

**UNIVERSITY OF TARTU**  
**DEPARTMENT OF ENGLISH STUDIES**

**THE COMPILATION AND LEXICOGRAMMATICAL ANALYSIS OF  
AN ESTONIAN SPOKEN MINI-CORPUS OF ENGLISH AS A  
LINGUA FRANCA**

**MA thesis**

**MERLI KIRSIMÄE**

**SUPERVISOR: *Lect.* JANE KLAVAN, *PhD***

**TARTU**

**2017**

## ABSTRACT

Although the research of English as a lingua franca (ELF) is a relatively recent development, it has nevertheless rapidly become an acknowledged field of study that also causes considerable debate among linguists and English language teaching professionals. Over the past decade or so, ELF scholars have compiled several spoken corpora to explore the features of ELF and have described them from the perspective of pronunciation (Jenkins 2000, Walker 2010), pragmatics (Kaur 2011, Seidlhofer 2011, Walkinshaw & Kirkpatrick 2014) and lexicogrammar (Seidlhofer 2004, Dewey 2007b, Breiteneder 2009, Önen 2014). Jenkins (2007) and Seidlhofer (2011) have both elaborated on the topic of attitudes towards ELF and identity regarding ELF. The sociolinguistic aspects of ELF have also been explored in the Estonian context (Soler-Carbonell 2014, 2015), but the lexicogrammatical aspects have not. To fill that gap, a mini-corpus of spoken ELF was compiled for this thesis to see whether any of the lexicogrammatical features identified by previous ELF research occur with Estonian ELF speakers as well. Another aim of the thesis was to discuss the possible pedagogical implications of these findings.

The Introduction of the thesis presents a brief overview of previous research into ELF, its relation to pedagogy, the aims of the thesis and why the thesis is of relevance. The first chapter provides a more detailed account of the development and relevance of ELF research along with how spoken corpora have been and are compiled and used to describe ELF. In addition to that, the points of contact between ELF and English language teaching and the need for reconciliation between the two paradigms are discussed together with a short summary of the criticism that has been directed at the concept of ELF. The second chapter provides a description of the compilation process of the spoken ELF mini-corpus, the participants involved and the transcription and annotation of the interviews that were recorded to compile the corpus. The corpus is analysed from the perspective of innovations in the use of articles, prepositions and collocations with verbs of high semantic generality. Comparisons are drawn with similar studies by other scholars. The thesis ends with a discussion on the underlying motives for the innovations found in the corpus and their pedagogical implications. The limitations and possibilities for further research are also outlined together with a summary of the findings of the study in light of the theoretical framework presented in the thesis.

## TABLE OF CONTENTS

ABSTRACT .....	2
LIST OF ABBREVIATIONS .....	4
INTRODUCTION .....	5
1. LITERATURE REVIEW .....	10
1.1 ENGLISH AS A LINGUA FRANCA: RELEVANCE AND DEVELOPMENT .....	10
1.2. SPOKEN CORPUS RESEARCH AND ELF CORPORA .....	13
1.3 ELF AND PEDAGOGY .....	17
1.4 CRITICISM OF ELF.....	20
2. THE COMPILATION AND LEXICOGRAMMATICAL ANALYSIS OF THE SPOKEN ELF MINI-CORPUS .....	23
2.1 DATA COLLECTION.....	23
2.2. PARTICIPANTS .....	25
2.3 TRANSCRIPTION AND ANNOTATION .....	27
2.4 RESULTS .....	29
2.4.1 ARTICLES.....	29
2.4.2 PREPOSITIONS .....	32
2.4.3 COLLOCATIONS WITH WORDS OF HIGH SEMANTIC GENERALITY .....	35
2.5 DISCUSSION .....	37
CONCLUSION .....	43
REFERENCES .....	47
APPENDIX 1: INTERVIEW QUESTIONS.....	52
APPENDIX 2: PARTICIPATION INFORMATION SHEET AND CONSENT FORM ..	53
APPENDIX 3: PARTICIPANT INFORMATION QUESTIONNAIRE .....	55
APPENDIX 4: TRANSCRIPTION CONVENTIONS .....	56
APPENDIX 5: EXAMPLE INTERVIEW .....	61
RESÜMEE .....	71

## **LIST OF ABBREVIATIONS**

ACE – Asian Corpus of English

CEFR – the Common European Framework of Reference for Languages

EFL – English as a Foreign Language

ELF – English as a Lingua Franca

ELFA – Corpus of English as a Lingua Franca in Academic settings

ELT – English Language Teaching

ENL – English as a Native Language

ESOL – English for Speakers of Other Languages

L1 – First Language / Mother Tongue

L2 – Second Language

NS – Native Speaker

NNS – Non-Native Speaker

OCD – Oxford Collocations Dictionary for Students of English

SE – Standard English

VOICE – Vienna-Oxford International Corpus of English

## INTRODUCTION

Over the past two decades, research into English as a lingua franca (ELF) has steadily developed into a thriving field. ELF has been studied from the perspective of pronunciation, lexis, grammar and pragmatics (see, for example, Dewey 2007b, Jenkins 2000, Jenkins 2007, Kaur 2011, Önen 2014, Seidlhofer 2004, Seidlhofer 2011, Walker 2010, Walkinshaw & Kirkpatrick 2014). This research proceeds from two basic assumptions: first, that English does not *belong* to native speakers, since there are now more non-native than native speakers of English (Cogo & Dewey 2006, Graddol 2006, Seidlhofer 2011) and second, that ELF is not a *language* as such, but rather “a means of communication not tied to particular countries and ethnicities, a linguistic resource that is not contained in, or constrained by, traditional (and notoriously tendentious) ideas of what constitutes ‘a language’” (Seidlhofer 2011: 81). Jenkins (2012: 487) argues that ELF is something that should be ‘additionally acquired’ by L1 English speakers too, to be able to communicate with L2 English, that is, ELF speakers. In general, the appreciation of non-native speakers of English and the role that they have played in the globalisation of English (Haberland 2013: 195) is called for by ELF scholars.

One of the ways to credit L2 English (or ELF) speakers is to not consider them learners by default, but rather see them as users of the language in their own right (Seidlhofer 2011: 24). This is certainly one of the main agendas of ELF scholars along with reconsidering what in mainstream English language teaching (ELT) is seen as erroneous language and seeing this as possibly innovative (see, for example Dewey 2007b: