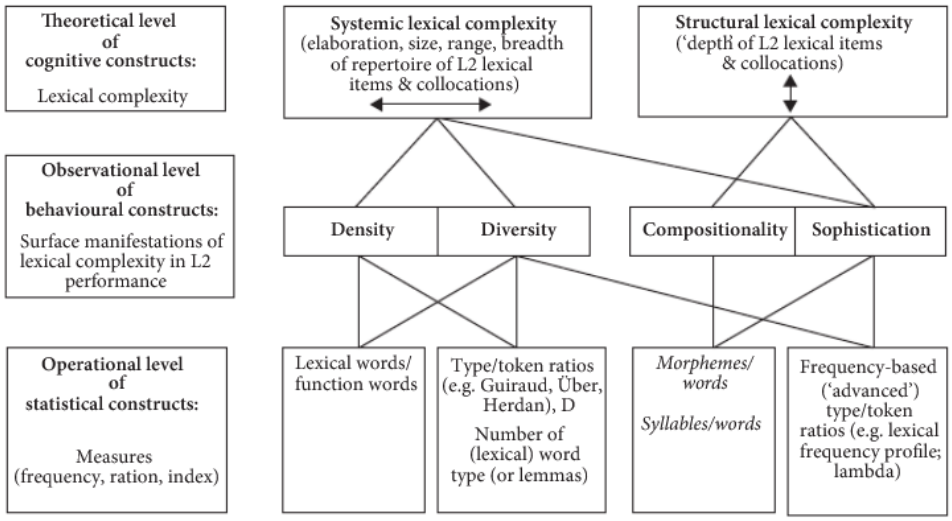


Figure 3: Lexical complexity (Bulté and Housen 2012: 28)

Other available definitions of complexity range from "grammatical and lexical complexity mean that a wide variety of both basic and sophisticated structures and words are available to the learner" (Wolfe-Quintero, Inagaki and Kim 1998) to "complexity is the extent to which learners produce elaborated language" (Ellis and Barkhuizen 2005: 139).

Palloti (2015) declares that there are three elements of complexity: structural complexity, cognitive complexity and developmental complexity. Structural complexity refers to the number of linguistic items and their relational patterns, cognitive complexity is connected with processing costs and developmental complexity deals with "the order in which linguistic structures emerge and are mastered" (ibid.). Palloti (2015: 125) argues that "a text with a wide variety of lexemes will be said to be more complex than one where the same few words are repeated over and over". Lexical complexity is thus, according to Palloti, more connected with lexical diversity.

In current SLA research, lexical complexity can function either as an independent variable or a dependent variable (depending for instance on age, sex or the type of instruction). The studies show that L2 learners´ complexity improves as the general language proficiency increases in time (Bulté and Housen 2014, Durán et al. 2004).

2.4.1 The CAF Model

Housen et al. (2012) suggest that language proficiency and performance is inherently multi-dimensional. It is the CAF model which meets this view. The CAF triad (complexity, accuracy and fluency) emerged from Brumfit´s (1984) two-dimensional teaching model distinguishing separate accuracy-focused and fluency-focused teaching activities. Later on, complexity was added to the triad, as the last dimension, by Skehan in the 1990s. Housen et al. (2012) have defined the CAF triad in the following way: complexity refers to the ability to use a wide and varied range of sophisticated structures and vocabulary, accuracy is the ability to produce target-like and error-free language and finally, fluency depicts the ability to speak with native-like rapidity, without over-use of pause, hesitation, or reformulation.

Although studies on this topic are frequent, the problem is that the vast number of measures employed for each dimension of the CAF triad decreases the comparability of findings. Another problem of the CAF model is that a speaker can produce a written text with a high score in complexity, accuracy and fluency but without "putting the message across" (Pallotti 2009). This communicative inadequacy or inappropriateness is not measurable by the CAF model.

2.4.2 Lexical Diversity – Concept and Studies

Lexical diversity is a measurement of how many different lexical words there are in a text, it is a text-internal metric. Lexical words are words such as nouns, adjectives, verbs, and adverbs that convey meaning in a text, for example, 'dog', 'blue', 'run' and 'usually'. These are different from grammatical words, like articles, pronouns, and conjunctions, that hold the text together and show relationships, for example, 'the', 'her' and 'or' (Quirk 1985: 72; Bax 2021). Lexical diversity focuses on lexical words whereas grammatical words are not considered.

Another view on lexical diversity is expressed by Duran et al. (2004a): "…lexical diversity is about more than vocabulary range. Alternative terms, 'flexibility', 'vocabulary richness', 'verbal creativity', or 'lexical range and balance' indicate that it has to do with how vocabulary is deployed as well as how large the vocabulary might be".

The writing skill of L2 English learners belongs to productive skills, next to a speaking skill. On the other hand, reading and listening skills form receptive language skills. Milton states (2009: 117) that the productive vocabulary of L2 learners tends to be smaller than receptive vocabulary and that there is a higher chance that low-frequency words would be recognised than actively produced. The comparison of low-frequency words between texts with productive and receptive lexis was not analysed in this thesis as its focus was only on English orthography.

The type/token ratio (TTR) is used as the simplest method how to measure lexical diversity. The TTR is "the total number of different words used in the text, which are referred to as types, divided by the overall number of words in the piece of discourse, labelled as tokens" (Czwenar 2013: 83). In other words, the TTR is the indicator of word repetition in the text. The advantage of TTR is its simplicity in its calculation (many automated tools can be used in this way), but it is criticised for its sensitivity to text length (ibid; Kuiken and Vedder 2012). In consequence, many composite indicators for measuring lexical diversity have been developed to compensate for the weakness of TTR. In this thesis, two composite indicators are employed – VOCD-D and MTLD. These indicators are more reliable than a single TTR indicator and their details are described in the next two chapters.

2.4.2.1 VOCD-D

Duran et al. (2004a: 239) have developed an advanced VOCD application with an index D as an indicator of lexical diversity and have argued that "D offers a robust metric of 'vocabulary range and balance' for research and for application where quantification of lexical diversity is required." The formula which is used in VOCD application with a D parameter is:

$$TTR = \frac{D}{N}\left[\left(1 + 2\frac{N}{D}\right)^{\frac{1}{2}} - 1\right]$$

, where TTR stands for type/token ratio and N is the number of tokens (ibid. 224).

The larger D, the greater the lexical diversity is. VOCD presents a "valid index, D, which can be calculated with good reliability even for shorter transcripts" Duran et al. (2004a: 237). The index D is determined from 1,600 text calculations. Initially, 100 trials of random samplings of 35 tokens from throughout the text are made and then their average TTR is

calculated. The same is repeated for 36 tokens and so on up to 50 tokens, which total 16 x 100 calculations (ibid. 224-225).

Duran et al. (2004a: 227-237) have elaborated a detailed analysis of carefully selected language samples. The result can be seen in Figure 4, where the y-axis represents various sample cohorts (months corresponds to the age of L1 speakers of English) and the x-axis shows the interval values of index D.

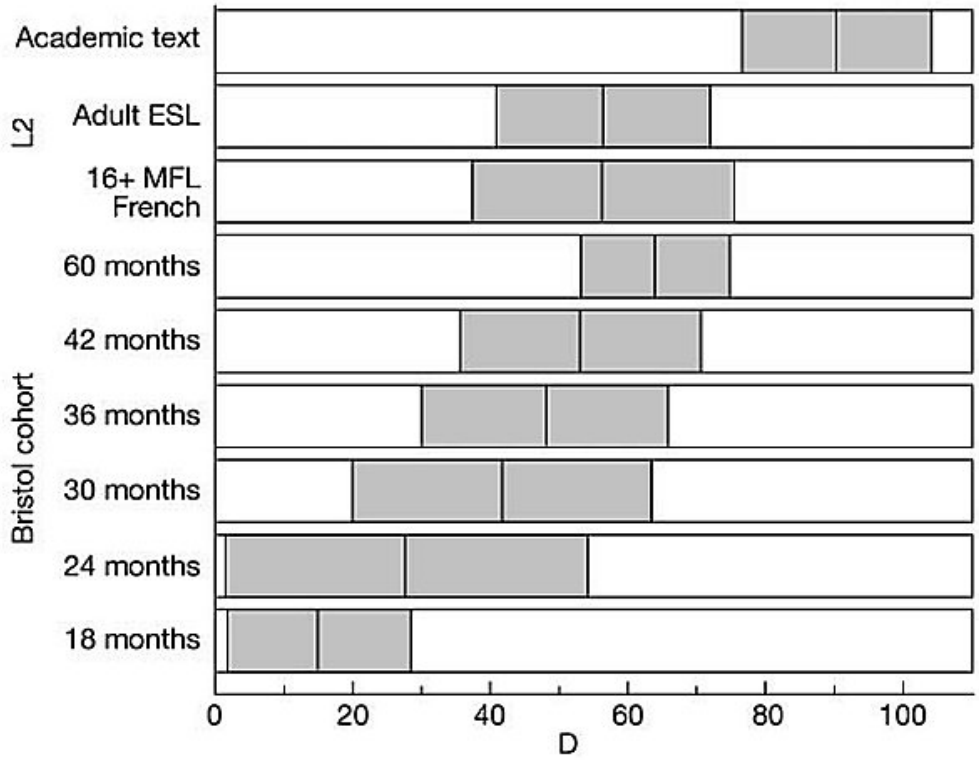


Figure 4: VOCD means and sub-ranges (10th – 90th percentiles) of D for various cohorts (Duran et al. 2004a: 238)

The two topmost cohorts are of utmost importance for the purpose of this thesis research as they depict the D-values for adults. The first is the adult second learner group (see *Adult ESL* in figure 4), where D is typically somewhere between 40–70. The second is the adult native speaker cohort writing an academic text (see *Academic text* in figure 4), where lexical diversity, represented by index D, typically have a measure of between 80–105.

2.4.2.2 MTLD

McCarthy et al. (2010) claim that the only index not found to vary as a function of text length is MTLD (Measure of Textual Lexical Diversity). This index is constructed as the quotient of the number of words divided by the total factor count. The final MTLD number is the mean of the two values, one for forwarding processing and one for reverse processing. It is a real number between 0 and 1. McCarthy et al. (2010: 5) state that "various studies have shown MTLD to be at least as effective as the industry standard vocd-D index, and even one of the most informative and distinguishing variables in the entire arsenal of several hundred Coh-Metrix indices".

According to Limbrick (2008: 13), "reliability of data can be affected by issues such as students' response to the atypical writing context of a formal assessment task, students' interest in the standard writing topic and text form of the task at any one assessment point, and student attitudes on the day". It can be inferred that not only language competencies are at stake but also some psychological constraints.

Another study by Foster and Tavakoli (2009) showed that there is no particular difference between native and non-native speaker groups in the variance in scores for lexical diversity when they were given the tasks to write different narrative stories based on given pictures. The learners, who were based in London and Tehran, "did not differ in their performance when compared to each other, except in lexical diversity, where the learners in London were close to native-speaker levels".

2.4.3 Lexical Sophistication – Concept and Studies

Text lexical complexity is not only about lexical diversity, but it also depends on other factors including how these lexical words are used, which can be represented by lexical sophistication. Lexical sophistication, as opposed to lexical diversity, is the type of text-external metrics. It measures the depth of the text by comparing it to external frequency word lists like EVP (English Vocabulary Profile), AWL (Academic Word List), BNC (British National Corpus) or COCA (Corpus of Contemporary American English). In this thesis, an EVP frequency word list is used, and lexical sophistication is measured by the number of low-

frequency words in texts (concrete reasons for this option are provided in chapter 3.2.1.3 English Vocabulary Profile).

Milton (2009: 137) claims that measures of lexical sophistication are problematic since they are sensitive to the genre, register and even individual stylistic choices of L2 English users. Lexical sophistication would be probably higher in written texts than in oral production as they usually contain more infrequent lexis (ibid.). He argues that despite these shortcomings the measures of lexical sophistication are still useful within genre boundaries for measuring lexical text 'quality'. The study of Yoon and Polio (2017: 288) found that the genre of written texts influences word frequencies: argumentative essays contained more low-frequency words than narrative texts.

Palfreyman (2019) discovers in his study that undergraduate student writers (with L1 Arabic and L2 English) have significant differences in their students' use of vocabulary from different frequency bands of vocabulary in their L1 writing and their L2 writing. He also claimed that "there is the possibility of a positive correlation between lexical sophistication in the two languages".

In another study, Tabari et al. (2021) found that task sequencing on L2 written production plays a significant role in the increase in syntactic complexity, accuracy, lexical complexity, and fluency. The research was randomly divided into two groups of participants: one performing three sequenced tasks from simple to complex sequence, the other writing only one task (either simple, medium or complex).

Based on previous aforementioned studies and research, the following research questions are considered:

- Is lexical diversity of English native speakers higher than that of advanced non-native speakers of English (with L1 Czech)?
- Is lexical sophistication of English native speakers higher than that of advanced non-native speakers of English (with L1 Czech)?

3 Empirical Part

3.1 Data

3.1.1 Data Selection

In order to answer the research questions, two small corpora of written English will be investigated and compared. The data consists of 20 written samples of L2 adult students of English (L2 texts) and 20 written samples of L1 English journalists (L1 texts), which altogether makes the set of 40 texts.

The L2 students were from Septima and Oktava classes of the International Baccalaureate Diploma Programme where the whole programme is taught fully in English through the method of EMI (English as a Medium of Instruction). The advantage of the student´s group is their homogeneity in the dimension of age (18 to 19 years-old students), and the dimension of their English proficiency level. According to CEFR, the students´ language proficiency level is C2 (they all successfully passed the CPE examination). The written student samples are essays from the subject of economics. The assignment instructed students to write up to 800-word commentaries based on a chosen article from the English news media to apply their knowledge of economics and to demonstrate the ability to analyse and evaluate scientific text with the help of economic diagrams from economics. In this way, all students´ written commentaries are of a similar aim, extent, and genre. Two examples of students´ written work can be seen in appendices 1 and 3.

According to Diploma Programme Economics Guide (2020: 67), the articles (L1 texts) may be from a newspaper, a journal or the Internet, but must not be from television or radio broadcast. They were published in a wide range of newspapers and magazines, for instance, The Financial Times, The Economist, The Wall Street Journal, and are not restricted to their length: "if a student includes a relatively lengthy article, which is discouraged, the student must highlight the section(s) of the article upon which the commentary is based" (DP Economics Guide 2020: 67). All articles are written by native or near-native adult English speakers.

The mean of **L1 text tokens** reached the mean value of 654 with a minimum value being 279 tokens and a maximum value being 1,240 tokens ($\bar{x}$ = 654, min = 279, max = 1,240, SD =

320.4). All L1 texts are comparable to their extent (shorter texts), genre (economics) and aim (infotainment). Two examples of L1 texts can be found in appendices 2 and 4.

The mean of **L2 text tokens** reached the mean value of 781 with a minimum value being 722 tokens and a maximum value being 886 tokens ($\bar{x}$ = 781, min = 722, max = 866, SD = 37.4). All L2 text topics are from the realm of economics (microeconomics, macroeconomics, or the global economy). They are comparable to their extent (shorter texts), genre (economics) and aim (commentary), as are L1 texts. Two L2 text examples can be found in appendices 1 and 3. The next table shows the overview of L1 and L2 text tokens.

L1 text	1A	2A	3A	4A	5A	6A	7A	8A	9A	10A	articles
tokens	645	932	607	438	471	520	1,153	660	466	634	
L1 text	11A	12A	13A	14A	15A	16A	17A	18A	19A	20A	
tokens	279	599	486	377	1,218	1,240	327	401	1,235	391	
L2 text	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C	comment.
tokens	768	722	754	744	745	760	763	781	767	797	
L2 text	11C	12C	13C	14C	15C	16C	17C	18C	19C	20C	
tokens	803	845	806	795	770	886	763	807	754	783	

Table 2: Number of tokens in L1 and L2 texts

Both groups of written samples (L1 and L2 texts) are homogenous by their extent (the mean of L1 texts – 654 tokens, the mean of L2 texts – 781 tokens), genre (economics), and language (English), but heterogeneous by their authorship (L2 English students with L1 Czech vs. L1 English adults) and by their aim (internal assessment written for examiners vs. news-media article written for readers). As the basic features of these 40 texts are similar, the same analytical methods could be applied.

3.1.2 Data Processing

All sample texts had to be manually prepared for their use in the Text Inspector (detailed description in chapter 3.2.1) as the software works only with plain text. The preparation of texts in steps was as follows:

1/ A cover sheet for the assignment was erased.

2/ All unnecessary and unrelated words were deleted, like ‘article:’ or ‘commentary:’.

3/ All figures (diagrams, tables) were eliminated.

4/ A bibliography section was deleted.

5/ All footnotes were deleted and not considered for analysis.

6/ Finally, all redundant spaces between paragraphs were deleted.

Having eliminated all unnecessary objects in a text, the plain text was copied to the Text Inspector and the thorough analysis started.

Data analysis was performed in a three-step approach. First, the L2 English student´s texts were analysed. In the next step, the L1 journalist´s texts were explored and finally, the comparison of both these text groups was applied.

A detailed analysis of the data was carried out. Both the basic statistical analysis and the in-depth quantitative analysis were conducted. The basic quantitative analysis included statistical indicators such as a number of sentences, a number of tokens, a number of types, an average sentence length, a type/token ratio, and a number count. The in-depth analysis focused on lexical diversity, which was measured through VOCD-D and MTLD complex indicators, and on lexical frequencies, for which the EVP Text Inspector tool was applied. All data was manually typed into an Excel spreadsheet where the data was analysed, and the appropriate figures and tables were created.

3.2 Methodology

3.2.1 The Text Inspector

Text Inspector is one of many possible online website tools which can be used for measuring lexical complexity. It presents itself as an “award-winning[1] professional text analysis tool for teachers, examiners, academics and you” (Bax 2021).

Text Inspector is available on https://textinspector.com/ and is free of charge for the analysis of up to 250-word long texts. In the case of longer texts, there are paid subscription

[1] the winner of the 2017 British Council ELTons Digital Innovation Award

plans on offer. The advantage of Text Inspector is that it can be used by a wide range of researchers as the software is free (or for the small subscription fee for longer texts) and easily accessible on the Internet and thus the replicability of research is easier than in the case of other not so easily accessible tools.

The Text Inspector software, in its paid version, was used for lexical complexity analysis of the 40 written texts in this thesis.

The web tool enables a multitude of online tools:

- Statistics and Readability
- Lexical Diversity
- British National Corpus (BNC)
- Corpus of the Contemporary American English (COCA)
- Academic Word List (AWL)
- English Vocabulary Profile (EVP)
- Scorecard
- Parts of Speech Tagger
- Metadiscourse Markers
- Errors.

Only some Text Inspector tools were utilised considering the aims of the thesis. For this reason, the following description only focuses on the tools which were used.

3.2.1.1 Statistics and Readability

The "Statistics and Readability" function offers an overview of the language level and complexity of a text. It offers a basic quantitative analysis of the text, like a number of sentences, a number of tokens, a number of types, a syllable count, a type/token ratio (TTR), an average sentence length in words, a number count, a number/percentage of words with more than two syllables, a number of average syllables per word/per sentence/per hundred words. Except for these measures, readability scores are presented, namely a Flesch Readability Ease, a Flesch-Kincaid Grade and a Gunning Fog Index.

Some basic quantitative analysis using this tool is implemented in the thesis.

3.2.1.2 Lexical Diversity

In the Text Inspector, lexical diversity is represented by two measures: VOCD-D and MTLD. Detailed information regarding the definition and calculation of these measures can be found in chapters 2.4.2.1 and 2.4.2.2.

In the thesis, both VOCD-D and MTLD measures for calculating the lexical diversity of texts are used.

3.2.1.3 English Vocabulary Profile

The English Vocabulary Profile (EVP) function analyses the text according to the EVP from Cambridge University Press[2]. It uses the Cambridge Learner Corpus (CLC); a collection of hundreds of thousands of examination scripts from English language learners across the world to help analyse texts in terms of their CEFR level. "The English Vocabulary Profile offers reliable information about which words (and importantly, which meanings of those words) and phrases are known and used by learners at each level of CEFR" (English Vocabulary Profile 2015). Lexical frequency is then displayed in a table and a diagram format.

The Text Inspector also offers another tool for measuring text lexical sophistication – Academic Word List (AWL). In considering which tool is more appropriate for the thesis, the following arguments were taken into consideration:

1. AWL is categorised into ten sublists according to their frequency; in comparison, EVP is closely connected to the CEFR levels (A1-C2) and in this way more suitable for analysing C2 level proficiency of English learners
2. AWL contains only over 3000 words occurring frequently in academic texts; EVP is based on extensive Cambridge corpora: Cambridge Learner Corpus – 48 million words, which includes student writing at all six CEFR levels, and Cambridge International Corpus – a 1.2-billion-word collection of written and spoken English

[2] EVP is the part of English Profile Project. Another part is English Grammar Profile. Available on http://www.englishprofile.org/.

3. The EVP tool enables the comparison with Rálišová's research (2020). The AWL tool was not considered in her work either (ibid).

3.2.1.4 Other Text Inspector tools

As other Text Inspector online tools were not used in this thesis, they are only briefly described.

The British National Corpus (BNC) tool provides word frequencies that are based on the British National Corpus. Data are available both in a table and a diagram format.

Similarly, the Corpus of Contemporary American English (COCA) tool provides word frequencies that are based on the Corpus of Contemporary American English. Data are again available both in the form of a table and a diagram.

The Academic Word List (AWL) tool analyses a text according to the AWL which contains over 3,000 items (words that occurs frequently in academic texts). The AWL consists of ten sublists which are organised according to their frequency. The displayed list of lexical frequency is in a table and a diagram variant.

The "Scorecard" tool classifies texts according to a CEFR level. It provides a general idea of the learner´s level of English.

The "Parts of Speech Tagger" identifies word classes, such as nouns, verbs, or adjectives. The tool is based on a modified version of TreeTagger which was developed by Helmut Schmid at the Institute for Computational Linguistics of the University of Stuttgart (Bax 2021).

The "Metadiscourse Markers" are "also known as 'transitions'; these are words and phrases such as 'firstly' and 'in conclusion' that add extra information to a text" (Bax 2021).

The "Errors" tool identifies spelling errors in the sample.

3.2.2 Method

As the first representation of the analysed texts, the following indicators would be used: the number of sentences, the number of tokens, the number of types, the average sentence length in words, the type/token ratio, and the number count. The number count was specifically

included as it was expected that a high proportion of numbers would be represented in economic texts. The statistics will be calculated by the Text Inspector "Statistics and Readability" tool, the data will be exported and adjusted in MS Excel software.

The lexical complexity of the whole corpus (20 L1 texts and 20 L2 texts) will be analysed in two dimensions, in the dimension of lexical diversity and the dimension of lexical sophistication. Lexical diversity will be measured by two independent measures – VOCD-D and MTLD. By using two independent indices of lexical diversity in the thesis, instead of only one, the reliability of results will be increased. Furthermore, the reliability of results will be improved by three independent measurements in the case of the VOCD indicator. But still, there are no perfect measures of lexical diversity (see chapter 3.3.3). Lexical sophistication will be measured by the number of low-frequency words using the EVP frequency word list. The reasons for choosing the EVP tool are discussed in chapter 3.2.1.3.

3.3 Results

3.3.1 Quantitative analysis

The next two tables (tables 3 and 4) show relevant statistical data for the whole sample of 40 texts. The first table contains the statistics for 20 commentaries (1C – 20C), which were produced by L2 English learners (L2 texts), and the second table provides the statistics for 20 articles (1A – 20A), which were written by L1 English adults (L1 texts).

(1) Commentary	(2) No. of sentences	(3) No. of tokens	(4) No. of types	(5) Average sentence length	(6) Type/token ratio	(7) Number count
1C	27	768	342	28.44	0.45	6
2C	35	722	253	20.63	0.35	9
3C	32	754	241	23.56	0.32	12
4C	38	744	232	19.58	0.31	2
5C	29	745	335	25.69	0.45	6
6C	27	760	298	28.15	0.39	3
7C	35	763	310	21.80	0.40	18

(1) Commentary	(2) No. of sentences	(3) No. of tokens	(4) No. of types	(5) Average sentence length	(6) Type/token ratio	(7) Number count
8C	32	781	373	24.41	0.48	11
9C	20	767	355	38.35	0.46	4
10C	30	797	256	26.57	0.32	6
11C	33	803	263	24.33	0.33	12
12C	44	845	295	19.20	0.35	1
13C	27	806	272	29.85	0.34	10
14C	34	795	307	23.38	0.39	9
15C	28	770	326	27.50	0.42	8
16C	33	886	306	26.85	0.35	14
17C	37	763	313	20.62	0.41	2
18C	34	807	295	23.74	0.37	0
19C	39	754	283	19.33	0.38	4
20C	32	783	278	24.47	0.35	1
Mean	32.30	781	297	24.82	0.38	6.90
SD	5.27	37.4	38	4.51	0.05	4.93

Table 3: L2 text quantitative analysis

In the case of L2 texts (= commentaries), numerals, as a word class, were present in all texts, except for one. The mean is approximately 7 words, with a maximum of 18 words. Presumably, numerals in these texts are not often repeated and thus their impact is that they slightly decrease the value of TTR. The number of tokens across the whole sample is very similar, with a mean of 781 tokens, a minimum value of 722 tokens and a maximum value of 866 tokens ($\bar{x}$ = 781, min = 722, max = 866, SD = 37.4). The reason for such a similar length is the prescribed word count which was set to the students (to write a 800-word commentary). For the same reason, a very similar situation is to the number of sentences in a text ($\bar{x}$ = 32.3, min = 20, max = 44, SD = 5.3). Another interesting finding is that TTR is in the narrow range of 0.31 – 0.48, with a mean of 0.38.

(1) Article	(2) No. of sentences	(3) No. of tokens	(4) No. of types	(5) Average sentence length	(6) Type/token ratio	(7) Number count
1A	24	645	332	26.88	0.51	17
2A	38	932	407	24.53	0.44	5
3A	31	607	302	19.58	0.50	9
4A	21	438	256	20.86	0.58	13
5A	16	471	233	29.44	0.49	4
6A	18	520	263	28.89	0.51	8
7A	59	1153	539	19.54	0.47	13
8A	26	660	291	25.38	0.44	11
9A	16	466	249	29.12	0.53	13
10A	38	634	345	16.68	0.55	18
11A	10	279	136	27.90	0.49	22
12A	24	599	271	24.96	0.45	25
13A	25	486	255	19.44	0.53	6
14A	16	377	220	23.56	0.59	26
15A	55	1218	491	22.15	0.41	42
16A	51	1240	484	24.31	0.39	16
17A	11	327	165	29.73	0.50	1
18A	16	401	219	25.06	0.55	5
19A	59	1235	548	20.93	0.44	17
20A	13	391	239	30.08	0.61	19
Mean	28.35	654	312	24.86	0.50	14.5
SD	16.20	320	120	3.70	0.06	9.6

Table 4: L1 text quantitative analysis

In the case of L1 texts (= articles), numerals are represented even more than it was in the case of L2 texts ($\bar{x}$ = 14.5, min = 1, max = 42, SD = 9.6). Again, the high proportion of numerals in L2 texts (2.2 %) is given by the fact that the subject of these texts is economics. The length of the texts is more varied here ($\bar{x}$ = 654, min = 279, max = 1,240, SD = 320.3) as there was no limitation to the article choice; only a recommendation was given that the article should be shorter in its nature. As for the TTR, the values are higher than it was in the case of student´s texts ($\bar{x}$ = 0.5, min = 0.39, max = 0.61, SD = 0.06).

A detailed comparison of statistical data between L2 texts and L1 texts is presented in the next subchapters (all data is taken from table 3 and 4).

3.3.1.1 Number of Sentences

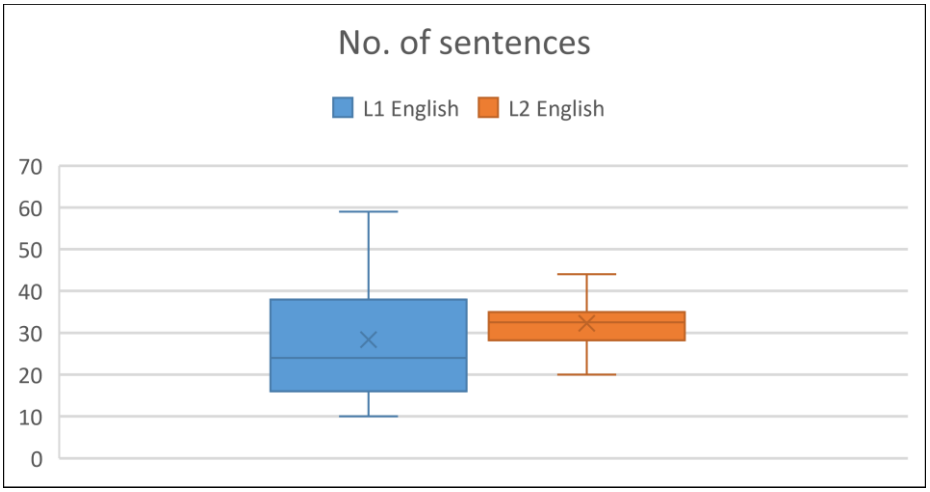


Figure 5: A comparison of sentence count between L1 and L2 texts

It could be inferred that the number of sentences in L1 texts is smaller than in L2 texts – it is true both for the mean and the median. The number of sentences in L2 texts ($\bar{x}$ = 32.3, min = 20, max = 44, SD = 5.3) is less varied because it was influenced by instructions for the assignment (a set word count).

A non-parametric statistical test, namely Wilcoxon Signed-Rank test was carried out. The test showed no significant difference between the two groups ($Z = -1.06$, $p > 0.05$). The results are, however, hardly surprising as the length of the assigned L2 tasks was restricted by one of the testing criteria.

3.3.1.2 Average sentence length

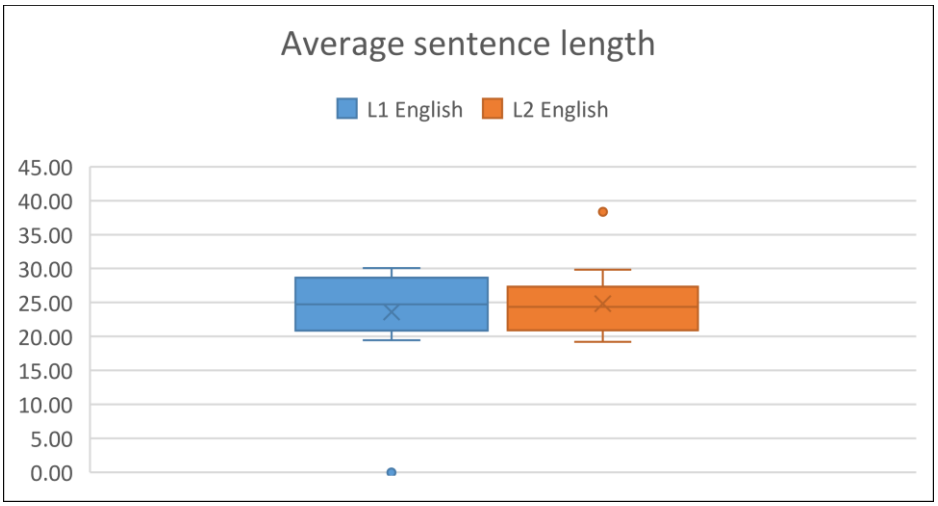


Figure 6: A comparison of average sentence length between L1 and L2 texts

If we see the average sentence length indicator for both groups of texts, the results are surprisingly very similar. The mean for L1 texts is 24.86 words per sentence and the mean for L2 texts is 24.82. Even if we investigate all other values, a standard deviation included, they confirm "the close similarity" of the two groups as for their average sentence length. The possible reason for this could be the students´ effort to imitate the original article in their commentaries.

The non-parametric statistical test, Wilcoxon Signed-Rank test showed that the difference is not statistically significant ($Z = -0.05$, $p > 0.05$).

3.3.1.3 Number of Tokens

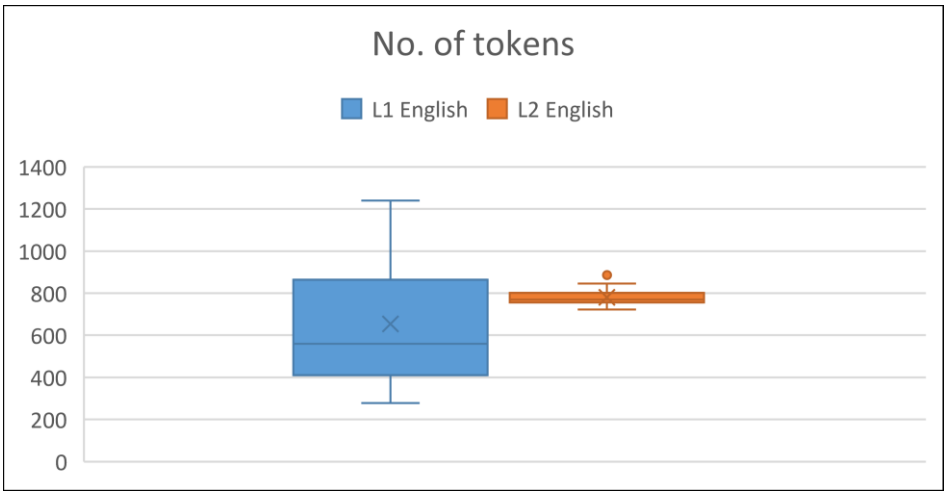


Figure 7: A comparison of token count between L1 and L2 texts

The number of tokens in L2 texts is extremely similar ($\bar{x}$ = 781, min = 722, max = 866, SD = 37.4). On the other hand, we can see much greater variation in L1 texts ($\bar{x}$ = 654, min = 279, max = 1,240, SD = 320.3). The reason for this is the same as was in the case of the number of sentences – the limitation of the assignment by a predetermined word count and an unlimited choice of articles regardless of their length.

Again, Wilcoxon Signed-Rank test showed that the difference is not large enough to be statistically significant ($Z = -1.38$, $p > 0.05$).

3.3.1.4 Number of Types

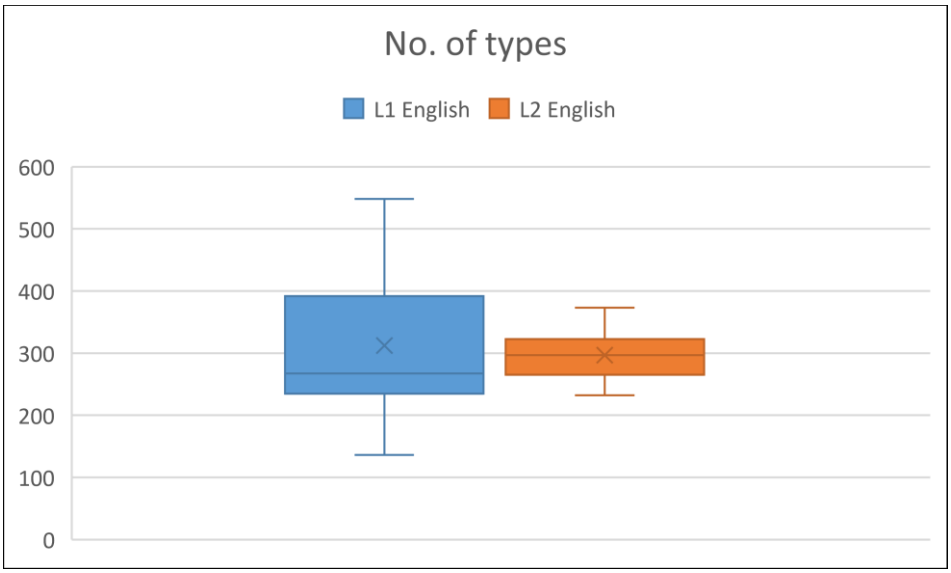


Figure 8: A comparison of type count between L1 and L2 texts

The number of tokens in the L1 texts was higher than that in the L2 texts (781 vs. 654) by 16%, therefore it could be expected that the number of types in L2 texts would be similarly higher. However, the opposite is true: L2 text values are ($\bar{x}$ = 297, min = 232, max = 373, SD = 37.9) and L1 text values are ($\bar{x}$ = 312, min = 136, max = 548, SD = 120.4). But if we analyse it closely, we can see that the median for L2 texts is higher than for L1 texts (297 vs. 267). This phenomenon has to be understood in the light of the fact that some L1 texts are longer, which affects the mean. The evidence can be seen in figure 8 and table 4.

Wilcoxon Signed-Rank test displayed that the difference is not large enough to be statistically significant ($Z = 0.22$, $p > 0.05$).

3.3.1.5 Type/Token Ratio

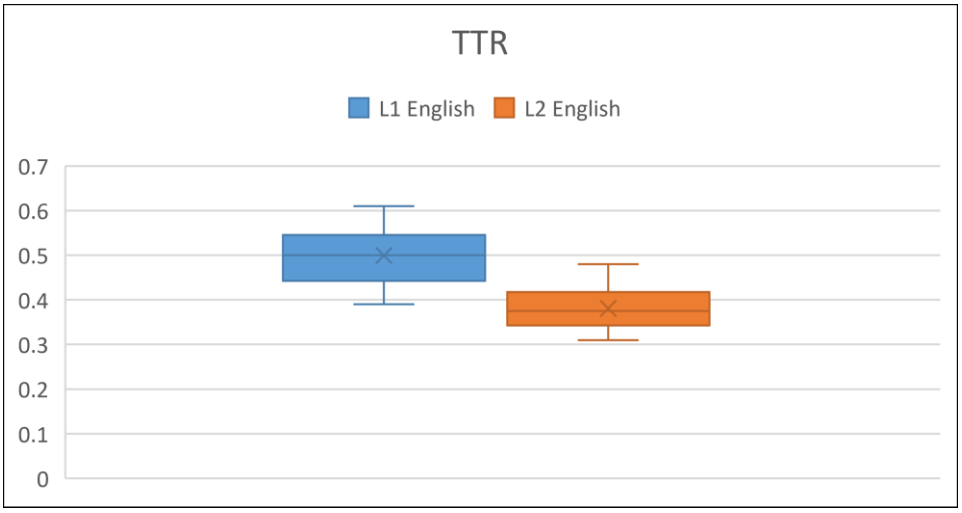


Figure 9: A comparison of TTR between L1 and L2 texts

The type/token ratio (TTR) synthesizes the previous two indicators into one single indicator. Due to its sensitivity to text length (Kuiken and Vedder: 2012), the L1 text TTR could be less reliable, as the text length more varies ($\bar{x}$ = 654, min = 279, max = 1,240, SD = 320.3), than for L2 texts, where the text length is much more homogenous ($\bar{x}$ = 781, min = 722, max = 866, SD = 37.4). The L2 text TTR is 0.38 whereas the L1 text TTR is 0.50. The possible interpretation is that L2 texts incorporate more word repetition than L1 texts. In L1 texts, on average each word is repeated twice, whereas in L2 texts it is nearly three times. We can deduce that students tend to use the same lexis more and they use fewer synonyms than their L1 English counterparts.

In the case of TTR, Wilcoxon Signed-Rank test showed that the difference is statistically significant (Z = 3.77, p = 0.00016, the effect size = 0.84). The results of better composite indicators, such as VOCD-D and MTLD, which are not sensitive to the text length, are presented in chapter 3.3.2.1.

3.3.1.6 Number Count

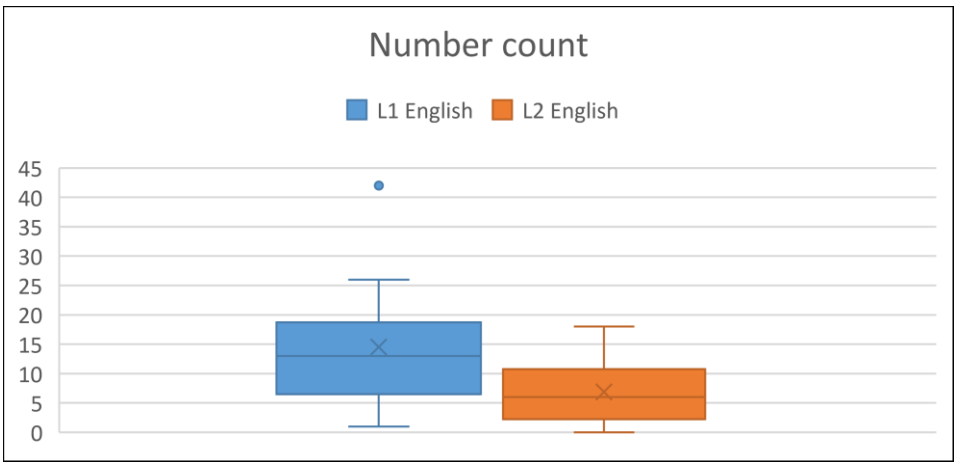


Figure 10: A comparison of number count between L1 and L2 texts

As the texts are of scientific nature from economics, they would presumably comprise some numerals. The proportion of numerals in analysed texts is vital as it has some impact on TTR values. We can expect that numerals would be in texts repeated only occasionally and that the numbers in the texts are in majority hapax legomena – occurring only once. The overall effect of hapax legomenon is that it decreases the value of TTR. The relative frequency of numbers in texts is very low though (0.8% in the case of L2 texts and 2.2% in the case of L1 texts).

Here, Wilcoxon Signed-Rank test displayed that the difference between the two samples is statistically significant ($Z = 2.73$, $p = 0.0061$, the effect size = 0.63).

3.3.2 Lexical complexity analysis

Lexical diversity and lexical sophistication will be analysed in this section.

3.3.2.1 Lexical diversity

The VOCD-D and MTLD indicators were selected for lexical diversity examination as these indicators are more reliable than a single TTR indicator.

In the case of VOCD-D, three different analyses were applied to decrease the weakness of sampling randomness producing each time slightly different figures. From these three independent measures, the average value was calculated as a final value for texts. Results and calculations can be seen in table 5.

A/C No	VOCD1	VOCD2	VOCD3	Average
1A	120.33	119.05	120.45	119.94
1C	128.42	126.92	128.45	127.93
2A	102.32	102.73	102.47	102.51
2C	62.43	62.51	62.64	62.53
3A	124.83	124.86	125.74	125.14
3C	54.35	55.06	54.69	54.70
4A	126.47	126.23	127.51	126.74
4C	65.80	65.16	66.34	65.77
5A	89.22	89.64	91.18	90.01
5C	87.99	88.35	89.03	88.46
6A	95.69	95.07	95.44	95.40
6C	73.43	72.74	72.97	73.05
7A	138.22	138.27	136.76	137.75
7C	84.54	84.62	83.69	84.28
8A	80.34	80.19	81.64	80.72
8C	123.91	123.29	122.60	123.27
9A	113.37	115.06	113.14	113.86
9C	87.45	87.55	88.13	87.71
10A	139.67	138.57	138.90	139.05
10C	49.04	48.83	48.83	48.90
11A	73.72	75.13	74.19	74.35
11C	57.93	57.80	58.04	57.92
12A	98.58	98.85	99.37	98.93
12C	70.44	71.24	71.27	70.98
13A	119.68	117.79	119.54	119.00
13C	81.38	82.29	82.44	82.04
14A	104.56	105.21	104.82	104.86
14C	66.59	66.25	66.26	66.37
15A	112.18	111.62	113.60	112.47
15C	79.76	80.38	79.16	79.77
16A	116.19	113.44	113.36	114.33
16C	64.67	64.37	64.73	64.59
17A	74.96	75.69	75.94	75.53
17C	81.67	82.52	82.27	82.15
18A	124.38	121.77	121.89	122.68
18C	66.16	66.11	67.81	66.69
19A	128.45	128.61	127.32	128.13
19C	86.08	84.11	84.84	85.01

A/C No	VOCD1	VOCD2	VOCD3	Average
20A	150.30	147.92	150.15	149.46
20C	65.87	65.18	66.44	65.83

Table 5: D values for three independent measures

The table shows that the values of all three VOCD-D measures for each text are quite similar and they are thus representative. No value was significantly different for any L1 text or L2 text. Kruskall-Wallis test revealed no significant difference between the groups ($H = 0.018$, $p > 0.05$).

Average D values are given in the next table and figure 11.

VOCD-D (average)	Article No	VOCD-D (average)	Comment. No
74.35	11A	48.90	10C
75.53	17A	54.70	3C
80.72	8A	57.92	11C
90.01	5A	62.53	2C
95.40	6A	64.59	16C
98.93	12A	65.77	4C
102.51	2A	65.83	20C
104.86	14A	66.37	14C
112.47	15A	66.69	18C
113.86	9A	70.98	12C
114.33	16A	73.05	6C
119.00	13A	79.77	15C
119.94	1A	82.04	13C
122.68	18A	82.15	17C
125.14	3A	84.28	7C
126.74	4A	85.01	19C
128.13	19A	87.71	9C
137.75	7A	88.46	5C
139.05	10A	123.27	8C
149.46	20A	127.93	1C
111.54	Mean	76.90	Mean

Table 6: VOCD-D average values

L2 text VOCD-D mean value is 76.90 and L1 text VOCD-D mean value is substantially higher – 111.54. This result is in accordance with the previously stated hypothesis.

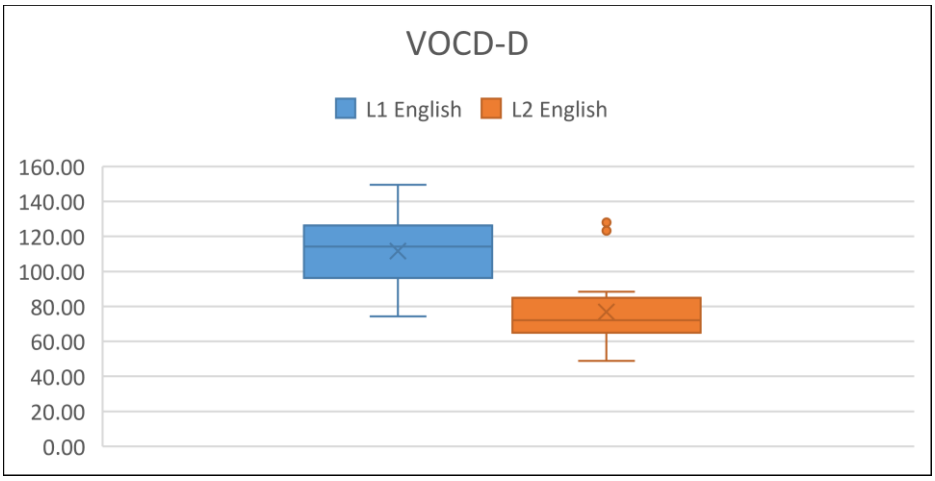


Figure 11: A comparison of VOCD-D between L1 and L2 texts

A detailed look at D values shows a large variance for **L1 texts/articles ($\bar{x}$ = 111.54, min = 74.35, max = 149.46, SD = 21.1)**. This variance could be influenced by the different L1 text topics and by their different aim. The distribution of VOCD-D values across L1 texts is even, which is proven by the close average (111.54) and the median (114.10).

The same even distribution is for L2 texts, with the average being 76.90 and the median 72.02. The data for **L2 texts/commentaries** is as follows **($\bar{x}$ = 76.90, min = 48.90, max = 127.93, SD = 20.1)**. Again, there is a large variance in values. The interesting fact is that the two highest values are very distinctive (1C – 127.93, 8C – 123.27) as they are even higher than the average value for L1 texts. The third maximum value is 88.46, which is far less than the two highest values.

In the case of MTLD, the values and their distribution are summarised in the next table and the next figure.

MTLD	Article No	MTLD	Comment. No
60.80	11A	46.46	11C
78.53	8A	48.37	10C
90.09	12A	54.93	18C
92.11	17A	56.50	3C
93.86	6A	57.95	2C
101.20	18A	58.27	4C
104.01	15A	61.61	19C
105.71	5A	63.08	16C
111.27	3A	65.39	20C

MTLD	Article No	MTLD	Comment. No
114.16	2A	67.98	12C
115.02	16A	69.68	14C
118.35	1A	71.51	7C
127.03	14A	71.66	13C
133.23	9A	79.39	17C
135.26	13A	80.90	6C
136.03	19A	82.89	15C
144.79	7A	87.45	5C
153.83	4A	88.08	9C
162.97	10A	112.84	1C
188.47	20A	122.83	8C
118.34	Mean	72.39	Mean

Table 7: MTLD values

MTLD values for **L2 texts/commentaries** represented by the mean value, the minimum value, the maximum value and the standard deviation are **($\bar{x}$ = 72.39, min = 46.46, max = 122.83, SD = 19.7)**. On the other hand, MTLD values for **L1 texts/articles** reach higher values **($\bar{x}$ = 118.34, min = 60.80, max = 188.47, SD = 30.3)**.

As was previously the case with VOCD-D measurement, here again, MTLD levels are higher for L1 texts (articles) than those of L2 texts (commentaries).

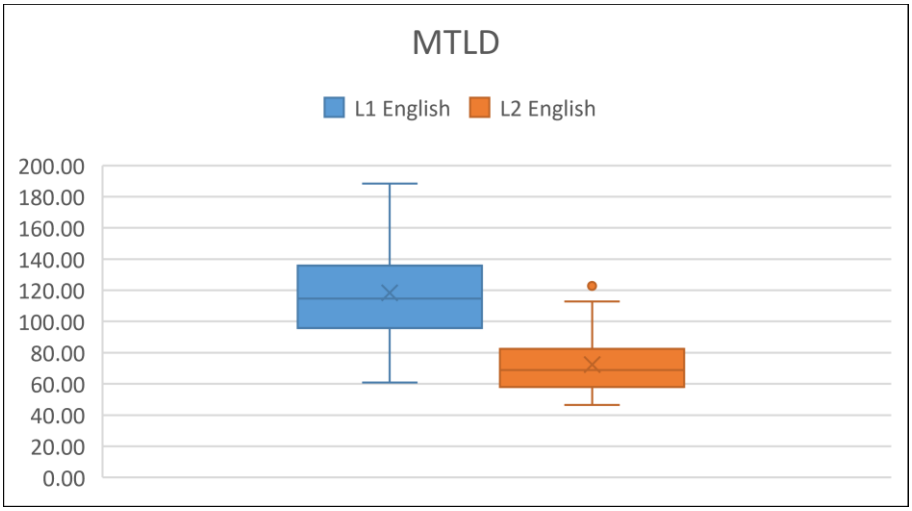


Figure 12: A comparison of MTLD between L1 and L2 texts

The last view on lexical diversity is the comparison of VOCD-D and MTLD results together.

Article No	VOCD-D (average)	MTLD	Comm. No	VOCD-D (average)	MTLD
1A	119.94	118.35	1C	127.93	112.84
2A	102.51	114.16	2C	62.53	57.95
3A	125.14	111.27	3C	54.70	56.50
4A	126.74	153.83	4C	65.77	58.27
5A	90.01	105.71	5C	88.46	87.45
6A	95.40	93.86	6C	73.05	80.90
7A	137.75	144.79	7C	84.28	71.51
8A	80.72	78.53	8C	123.27	122.83
9A	113.86	133.23	9C	87.71	88.08
10A	139.05	162.97	10C	48.90	48.37
11A	74.35	60.80	11C	57.92	46.46
12A	98.93	90.09	12C	70.98	67.98
13A	119.00	135.26	13C	82.04	71.66
14A	104.86	127.03	14C	66.37	69.68
15A	112.47	104.01	15C	79.77	82.89
16A	114.33	115.02	16C	64.59	63.08
17A	75.53	92.11	17C	82.15	79.39
18A	122.68	101.20	18C	66.69	54.93
19A	128.13	136.03	19C	85.01	61.61
20A	149.46	188.47	20C	65.83	65.39
Mean	111.54	118.34	Mean	76.90	72.39

Table 8: VOCD-D and MTLD values

In the case of L1 texts (articles), minimum (light blue cells in the table) and maximum values (yellow cells) were the same, regardless of whether the VOCD-D or MTLD measurement was used. In the case of L2 texts, minimum values were the same but for maximum values slightly different. In each case, the maximum value was relevant to the second-highest value respectively.

Overall, average values for both lexical diversity measurements are nearly the same (the difference is less than 6%). Interestingly, VOCD-D average values (76.90) were higher than MTLD values (72.39) for L2 texts and lower for L1 texts (111.54 vs. 118.34), but the difference is marginal.

In conclusion, we can assume that L1 English lexical diversity is higher than L2 English lexical diversity for both measures (VOCD: 111.54 vs. 76.90; MTLD: 118.34 vs. 72.39). The gap between native speakers´ and C2 English learners´ lexical diversity materialised greater in the case of MTLD (38%). In the case of VOCD, the difference reached circa 31%.

Mann–Whitney U test was carried out to compare the data for L1 texts and L2 texts. The difference between the randomly selected values of VOCD-D and the MTLD populations is not large enough to be statistically significant ($Z = 0.21$, $p > 0.05$).

3.3.2.2 Lexical sophistication

In this thesis, lexical sophistication is measured through the Text Inspector´s "Lexis: EVP" tool. The tool investigates the number of words in distinctive CEFR categories (here only six categories A1-C2, the pre-A1 category is not included) both in absolute numbers and in relative values – percentages. The instances of words according to CEFR categories, which were taken from the L2 text No. 11C, are: (A1) *the*, *and*, *on*; (A2) *gas*, *able*, *could*; (B1) *drops*, *quantity*, *raising*; (B2) *loss*, *specific*, *theory*; (C1) *comparable*, *formula*, *hence*; (C2) *allocation*, *marginal*, *prevailing*. The Text Inspector software enables chunking words that collocate or are part of phrasemes, that is why lexical frequency lists also incorporate collocations or phrasemes. For example, *took advantage of* (B1, text 2A), *according to* (B1, text 8A), *designed for* (B2, text 8A) or *resulting in* (B2, text 15C).

In the next table, an overview of the L1 text lexical sophistication, measured through English Vocabulary Profile frequency lists, is demonstrated.

Article No	A1 types	A1 types (%)	A2 types	A2 types (%)	B1 types	B1 types (%)	B2 types	B2 types (%)	C1 types	C1 types (%)	C2 types	C2 types (%)	Unlisted types	Unlisted types (%)
1A	66	19.9	49	14.8	56	16.9	60	18.1	12	3.6	5	1.5	82	24.7
2A	112	27.5	65	16.0	83	20.4	59	14.5	13	3.2	12	2.9	60	14.7
3A	100	33.1	52	17.2	43	14.2	37	12.3	9	3.0	8	2.6	52	17.2
4A	88	34.4	45	17.6	28	10.9	21	8.2	5	2.0	10	3.9	57	22.3
5A	66	28.3	25	10.7	49	21.0	35	15.0	14	6.0	3	1.3	40	17.2

Article No	A1 types	A1 types (%)	A2 types	A2 types (%)	B1 types	B1 types (%)	B2 types	B2 types (%)	C1 types	C1 types (%)	C2 types	C2 types (%)	Unlisted types	Unlisted types (%)
6A	78	29.7	34	12.9	31	11.8	42	16.0	13	4.9	6	2.3	56	21.3
7A	148	27.5	87	16.1	76	14.1	77	14.3	25	4.6	20	3.7	103	19.1
8A	80	27.5	36	12.4	50	17.2	56	19.2	11	3.8	9	3.1	47	16.2
9A	75	30.1	48	19.3	31	12.4	29	11.6	9	3.6	8	3.2	46	18.5
10A	101	29.3	62	18.0	49	14.2	35	10.1	6	1.7	8	2.3	84	24.3
11A	39	28.7	24	17.6	20	14.7	10	7.4	3	2.2	1	0.7	39	28.7
12A	85	31.4	39	14.4	52	19.2	34	12.5	6	2.2	6	2.2	49	18.1
13A	94	36.9	44	17.3	24	9.4	36	14.1	8	3.1	4	1.6	43	16.9
14A	48	21.8	29	13.2	29	13.2	31	14.1	11	5.0	7	3.2	62	28.2
15A	141	28.7	85	17.3	71	14.5	69	14.1	25	5.1	8	1.6	95	19.3
16A	175	36.2	91	18.8	76	15.7	69	14.3	13	2.7	2	0.4	62	12.8
17A	48	29.1	24	14.5	33	20.0	22	13.3	6	3.6	5	3.0	26	15.8
18A	94	42.9	39	17.8	22	10.0	20	9.1	6	2.7	4	1.8	36	16.4
19A	146	26.6	87	15.9	98	17.9	67	12.2	28	5.1	17	3.1	108	19.7
20A	61	25.5	32	13.4	32	13.4	25	10.5	13	5.4	6	2.5	67	28.0
Mean	92	29.7	50	15.8	48	15.1	42	13.0	12	3.7	8	2.4	61	20.0

Table 9: L1 text lexical sophistication

It can be inferred from the table that there is a large proportion of words that are categorised as unlisted types (20%). If we see this phenomenon in detail by reading through all texts, the structure of these words in L1 texts mainly consists of four groups:

- numbers (e.g., *trillion, 2017*),
- proper nouns (e.g., *Trump, Indonesia*),
- abbreviations and acronyms (e.g., *GDP [Gross Domestic Product], MUP [Minimum Unit Pricing]*),
- specific economic lexis (e.g., tariff, treasury).

Numbers and proper names are generally more frequent in economic texts, which is why the percentage of unlisted types is higher than usual. Among specific economic terms which were not recognised as A1-C2 categories (A1-C2 types in the Text Inspector) by Text Inspector are, for example, *tariff*, *counterbalance*, *reallocate*, *treasury* or *exporter*. All these words were

even manually checked in the English Vocabulary Profile webpage database (2015) and were not present in the database.

There is a steady decline in the distribution from A1 types to C2 types in the L1 text sample as is demonstrated below. As the length of L1 and L2 texts is different, the interpretation is taken from percentages, not from absolute numbers.

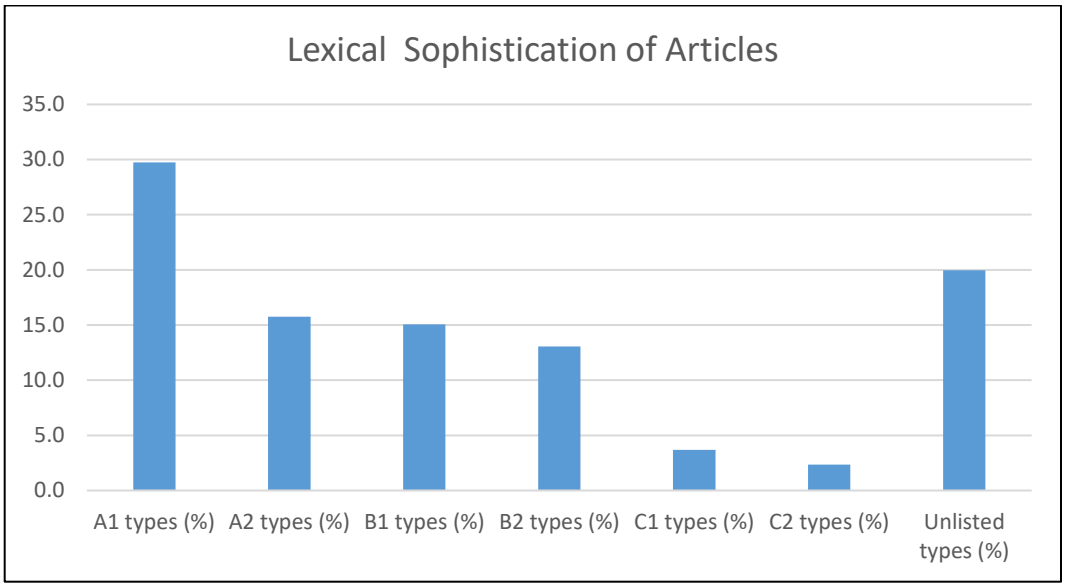


Figure 13: L1 text lexical sophistication

A1 word-types amount to 29.7%, A2 types 15.8%, B1 types 15.1%, B2 types 13.05%, C1 types 3.7%, C2 types 2.4% and unlisted types 20% on average.

The next table contains the values of the L2 text lexical sophistication.

Commentary No	A1 types	A1 types (%)	A2 types	A2 types (%)	B1 types	B1 types (%)	B2 types	B2 types (%)	C1 types	C1 types (%)	C2 types	C2 types (%)	Unlisted types	Unlisted types (%)
1C	82	24.0	42	12.3	66	19.3	64	18.7	16	4.7	4	1.2	61	17.8
2C	73	28.9	24	9.5	65	25.7	44	17.4	8	3.2	9	3.6	31	12.3
3C	62	25.7	36	14.9	48	19.9	40	16.6	12	5.0	3	1.2	40	16.6
4C	76	32.8	34	14.7	49	21.1	36	15.5	10	4.3	2	0.9	25	10.8
5C	64	19.1	49	14.6	64	19.1	69	20.6	25	7.5	9	2.7	55	16.4
6C	73	24.5	45	15.1	59	19.8	55	18.5	13	4.4	9	3.0	43	14.4

Commentary No	A1 types	A1 types (%)	A2 types	A2 types (%)	B1 types	B1 types (%)	B2 types	B2 types (%)	C1 types	C1 types (%)	C2 types	C2 types (%)	Unlisted types	Unlisted types (%)
7C	65	21.0	44	14.2	54	17.4	57	18.4	22	7.1	12	3.9	58	18.7
8C	61	16.4	43	11.5	61	16.4	84	22.5	23	6.2	16	4.3	82	22.0
9C	66	18.6	43	12.1	72	20.3	72	20.3	25	7.0	16	4.5	61	17.2
10C	77	30.1	48	18.8	50	19.5	42	16.4	9	3.5	5	2.0	24	9.4
11C	73	27.8	53	20.2	53	20.2	36	13.7	14	5.3	6	2.3	26	9.9
12C	95	32.2	48	16.3	62	21.0	48	16.3	7	2.4	8	2.7	27	9.2
13C	110	40.4	51	18.8	33	12.1	47	17.3	8	2.9	5	1.8	17	6.3
14C	83	27.0	50	16.3	59	19.2	56	18.2	16	5.2	3	1.0	37	12.1
15C	91	27.9	64	19.6	67	20.6	53	16.3	17	5.2	6	1.8	27	8.3
16C	97	31.7	45	14.7	62	20.3	50	16.3	15	4.9	9	2.9	23	7.5
17C	103	32.9	39	12.5	57	18.2	49	15.7	19	6.1	9	2.9	35	11.2
18C	97	32.9	44	14.9	59	20.0	57	19.3	11	3.7	5	1.7	23	7.8
19C	93	32.9	55	19.4	54	19.1	38	13.4	10	3.5	5	1.8	27	9.5
20C	75	27.0	43	15.5	51	18.3	51	18.3	10	3.6	4	1.4	46	16.5
Mean	81	27.7	45	15.3	57	19.4	52	17.5	15	4.8	7	2.4	38	12.7

Table 10: L2 text lexical sophistication

Here, the number of unlisted types is smaller than it was in the previous case (12.7% vs. 20%) but it is still a high proportion. The reason for such a high representation is due to the economic nature of all texts which include many numbers, proper nouns, abbreviations (predominantly acronyms) and specific economic terms which are not included in the English Vocabulary Profile database.

If we investigate the distribution of different types according to EVP, there is no steady decline as was the previous case in L1 texts. A2 word types are less frequent than B1 types and B2 types. Overall, A1-word types make 27.7%, A2 types 15.3%, B1 types 19.4%, B2 types 17.5%, C1 types 4.8%, C2 types 2.4% and unlisted types 12.7% on average.

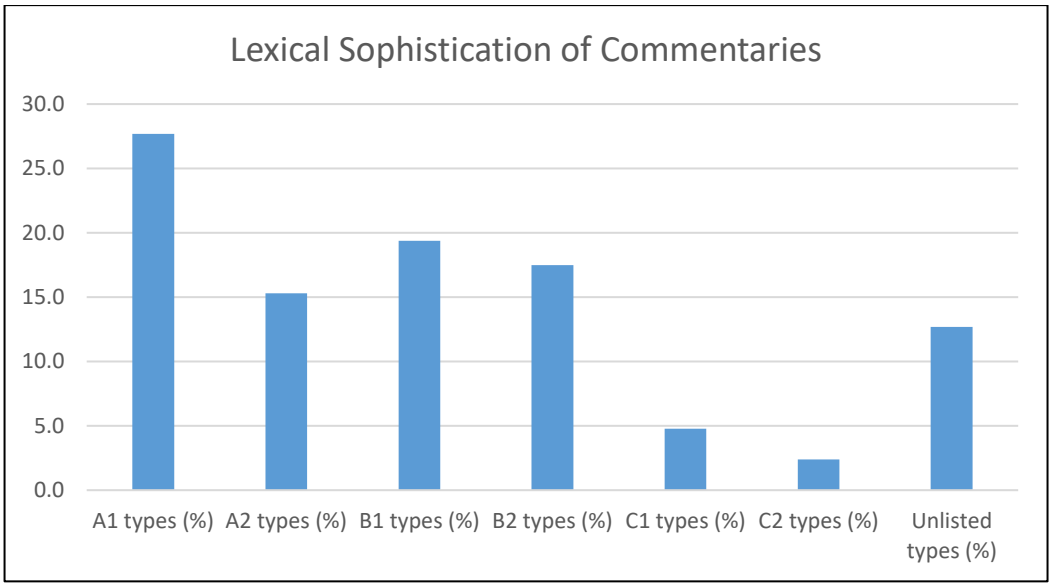


Figure 14: L2 text lexical sophistication

Finally, if we compare the lexical sophistication of L1 and L2 texts together, we get the following picture.

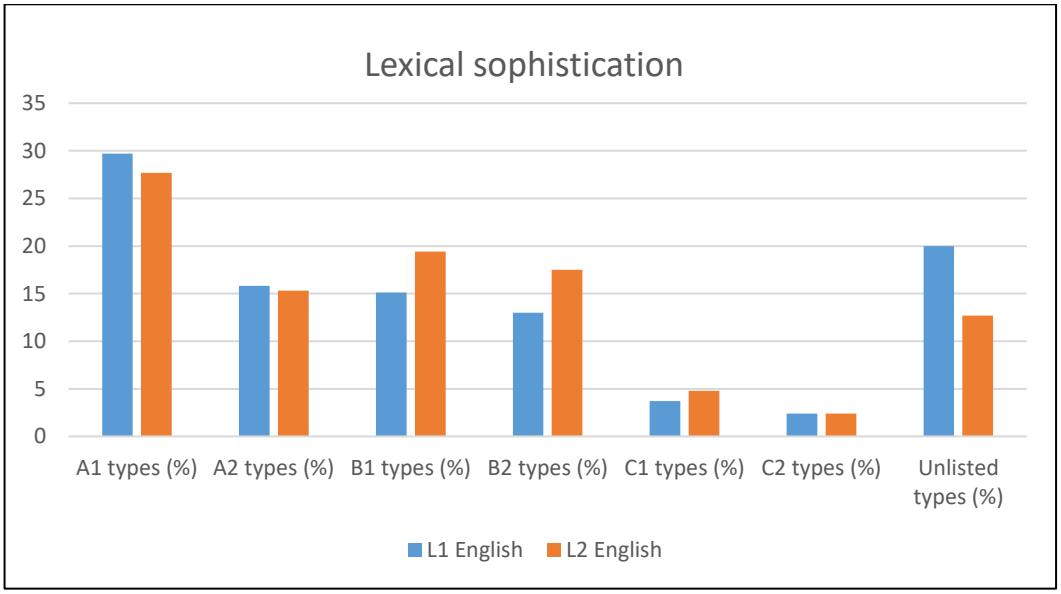


Figure 15: Comparison of lexical sophistication

Three important observations can be made from this available data.

1/ The number of low-frequency words, represented by C2 types, is low – 2.4% for both L1 and L2 texts. This number would be somewhat higher if we counted some words from the unlisted types where the proportion is quite high (20.0% and 12.7%). Therefore, the hypothesis

that the number of low-frequency words of L1 native speakers will be higher than those of L2 proficient learners was not proven.

2/ The distribution of L1 text lexical sophistication and of L2 text lexical sophistication is comparable: A1 types – 29.7%, 27.7%; A2 types – 15.8%, 15.3%; B1 types – 15.1%, 19.4%; B2 types – 13.05%, 17.5%; C1 types – 3.7%, 4.8%; C2 types – 2.4%, 2.4%. The largest difference is in B1 and B2 types.

3/ The overall trend of word types, using EVP frequency lists, is the decrease from A1 to C2 type (with a small exception of A2 type), according to Zipf´s law.

3.3.3 Discussion

The thesis examined lexical complexity in the written production of L2 proficient learners of English in two dimensions: lexical diversity by the means of composite indicators (VOCD-D and MTLD) and lexical sophistication by the means of frequency wordlists (EVP). Two main research questions were initially set:

- Is lexical diversity of English native speakers higher than that of non-native speakers (with L1 Czech)?
- Is lexical sophistication of English native speakers higher than that of non-native speakers (with L1 Czech)?

Based on scientific literature and various studies in the given area, the following hypothesis was defined: it is presumed that L2 proficient learners of English have lower lexical complexity than native speakers. Nevertheless, the hypothesis could not be proven and the results are inconclusive.

Even though measuring lexical sophistication is problematic due to many factors which can influence the results (see the next chapter 3.3.4 Limitations of the study), Milton argues that “lexical qualities of specific texts within a single genre” in controlled tasks could be still proven useful (2009: 138). According to him, examining a language learner´s lexical complexity can reveal the current state of affairs with all the strengths and weaknesses of the learner. Based on this analysis, a concrete plan for learner´s improvement can be suggested (ibid.).

Rálišová (2020) investigated lexical complexity in L2 speech of L1 Czech speakers in her thesis. She used the EVP and COCA frequency list approach with the aim to show the difference in lexical sophistication between B2 and C1 speakers. Nevertheless, the results were inclonclusive: "neither of the two lexical frequency metrics from Text Inspector […] proved that C1 speakers would indeed be the more prolific users of low-frequency vocabulary". According to Rálišová (2020), this inconclusiveness is due to variables such as "sensitivity to the oral genre, the low wordcount, too advanced CEFR levels of speakers bordering on each other and low number of speakers participating". Besides, the low occurrence of higher-proficiency vocabulary (i.e. low-frequency words) can be explained by the Zipf's law.

Another research on a lexical variety of oral production was conducted by Škutová (2020), where the correlation between the proficiency levels of speakers and their range of vocabulary was investigated. The results were inconclusive again as was the case in Rálišová's research: "it remains unclear to what extent lexical variety in oral production contributes to our perception of the speaker's proficiency or whether there is a clear correlation between the proficiency levels of speakers and the range of vocabulary they employ". To reach more conclusive results, a larger sample of speakers would be necessary (ibid.).

Due to its sensitivity to text length (Kuiken and Vedder: 2012), TTR was not regarded as the main indicator of text lexical diversity. Instead, the composite indicators VOCD-D and MTLD were used. To even improve the reliability of the measurement three VOCD-D independent analyses were performed and then the mean was calculated.

If we compare L2 text lexical diversity measured by VOCD-D in this thesis with Duran et al. (2004a) results, the mean value of D is 76.90 for L2 texts and somewhere between 40–70 for the adult second learner cohort (see figure 4). Duran´s (ibid.) categorisation does not distinguish different proficiency levels of the adult second learner group but if we assume that the lowest level (A1) would be with the value around 40 and the highest level (C2) circa 70, we can conclude that the value of 76.90 represents very proficient command of written English as it is even higher than the upper boundary of Duran´s research. Our adult second learner cohort language proficiency level is C2 which is the highest level of mastery. The difference is quite small (9.8%), so results with Duran et al. (2004a) are quite similar.

If we see the VOCD-D values for L1 texts, the mean value is 111.54. Duran et al. (2004a) found in their research that lexical diversity, represented by index D, typically have a measure of between 80-105 for adult native speaker cohort writing academic texts (see again figure 4).

By comparison, our D mean value for L1 texts is even higher than the maximum border which proves the fact that the lexical diversity of the texts produced by the journalists is probably higher because journalists´ writing competence is extremely high due to their specialised training.

We can assume that L1 English lexical diversity is higher than L2 English lexical diversity as VOCD is 111.54 and MTLD is 118.34 for L1 texts and 76.90 and 72.39 respectively for L2 texts. These results are in accordance with Duran et al. research (2004a). On the other hand, Foster and Tavakoli (2009) showed that there is no particular difference between native and non-native speaker groups in the variance in scores for lexical diversity in their research.

Palfreyman (2019), in his study, suggested that there is the possibility of a positive correlation between lexical sophistication in L1 and L2 written production. In his research, he focused on undergraduate English students with Arabic L1. This "transfer" from L1 to L2 was not investigated in the thesis but would be definitely interesting to research.

Milton (2009) argues that lexical sophistication would be probably higher in written texts than in oral production as they usually contain more infrequent lexis. If we assume that there is usually more time for written production than for oral one and written texts are considered more formal, we can predict that this hypothesis is true. Nevertheless, it would have to be followed by additional data analysis.

Tabari et al. (2021) found that task sequencing plays a significant role in L2 written output. This result is not surprising and has an important consequence, especially for teachers. According to this research, they should follow the path from easily written assignments to more complex ones.

3.3.3.1 Limitations of the study

There are several limitations to the study of lexical complexity in the written production of L2 proficient learners of English.

The first limitation is that the task to produce a written sample is influenced by the initial article choice from the news media. L2 writers are in this way primed by the L1 text lexis which was already used in the text. The **priming effect** is reinforced by the fact that the authors of news-media articles are considered to be experts and authorities on the given topic. The

consequence is that students have the tendency to repeat the same lexis and the impact on lexical diversity is substantial.

The second restriction is the length of produced academic written texts. The assignment instructions asked students to write up to 800 words (the mean was 781 words). It means that **the length of the written texts** must not have exceeded this word count, otherwise, there would be some penalty to a student´s mark. In other words, the results of this research would be more conclusive if the length of written texts were longer and not restricted by the word count.

The next limitation, which influences the validity of the results, is **the number of texts** researched. The larger the number of texts, the more conclusive the results would be. This limitation is very close to the previous one, which was the length of the texts. Together, they determine the whole sample of the study. By increasing either of these variables or both, the sample would be more reliable and results more conclusive.

Another restriction is that the **topic**, which students deal with, is very technical, in this case from the realm of economics. Generally, synonyms are less frequent in scientific writings than in general language. For instance, Milton (2009) argues that the **genre** has a strong impact on lexical diversity. Another research showed the influence of various disciplines (humanities, life sciences and physics) on linguistic complexity (Khany and Kafshgar: 2016). The study of Yoon and Polio (2017) found that the genre of written texts influences word frequencies: argumentative essays contained more low-frequency words than narrative texts.

The results of lexical sophistication testing can be influenced by Text Inspector´s tools too. The first weak point of the Text Inspector´s EVP tool is **word polysemy**. The EVP tool is not so precise as to the recognition of the right word meaning. If a word is polysemous, such as the word *mark*, the EVP tool will have difficulty recognising a lexical category and a concrete meaning. For instance, English Vocabulary Profile (2015) states 4 different verb meanings for the word *mark*:

1. *to check a piece of work or an exam, showing mistakes and giving a letter or number to say how good it is* (B2)
2. *to show where something is by drawing or putting something somewhere* (B2)
3. *to leave an area of dirt or damage on something* (C1)
4. *to represent or show a characteristic of a person or thing or feeling* (C2)

and 3 different noun meanings for the word *mark*:

1. *a number or letter that is written on a piece of work, saying how good the work is* (A2)
2. *a small area on the surface of something which is damaged, dirty or different in some way* (B2)
3. *a symbol which is used for giving information* (C1).

Perhaps, it would be easier to determine a word class from the sentence syntax than to guess the right meaning from the context. Consequently, the type can be determined wrongly (for instance the EVP tool may mark the verb *mark* incorrectly as B2 type instead of C2 type), and the results of lexical sophistication will be in this way distorted.

Another weak point of the Text Inspector´s EVP tool is that it does not incorporate all words and meanings in its corpus which results in a high proportion of **unlisted types**, for example, the word *crown* is not included in the corpus (English Vocabulary Profile: 2015; accessed 27th September 2021), and the word *capital* is only included in one meaning: *a city which is the centre of government of a country, state or region* (A2) (English Vocabulary Profile: 2015) and not in the second economic meaning: *money or property, especially when it is used to start a business or to produce more wealth* (Pearson Education 2021).

4 Conclusion

L1 English lexical diversity was substantially higher than L2 English lexical diversity both for VOCD-D (111.54 vs. 76.90) and MTLD (118.34 vs. 72.39) composite indicators despite the priming effect of the texts as was described above. The lexical diversity gap between native speakers and C2 English learners, considering these two composite indicators, has reached approximately 34.5%. Another interesting result regarding lexical diversity, which was described in chapter 3.3.2.1, is that L2 English learners/users tend to repeat the same vocabulary more than L1 English adults.

Lexical sophistication, represented by the number of low-frequency words (C2 types) was the same for both L1 and L2 texts, namely 2.4%. This number is, nevertheless, affected by the high proportion of unlisted types (20.0% in L1 texts and 12.7% in L2 texts). The unlisted types in EVP comprise numerals, proper nouns, and abbreviations and acronyms, which are all common to economic texts. Moreover, many specific economic terms are not present in the

English Vocabulary Profile database. Therefore, the hypothesis that the number of low-frequency words of L1 native speakers will be higher than those of L2 proficient learners was not proven.

The distribution of L1 text lexical sophistication and L2 text lexical sophistication is similar. The largest difference between them is in B1 and B2 word types. The overall trend of word type distribution, using EVP frequency lists, is decreasing from A1 to C2 types, which is in accordance with Zipf´s law. "Zipf´s law states that if words of a language are sorted in the order of decreasing frequency of usage, a word's frequency is inversely proportional to its rank, or sequence number in the list" (Manin 2009).

This thesis research could be supplemented by the investigation of lexical complexity differences of written texts and oral production in the same economic topic and by the exploration of low-frequency productive vocabulary (writing and speaking competences) and low-frequency receptive vocabulary (reading and listening competences). The preliminary hypotheses would be that the lexical complexity of written texts is higher than in the case of oral production and that the productive lexis covers more low-frequency words than receptive lexis.

Another interesting research supplement would be a lexical complexity investigation of different IB DP student cohorts across L1 nationalities. In this way, the lexical complexity of different English second speakers on the same topic and assignment could be analysed.

5 Bibliography

Bax, Stephen. TextInspector. *textinspector.com*. Last accessed on 27th October 2021.

Bialystok, E. (1994). Analysis and Control in the Development of Second Language Proficiency. *Studies in Second Language Acquisition*, 16(2), 157–168.

Bulantová, B. (2020). *Syntactic Complexity in the Speech of Learners of English: Issues in Operationalization*. Prague: Charles University. MA Thesis, consultant: PhDr. Tomáš Gráf, Ph.D.

Bulté, B. and Housen, A. (2012). Defining and operationalising L2 complexity. In: A. Housen, F. Kuiken and I. Vedder, ed., *Dimensions of L2 Performance and Proficiency. Complexity, Accuracy and Fluency in SLA.* Amsterdam: John Benjamins.

Bulté, B. and Housen, A. (2014). Conceptualizing and Measuring Short-Term Changes in L2 Writing Complexity. *Journal of Second Language Writing* 26. 42–65.

Brumfit, C. J. (1984). *Communicative Methodology in Language Teaching: The Roles of Fluency and Accuracy.* Cambridge: Cambridge University Press.

Connor, U., and Schneider, M. (1990). Analyzing topical structure in ESL essays. *Studies in Second Language Acquisition*, *12*(04), 411–427.

Chafe, W., and Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (p. 83–113). Academic Press.

Council of Europe. (2001). *Common European Framework of Reference for Languages. Learning, Teaching, Assessment.* Cambridge University Press.

Crystal, D. (2019). *The Cambridge Encyclopedia of the English Language* Cambridge: Cambridge University Press.

Czwenar, I. (2013). Analysing Spoken Language for Complexity, Accuracy and Fluency: Some Methodological Considerations. In Szubko-Sitarek, W. and Salski, L. and

Stalmaszczyk, P. (eds.) *Language Learning. Discourse and Communication.* New York: Springer.

Duran, P. and Malvern, D. and Richards, B. and Chipere, N. (2004a). Developmental Trends in Lexical Diversity. *Applied Linguistics* 25 (2). 220–242.

Duran, P. and Malvern, D. and Richards, B. and Chipere, N. (2004b). Lexical Diversity and Language Development. Palgrave Macmillan UK.

English Vocabulary Profile. (2015). *English Profile: The CEFR for English*. Cambridge University Press. (http://englishprofile.org/index.php/wordlists). Last accessed on 27th September 2021.

Ellis, R., and Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford University Press.

Foster, P., and Tavakoli, P. (2009). Native Speakers and Task Performance: Comparing Effects on Complexity, Fluency, and Lexical Diversity. *Language Learning: A Journal of research in Language Studies*, 59(4), 866–896.

Housen, A., and Kuiken, F. (2009). Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 32. Oxford University Press.

Housen, A., Kuiken, F., and Vedder, I. (2012). *Dimensions of L2 performance and Proficiency: Complexity, Accuracy, and Fluency in Second Language Acquisition*. (Language Learning and Language Teaching, v. 32). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.32.

Huang, Lan-fen, and Kubelec, Simon, and Keng, Nicole, and Hsu, Lung-hsun. (2018). Evaluating CEFR rater performance through the analysis of spoken learner corpora. *Language Testing in Asia* 8 (1).

International Baccalaureate Organization. (2020). *Diploma Programme Economics Guide.* United Kingdom.

Johnson, Mark D. (2017). Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency:A research synthesis and meta-analysis. *Journal of Second Language Writing* 37. 13–38.

Khany, R. and Kafshgar, N. D. (2016) Analysing Texts through their Linguistic Properties: A Cross-disciplinary Study. *Journal of Quantitative Linguistics*, 23 (3), 278-294, doi: 10.1080/09296174.2016.1169848.

Kuiken, F. and Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. *Dimensions of L2 Performance and Proficiency*, 143 – 170.

Kyle, Kristopher and Crossley, Scott A. (2015). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly* 49 (4). 757–786.

Larsen-Freeman, D. (2009). Adjusting Expectations: The Study of Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics,* 30(4), 579–589. Oxford University Press.

Limbrick, Libby, et al. (2008). *Enhancing Capacity to Analyse Students' Writing*. Teaching & Learning Research Initiative Nāu i Whatu Te Kākahu, He Tāniko Taku.

Lu, Xaofei (2010). A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development. *TESOL Quarterly*, 45 (1), 36–62.

Manin, D. Yu (2009). Mandelbrot's Model for Zipf's Law: Can Mandelbrot's Model Explain Zipf's Law for Language? *Journal of Quantitative Linguistics*, 16 (3), 274-285, doi: 10.1080/09296170902850358.

McCarthy, P.M., Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods* 42**,** 381–392. https://doi.org/10.3758/BRM.42.2.381.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring Vocabulary Diversity Using Dedicated Software, *Literary and Linguistic Computing.* 15(3), 323-337.

Milton, James. (2009). *Measuring Second Language Vocabulary Acquisition* (Second Language Acquisition). Bristol, UK ; Buffalo [N.Y.]: Multilingual Matters.

Oxford University Press (2020). *Oxford English Dictionary.* Oxford dictionary online.

Palfreyman, D.M. & Karaki, S. (2019). Lexical sophistication across languages: a preliminary study of undergraduate writing in Arabic (L1) and English (L2). *International journal of bilingual education and bilingualism*. 22(8), 992-1015.

Palloti, G. (2015). A Simple View of Linguistic Complexity. *Second Language Research*, 31(1), 117-134.

Pearson Education (2021). *Longman Dictionary of Contemporary English Online (LDOCE)*. (http://www.ldoceonline.com). Last accessed on 24th December 2021.

Quirk, R. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Rálišová, D. (2020). *Measuring lexical complexity in L2 speech with word frequency lists*. Prague: Charles University. MA Thesis, consultant: PhDr. Tomáš Gráf, Ph.D.

Read, John. (2000). *Assessing Vocabulary* (Cambridge Language Assessment Series). Cambridge: Cambridge University Press.

Skehan, P., and Foster, P. (1998). Task type and task processing conditions as influences on foreign language performance. Language Teaching Research, 1(3), 185-211.

Skehan, P. (2009). Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics*, 30(4), 510–532.

Skehan, P., and Foster, P. (2008). Complexity, accuracy, fluency and lexis in task-based perfor-mance: A meta-analysis of the Ealing research. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard, and I. Vedder (Eds.). *Complexity, accuracy and fluency in second language use, learning and teaching*, 201–222. Wetteren: Universa Press.

Škutová, Z. (2020). *Lexical variety in oral production of L2 learners of English as a factor in determining language proficiency*. Prague: Charles University. BA Thesis, consultant: PhDr. Tomáš Gráf, Ph.D.

Tabari, M., and Miller, M. (2021). Unraveling the Effects of Task Sequencing on the Syntactic Complexity, Accuracy, Lexical Complexity, and Fluency of L2 Written Production. *Canadian Journal of Applied Linguistics*, *24*(2), 1–29. https://doi.org/10.37213/cjal.2021.31306.

Wolfe-Quintero, K., Inagaki, S., and Kim, H.-Y. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, & Complexity.* Honolulu: Second Language Teaching & Curriculum Center, University of Hawaii.

Yoon, H., and Polio, C. (2017). The Linguistic Development of Students of English as a Second Language in Two Written Genres. *TESOL Quarterly*, 51(2), 275–301.

6 Czech Résumé

Diplomová práce si kladla za cíl analýzu lexikální komplexnosti písemného projevu nerodilých mluvčích anglického jazyka, kteří jsou na nejvyšší úrovni jazykové kompetence. Ke splnění cílů bylo nejprve nutné definovat dva klíčové pojmy, a to jazykovou pokročilost nerodilých mluvčí a lexikální komplexnost. Jazykové kompetence pokročilého mluvčího je možné například hodnotit podle modelu čtyř dovedností, kterými jsou psaní, mluvení, čtení a poslech. Dalším modelem, kterým je možné měřit úroveň jazykové kompetence je model složený ze tří komponent: komplexnosti, přesnosti a plynulosti jazykového projevu, tzv. CAF model (z anglického complexity, accuracy a fluency) (Housen et al. 2012). Kromě těchto dvou modelů se práce také zabývá hodnocením jazykové úrovně ve Společném Evropském Referenčním Rámci (SERR) (The Council of Europe 2001), který je v evropském kontextu stěžejní. Nejvyšší jazyková pokročilost je zde označována úrovní C2 a je měřena jak pomocí obecných deskriptorů jazykové kompetence, tak pomocí specifických deskriptorů pro jednotlivé jazykové dovednosti. Průnik použité metodiky v SERR s modely CAF a čtyř dovedností není jistě náhodný. Obecné deskriptory úrovně C2 zmiňují preciznost, vhodnost a lehkost použití jazyka (The Council of Europe 2001:37), které jsou analogií přesnosti, komplexnosti a plynulosti modelu CAF. Specifické deskriptory úrovně C2 rozpracovávají obecné deskriptory do úrovní produktivních a receptivních jazykových dovedností.

Zajímavým aspektem při měření nejvyšší pokročilosti jazykových kompetencí je jejich srovnávání s rodilými, případně bilingvními, mluvčími. Dokonce ILR stupnice (Interagency Language Roundtable scale) přímo označuje nejvyšší stupeň jazykové kompetence jako „rodilý nebo bilingvní“ mluvčí. Nicméně v SERR je podotknuto, že úroveň C2 (ačkoliv je nazvána jako „mastery“) neznamená kompetenci rodilého mluvčího (The Council of Europe 2001: 37). Z těchto důvodů je žádoucí se ve výzkumu zaměřit i na rozdíly v lexikální komplexnosti písemného projevu nerodilých mluvčích a rodilých mluvčí anglického jazyka.

Komplexnost, která je pravděpodobně nejproblémověji definovatelná z triády CAF modelu (Bulté a Housen 2012: 28), byla poslední komponentou, která přistoupila k původnímu dvourozměrnému modelu pracujícímu se složkami přesnosti a plynulosti (Brumfit 1984) až v 90. letech dvacátého století (Skehan 1998). Komplexnost jazykové produkce může být analyzována v rovině lexikologie, syntaxe, morfologie a syntaxe. Nicméně teoretické vymezení jazykové komplexnosti není v literatuře jednotné. Například Ellis and Barkhuizen (2005: 139)

interpretují jazykovou komplexnost jako míru použití sofistikovaného jazyka studenty. Možnou kategorizací komplexnosti je její členění na gramatickou a lexikální komplexnost (Bulté a Housen 2012: 27), přičemž gramatická obsahuje syntaktickou a morfologickou komplexnost, lexikální komplexnost lexikální hustotu (lexical density), lexikální rozmanitost/různorodost (lexical diversity), lexikální kompozionalitu (lexical composionality) a lexikální propracovanost/sofistikovanost (lexical sophistication) (ibid.: 28). Pro účely této práce je podstatná lexikální komplexnost v dimenzích lexikální rozmanitosti a propracovanosti.

Lexikální rozmanitost se zaobírá otázkou různorodosti lexikálních slov v textu (gramatická slova jsou opomíjena), tedy slov nesoucích sémantický prvek jako například slovní druhy podstatná jména, přídavná jména, slovesa či příslovce (Quirk 1985). Duran et al. (2004a) zdůrazňuje, že se jedná o více než jen rozsah slovní zásoby, o „umění" použití lexika. Pro měření lexikální rozmanitosti je možné využít různé metriky, kdy asi nejznámější a nejpoužívanější je TTR (Type/Token Ratio). TTR je ukazatel udávající počet sémanticky rozdílných slov v textu, kterým se říká „typ", dělený celkovým výskytem všech slov v textu, kterým se říká „token" (Czwenar 2013: 83). Jedná se o jednoduchý indikátor opakovanosti slov v textu. Výhodou ukazatele je jeho výpočetní jednoduchost, nicméně je kritizován pro svoji citlivost na délce zkoumaného textu (ibid; Kuiken a Vedder 2012). Z tohoto důvodu byly vyvinuty kompozitní indikátory (které mnohdy vycházejí z TTR) jako alternativa měření lexikální rozmanitosti. V diplomové práci jsou využity dva kompozitní indikátory VOCD-D a MTLD, přičemž u VOCD-D byl navíc pro snížení rozptylu výsledků daný „náhodností" vybíraných vzorků z textu počítán průměr ze tří samostatně provedených měření.

Dimenze lexikální propracovanosti měří nikoliv „šíři" lexikální komplexnosti jako v případě lexikální rozmanitosti, ale její „hloubku". Lexikální propracovanost je obvykle měřena pomocí srovnávání s externími seznamy slov (Bulté a Housen 2012), kterými jsou například Academic Word List (AWL), English Vocabulary Profile (EVP) nebo British National Corpus (BNC). V této práci byl využitý nástroj EVP a byl zkoumán výskyt slov s nízkou četností v textu.

Milton (2009: 137) poukazuje na problematičnost měření lexikální komplexnosti v důsledku její citlivosti na daný žánr a registr. Předpokládá (ibid.) vyšší lexikální rozmanitost (vyšší výskyt málo frekventních slov) u psané produkce oproti mluvené produkci. Také výzkum Yoon a Polio (2007: 288) ukázal závislost lexikální rozmanitosti na žánru, a to při zkoumání argumentačních esejů a narativních textů. Výzkum v oblasti lingvistické L2 komplexnosti také

poukazuje na rostoucí jazykovou komplexnost v čase (Bulté a Housen 2014; Duran et al. 2004) (předpokládá se ovšem stálé učení se jazyku, v opačném případě dochází ke stagnaci či poklesu jazykové kompetence) a důležitou roli postupného učení v oblasti dovednosti psaní (Tabari et al. 2021).

Podstatnou součástí této práce byl samostatný výzkum. Základní výzkumná hypotéza předpokládala menší lexikální komplexnost nerodilých mluvčí jazyka s nejvyšší možnou jazykovou kompetencí na úrovni C2, než je tomu v případě rodilých mluvčí. V rámci jednotlivých skupin uživatelů jazyka (L2 autoři komentářů a L1 autoři článků) se očekávaly srovnatelné výstupy. Další pomocnou hypotézou byl nižší předpokládaný počet málo frekventních slov u skupiny L2.

Soubor dat, které byly podrobeny analýze, dohromady obsahuje 40 textů, přičemž 20 textů bylo napsáno českými nerodilými mluvčími anglického jazyka na úrovni C2 a 20 textů rodilými mluvčími anglického jazyka. Texty L2 dospělých studentů byly homogenní co do jejich rozsahu (s průměrnou délkou 781 slov), žánru (ekonomie), cíle (napsání komentáře k publikovanému článku) a struktury (doporučená struktura komentáře). Texty L1 novinářů měly podobné parametrické vlastnosti v rozsahu (s průměrnou délkou 654 slov), žánru (ekonomie) a cíli (článek v časopise, novinách nebo na Internetu). Data tak mohla být analyzována ve svých skupinách (L2 a L1) samostatně a také vzájemně mezi sebou. Pro analýzy byl využit automatizovaný webový nástroj Text Inspector, jehož bližší popis je také součástí práce.

Metodologie práce zahrnovala manuální přípravu textů pro jejich automatické zpracování, tzn. uvedení souborů do podoby prostého textu: vymazání úvodní strany, grafů, tabulek, referencí, poznámek pod čarou, nepatřičných mezer a jakýchkoliv nerelevantních slov (jako například uvozující slova „article“ a „commentary“). Podrobná analýza byla provedena nejdříve u L2 textů, následně L1 textů a nakonec bylo provedeno srovnání mezi oběma skupinami. Jednoduchá kvantitativní analýza obsahuje indikátory počtu vět, „token“ počtu, „type“ počtu, indikátor TTR, průměrnou délku vět a počet číslovek. Analýza v oblasti lexikální rozmanitosti se zaměřila na kompozitní indikátory VOCD-D a MTLD a v oblasti lexikální propracovanosti na měření málo frekvenčních slov pomocí seznamu slov z EVP. Grafy a tabulky práce byly vytvořeny prostřednictvím tabulkového procesoru Excel.

Výsledky kvantitativní analýzy potvrdily předpoklad, že ekonomické texty obsahují často číslovky. Číslovky se nacházely v 39 textech, výjimkou byl jediný komentář, který číslovku

neobsahoval. V procentuálním vyjádření se jednalo o výskyt ve výši 2,2% „tokenů“ u článků a 0,8% u komentářů. Důležitost čísel v ekonomických textech je zřejmá a ve velké většině se jedná o jev „hapax legomenon“. Z tohoto důvodu číslovky snižují hodnotu TTR, nicméně v souboru dat byl tento vliv malý. Lexikální rozmanitost komentářů, měřena indikátorem TTR, byla nižší ($\bar{x}$ = 0,38, min = 0,31, max = 0,48, SD = 0,05) než-li tomu bylo v případě článků ($\bar{x}$ = 0,50, min = 0,39, max = 0,61, SD = 0,06). Jedním z důvodů je vyšší využívání synonymie v případě L1 mluvčích. Počet vět a počet „tokenů“ potvrdil vysokou homogenitu textů v parametru jejich délky, a to především u komentářů.

Analýza v oblasti lexikální rozmanitosti potvrdila její vyšší hodnotu v případě L1 textů. Výsledky obou kompozitních indikátorů VOCD-D a MTLD byly srovnatelné. Rozdíl lexikální rozmanitosti v případě komentářů a článků byl přibližně třetinový. Výsledky VOCD-D indikátoru byly dále porovnány s výsledky dřívějšího výzkumu, který provedl Duran et al. (2004a: 227-237). Pro obě analyzované skupiny (L2 mluvčí a L1 autoři akademického textu) vyšly výsledky D indexu v této práci vyšší, než byl udávaný interval (ibid.). Nicméně statistická odchylka není významná a z důvodu neznalosti konkrétního složení původních zkoumaných skupin a specifičnosti zkoumaných skupin v této práci (L2 mluvčí nejvyšší úrovně, akademický text; L1 novinář s nejvyšší možnou jazykovou kompetencí, odborný akademický text) se dá vyvodit závěr, že výsledky spadají do rámce původního výzkumu. Vyšší výsledky VOCD-D v případě článků mohou být také ovlivněny větší skupinovou spoluprací na konečném výsledku – je obvyklé, že v „renomovaných“ časopisech a novinách probíhají nezávislé korektury textů před jejich publikováním.

Lexikální propracovanost textů byla měřena prostřednictvím rozdělení slov do kategorií a jejich srovnání se seznamy slov v EVP. Výhodou využití EVP je rozdělení slov do kategorií SERR, které byly využity pro zadefinování jazykové kompetence nerodilých mluvčí, a také to, že EVP dokáže pracovat i s kolokacemi a frázemi. Naopak jako problematické se jevila vysoká míra slov textu, které nebyly uvedené v seznamu slov EVP (tzv. unlisted types). Podrobným zkoumáním bylo zjištěno, že důvodem této skutečnosti jsou čtyři základní kategorie neuvedených slov: čísla, vlastní jména, zkratky (převážně akronymy) a specifické ekonomické termíny. Všechny tyto kategorie jsou typické pro ekonomický text a potvrzují tak citlivost lexikální komplexnosti textu na žánru, na což již bylo poukázáno v dřívějších výzkumech (Milton 2009; Yoon a Polio 2007). Analýza lexikální propracovanosti prokázala obdobné statistické rozdělení podle typů slov (A1-C2) u článků i komentářů. Celkový trend rozdělení

podle typů slov je klesající od A1 do C2 v souladu se Zipfovým zákonem. Počet slov s nízkou četností v textu, měřený pomocí typu C2, je obdobný v obou typech zkoumaných textů. Z tohoto důvodu hypotéza o vyšším výskytu slov s nízkou četností v textu u článků (L1 autoři) nemohla být potvrzena. Výzkum prokázal menší lexikální komplexnost nerodilých mluvčí jazyka s nejvyšší možnou jazykovou kompetencí na úrovni C2 jen v dimenzi lexikální rozmanitosti.

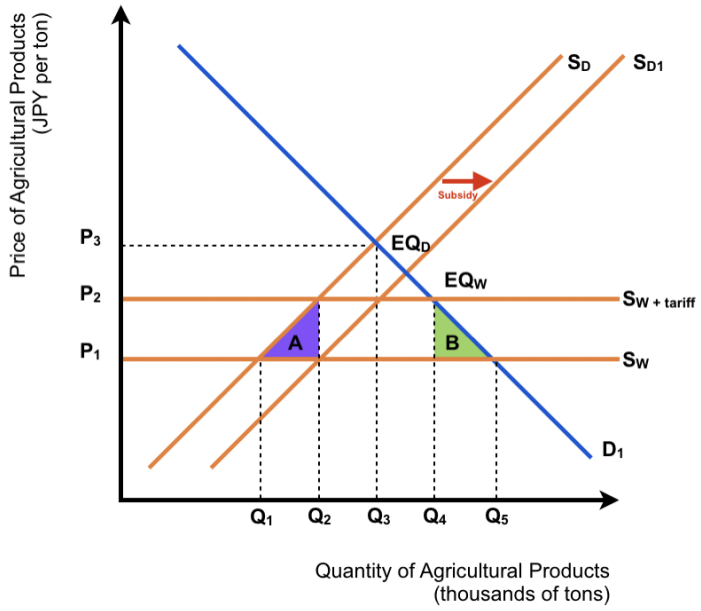
Na výsledky práce mohly mít vliv některé vlivy. Předně lexikální komplexnost může být ovlivněna typem úlohy. Dalším faktorem majícím vliv je, že studenti píší komentář vybraného ekonomického textu a mohou tak být ovlivněni výběrem slovní zásoby s tendencí opakovat „renomované“ autory. Tento tzv. „priming effect“ ve věci ovlivnění výběru slovní zásoby u textů komentářů nebyl předmětem výzkumu. Možným omezením s dopady na výsledky výzkumu je i množství a délka textů, která nedosahovala velké rozsáhlosti. S větším množstvím dat by výsledky výzkumu dosáhly lepší míry spolehlivosti. Dalším vlivem je žánr textů a problematičnost možného srovnání s podobně zaměřenými výzkumy s texty jiného žánru, neboť každý výzkum v oblasti lexikální komplexity je specifický a liší se v jednotlivých parametrech. Závěrečným ovlivněním výsledků byl výběr nástroje EVP, kde byl prokázán problém se slovní polysémií a množstvím slov, která nejsou v databázi EVP obsažena.

Zajímavým rozšířením práce by byl výzkum další L2 skupiny na úrovni C2, tentokráte s jinou národní příslušenství, a porovnání výsledků s českou L2 skupinou. Také výzkum „priming“ efektu by byl užitečným doplněním předkládaného výzkumu. V neposlední řadě by byl přínosným a zajímavým vědecký výzkum zkoumající rozdíly u slov s nízkou četností v psané a mluvené produkci, a v produktivní slovní zásobě (lingvistické kompetence psaní, mluvení) a receptivní slovní zásobě (čtení a poslech).

7 Appendix 1 – L2 English Sample Text (1C)

The article outlines the main impacts of the Regional Comprehensive Economic Partnership (RCEP) on Japan and its relations to other signatories. While described in the article as a “free trade zone”, this is not entirely accurate. A free trade area, in economic terms, means an agreement between signatory countries which effectively eliminates tariffs, whereas the RCEP only cuts some. Additionally, the RCEP includes provisions characteristic of a common market, in that the signatory countries have agreed to “common rules in areas such as e-commerce and intellectual property”. Still, the RCEP best fits the model of a preferential trading area, as it does not eliminate all tariffs nor establish common external barriers or entirely standardize product regulations.

fig. 1: Japanese Agricultural Market



With tariffs eliminated, the price of RCEP member states’ agricultural goods in Japan effectively decreases from P_2 to P_1. Where before, domestic producers (S_D) were

willing and able to supply Q_2 of agricultural products at a price lower than the world price of P_2, domestic producers will now only supply Q_1. As a result, RCEP producers will supply [Q_5-Q_1] agricultural products, increasing their revenues. Revenues of domestic producers, in contrast, will fall to [P_1×Q_1], and a loss of producer surplus of [A] will be incurred. Consumer surplus will, however, increase by [B] and [Q_5-Q_4] more goods will be demanded.

These results, however, could be mitigated. Assuming Japan were to subsidize domestic agricultural production, which it already does to a significant extent, with a per-unit subsidy equal to the decrease in end consumer price the tariff's elimination leads to, the market would return to the initial situation. Through this subsidy, domestic supply of agricultural goods would shift from S_D to S_{D1}, meaning Japanese agricultural producers would, again, supply Q_2 goods. The benefits to consumers would remain, however, as the lower price of P_1 would be maintained and Q_5 of goods still demanded, though the Japanese government would face lower net revenues and a high opportunity cost, having used tax-payer money for subsidies rather than, for example, social services or infrastructure.

Additionally, it could be argued that this tariff's effect will be lesser than indicated above simply due to Japanese consumers not seeing foreign goods as substitutes. Setting aside the issue of physiological differences between crops from, for example, Cambodia and Japan, Japanese consumers may tend to purchase local goods rather than imports, causing the actual quantity of foreign agricultural goods purchased to be less than Q_2, as consumers would see their utility as lower than that of higher-priced domestic products. Additionally, the RCEP agreement does not affect tariffs

on “rice, wheat, dairy products, sugar, and beef and pork”, calling into question its significance for agricultural imports into Japan, when such significant product categories have been excluded.

The second issue the RCEP addresses, according to the article, is that of Japan’s industrial exports, specifically making the “supply chains more efficient”. Assuming the below situation, where Japan produces economic technological goods with supply chains of its firms extending to other RCEP countries, where it is more efficient to produce components for them, multiple conclusions can be drawn.

fig. 2: Theoretical RCEP Market for Technological Goods (Production Possibilities)

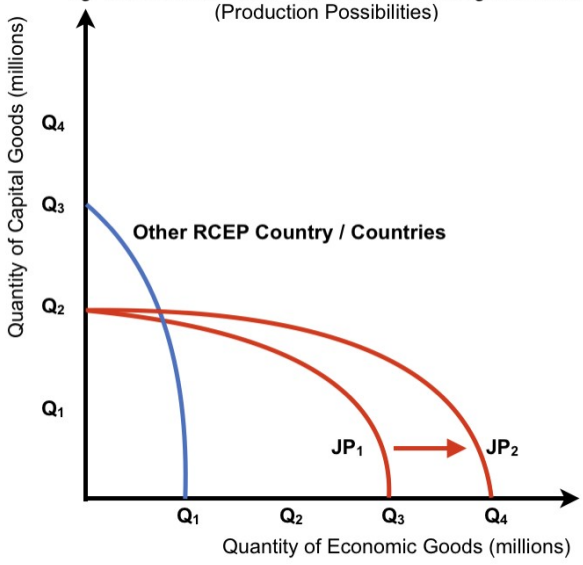


fig. 3: Production Outcomes from fig. 2

	Japan	Other RCEP
Capital Good Production (max)	Q_2	Q_3
Opportunity Cost (of Cap. Goods in terms of Econ. Goods)	3/2 Q_1	1/3 Q_1
Econ. Good Production (max)	Q_3	Q_1
Opportunity Cost (of Econ. Goods in terms of Cap. Goods)	2/3 Q_1	3 Q_1
NOTE	The following applies: $Q_2 = 2\ Q_1$, $Q_3 = 3\ Q_1$, $Q_4 = 4\ Q_1$,	

In this situation, both countries have varying opportunity costs for capital and economic goods, with Japan having a comparative advantage in economic goods (by producing them at a lower opportunity cost of 2/3 Q_1) and Other RCEP countries having such an advantage (1/3 Q_1) in capital good production. With the RCEP limiting tariffs (up to 91.5% on Japanese industrial goods) and establishing “common rules” for e-commerce, costs for import both in terms of fees and administrative procedures

are limited. This allows firms to operate more cost-effectively transnationally and to exploit the comparative advantages of both regions. It is also possible that with Japanese firms having access to a greater amount of capital goods at a lower price, Japan's production possibilities for economic goods made from those capital goods will increase (from JP_1 to JP_2). In the long run, this could allow Japanese firms to produce Q_4 economic goods while other RCEP countries manufacture Q_3 capital goods, rather than both regions producing both, increasing allocative efficiency. The RCEP standardizing intellectual property rights could also further this goal. Before, firms had a stronger incentive to keep trade secrets in their home country, as moving manufacturing abroad to countries like China meant risking semi-legal appropriation of their patents and manufacturing of copycat products. With the RCEP in place, firms could be more confident this will not be the case.

8 Appendix 2 – L1 English Sample Text (1A)

Japan, China and other Asian nations sign massive RCEP trade pact

Asia-Pacific countries including Japan, China and the 10 members of ASEAN signed a regional trade deal on Sunday covering nearly a third of the global economy, wrapping up eight years of negotiations following the withdrawal of India.

The 15 signatories to the Regional Comprehensive Economic Partnership reached the agreement, aimed at cutting tariffs and establishing common rules in areas such as e-commerce and intellectual property, during a virtual leaders' summit.

RCEP — also including Australia, New Zealand and South Korea — will create Asia's biggest free trade zone encompassing about a third of the world's population.
It will be Japan's first trade deal with both China, its largest trading partner, and South Korea as negotiations for a trilateral pact have yet to be concluded.
Speaking to reporters after signing the deal, trade minister Hiroshi Kajiyama said the 15 countries were seeking to wrap up domestic procedures quickly and put the pact into effect "as quickly as possible."

"Through the tariff removals, I believe there'll be a major impact on improving Japan's exports and making the region's supply chains more efficient," Kajiyama said. "I strongly believe we are building free and fair economic rules through introducing new rules on data free flows and the banning of demands for technology transfers, as well as the protection of intellectual property."

Supporters of the trade pact, which covers 2.2 billion people with a combined GDP of $26.2 trillion, said it will bolster pandemic-weakened economies by reducing tariffs, strengthening supply chains with common rules of origin, and codifying new e-commerce rules.

"The completion of negotiations is a strong message affirming ASEAN's role in supporting the multilateral trade system," Vietnamese Prime Minister Nguyen Xuan Phuc said as he hosted the virtual signing ceremony. The agreement will contribute to "developing supply chains that have been disrupted due to the pandemic as well as supporting economic recovery," he said.

Negotiators pushed the deal across the finish line after India surprised participants late last year by abandoning the agreement. Prime Minister Narendra Modi said he pulled out over concerns about how RCEP would affect the livelihoods of Indians, particularly the most vulnerable. India, though, will be allowed to rejoin the trade pact.

"The clause allowing India to join at a later date is symbolic and shows China's desire to build economic bridges with the region's third-largest economy," said Shaun Roache, Asia Pacific chief economist at S&P Global Ratings.

Whether RCEP changes regional dynamics in favor of China depends on the U.S. response, experts said. The agreement underscores how U.S. President Donald

Trump’s 2017 decision to withdraw from a different Asia Pacific trade pact — the Trans-Pacific Partnership — diminished America’s ability to offer a counterbalance to China’s growing regional economic influence.

That challenge will shift to President-elect Joe Biden. Still uncertain is how the Biden team will approach trade deals and whether it tries to re-enter the 11-nation TPP. RCEP was expected to fall significantly short of the revised TPP or Japan’s trade deal with the European Union in cutting tariffs.
Despite RCEP’s historic size, it is surpassed by other major trade deals in the level of market access. The deal will eliminate tariffs on 91% of goods compared with 99.9% for the revised TPP.

Japan will eliminate 61% of tariffs on agriculture imports from ASEAN nations, Australia and New Zealand, 56% for China, and 49% for South Korea, while maintaining tariffs on five product categories — rice, wheat, dairy products, sugar, and beef and pork — to protect domestic farmers.

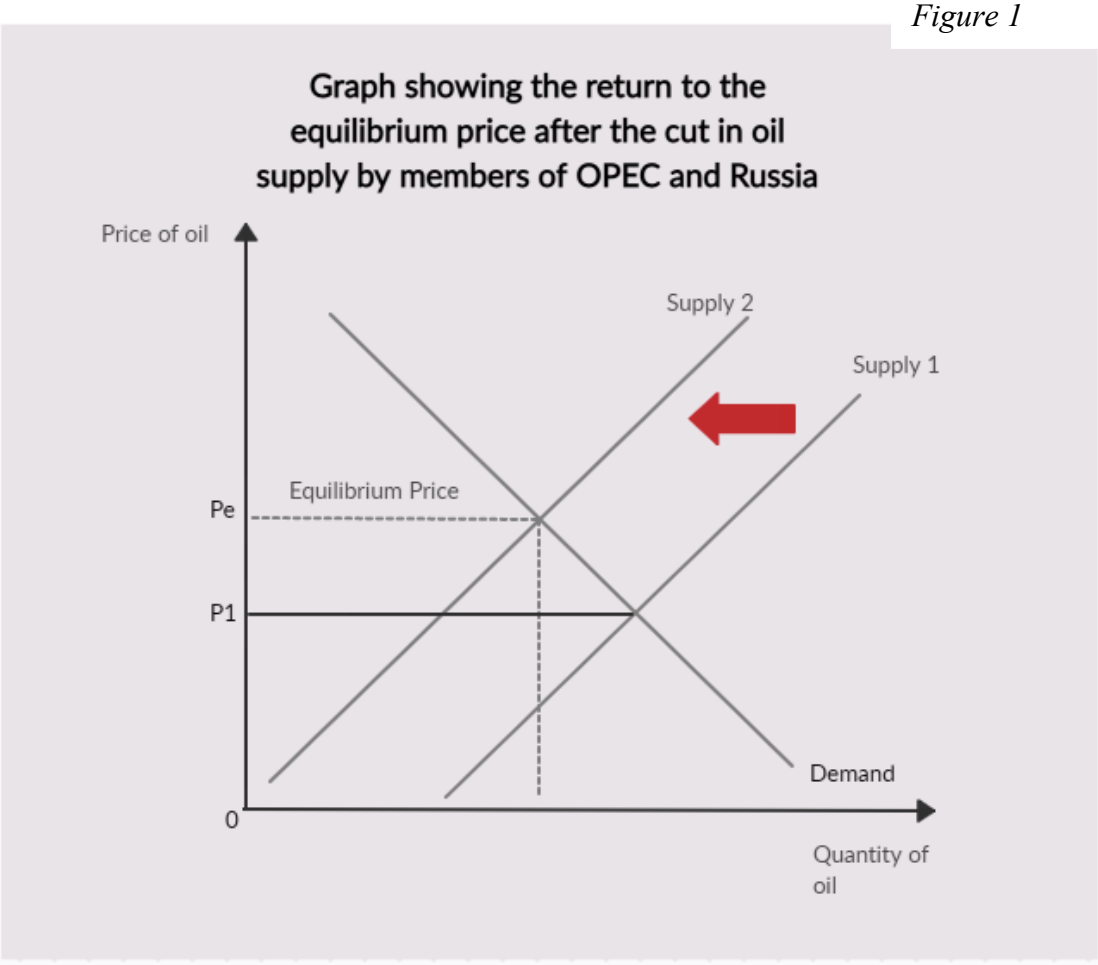
Meanwhile, the other countries will cut 91.5% of tariffs on Japanese industrial exports.

9 Appendix 3 – L2 English Sample Text (20C)

Due to a number of incidents which occurded worldwide, including the current COVID pandemic, the demand for oil in the world decreased. The **interdependence** of economies is an essential cause to the global drop in demand of oil. Not only the pandemic, which affects all the stakeholders taking part in the oil market, but a number of local events can massively influence the global economy.

As a mean to reach the previous demand levels, many international groups, including OPEC (The Organization of Petroleum Exporting Countries), which includes countries from the Middle East and Russia along with its allies have debated on reducing the supply. This solution to the drop in demand would be ideal, as it is a natural procedure in which to deal with a excess of supply. On the following graph, we can see the effects of this policy, which was discussed within the OPEC group and Russia. The initial price P_1 occurs during a excess of supply and the decrease in the supply would mean the return to an equilibrium price. At this point, MC=MB, meaning the market would be closer towards achiveing allocative efficiency.

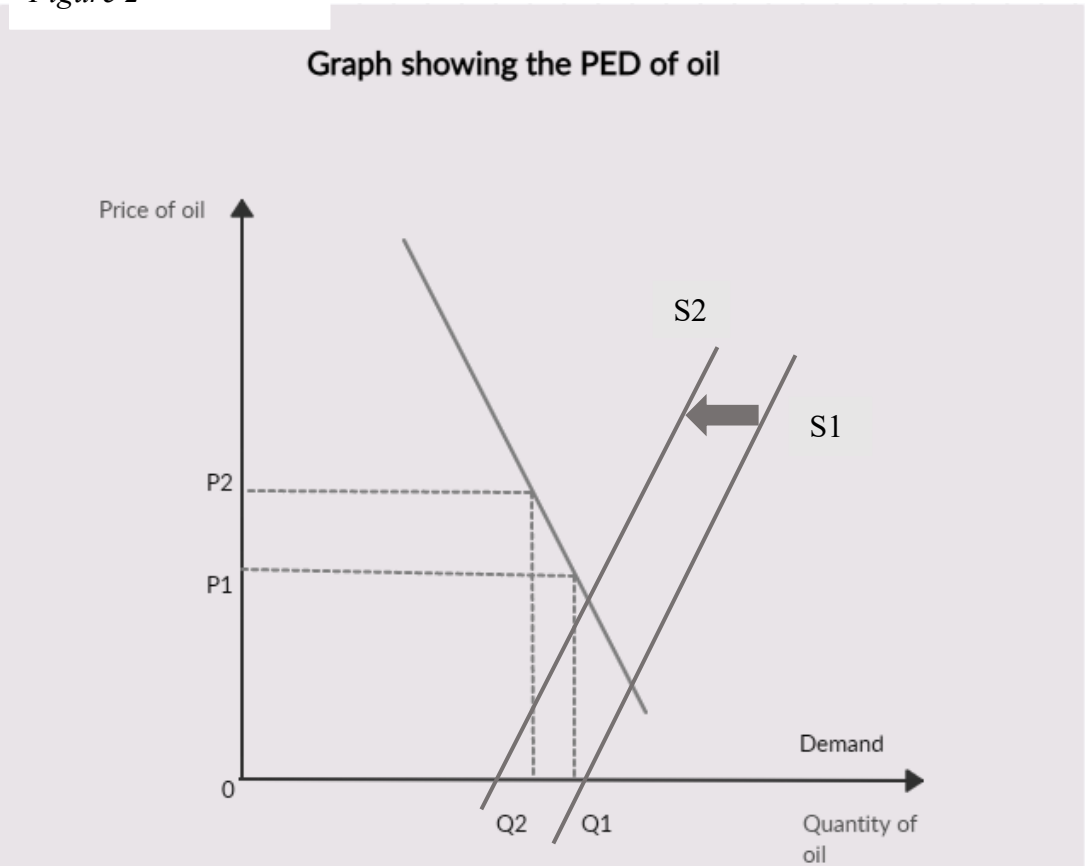
Figure 1



Nevertheless, as stated in the article, a force majeure was lifted on an oilfield in Sharara, resulting in an increases in production and therefore in increases in supply of oil. Although supply drops are not a viable option, as Giovanni Staunovo says, „It is all about ending production disruptions“, with the current demand being much lower than usual, the sudden return to the original production of oil would cause excess supply in the oil market, which would mean factors of production are not being used efficiently. Furthermore, Lybia's increases in oil production caused problems for the members of OPEC and Russia, because as previously mentioned, their target is to lower the supply. As all the countries, importing or exporting oil are **interdependent,** the increases in supply of oil by Lybia will cause problems for the countreis exporting oil. It would therefore be ideal to include Lybia in a deal, ensuring they will limit their oil production, if oil demand recovery continues to struggle to ensure that they will be able to control the price.

On the other hand, the shortage of oil is discussed as well with regards to the production disruptions in multiple countries. The labor strikes, along with the natural disaster, Hurricane Delta in the USA and restrictions related with COVID had all caused the decrease in oil production and therefore caused a shortage. As the demand for oil is worldwide, the countries all **interdependent** on each other regarding oil supply. The shortage of oil is a problem, as it causes prices to rise. As oil is a neccessity, its PED is inelastic and changes in price wont affect the demand by much. Although very advantegous for countries, which produce oil such as Norway and the USA, this would mean that countries which import oil would have to pay a still relatively high price for a small change in quantity, which is very disadvantegous for countries, which import large quantities of oil. As seen on the following graph, the change from Q1 to Q2 is smaller than the change from P1 to P2. The graph shows, that countries importing would have to pay price P2, if they would want to maintain steady imports of oil.

Figure 2



As the global oil market is **interdependent**, countries exporting oil would also suffer, as their losses in revenue would be equal to (Q1 x P1) – (Q2 x P2). Therefore, it is desirable for both parties to maintain the production outputs. As mentioned by Giovanni Staunovo, the solution is in ending production disruptions. Nevertheless, the disruption in oil production was in most cases caused by natural disasters (non-price determinant), as seen on Graph 2. In the case of the drop of price of the Brude crude and West Texas Intermediate, both were caused by the Hurricane Delta and by stricter restrictions regarding COVID. Right now, the global pandemic is causing stricter restrictions, causing further disruption oil production. For the production level to remain same, workers should be allowed into workplace to maintain the production level.

The OPEC and Russia are certainly trying to deal with the current situation in a sustainable way. The cut in supply by all the countries would help get closer to allocative efficiency and deal with the excess supply and achieve fairness in such a deal. Nevertheless, due to the COVID restrictions and other production disruptions, the supply of oil produced in the USA and Norway has decreased, causing shortage. Both governments have not dealt with the situation ideally, as production disruption will cause high prices, which will hurt both importers and exporters.

10 Appendix 4 – L1 English Sample Text (20A)

Oil prices fell on Monday as force majeure at Libya's largest oilfield was lifted, a Norwegian strike affecting production ended and U.S. producers began restoring output after Hurricane Delta.

Brent crude fell 52 cents, or 1.2%, to $42.33 a barrel by 1052 GMT and U.S. West Texas Intermediate was down 58 cents, or 1.4%, at $40.02.

"It's all about ending production disruptions ... (which) are not helpful in a period with ongoing demand concerns," said UBS oil analyst Giovanni Staunovo

Production in Libya, a member of the Organization of the Petroleum Exporting Countries (OPEC), is expected to rise to 355,000 barrels per day (bpd) after force majeure at the Sharara oilfield was lifted on Sunday.

Rising Libyan output will pose a challenge to OPEC+ - a group comprising OPEC and allies including Russia - and its efforts to curb supply to support prices.

"If oil demand recovery continues to struggle due to new or stricter COVID-related mitigation measures, the (OPEC+) producer group may need to reconsider the planned tapering of their voluntary supply cuts," said BNP Paribas analyst Harry Tchilinguirian.

Front-month prices for both contracts gained more than 9% last week in the biggest weekly rise for Brent since June. But both fell on Friday after Norwegian oil companies struck a deal with labour union officials to end a strike that had threatened to cut the country's oil and gas output by close to 25%.

Hurricane Delta, which dealt the greatest blow to U.S. Gulf of Mexico energy production in 15 years, was downgraded to a post-tropical cyclone at the weekend.

Workers headed back to production platforms on Sunday and French oil major Total was working to restart its 225,500 barrel per day Port Arthur refinery in Texas.

Prices were also pressured by a jump in new COVID-19 cases, which has raised the spectre of more lockdowns.

Infections are at record levels in the U.S. Midwest and in Britain Prime Minister Boris Johnson is expected to announce new measures on Monday while Italy is preparing fresh nationwide restrictions.

Goldman Sachs, meanwhile, said that the outcome of the U.S. presidential election would not impact its bullish oil and natural gas outlook and that an overwhelming Democratic victory could be a positive catalyst for these sectors.

11 Appendix 5 – Text Inspector´s Print-Screens for EVP (10A)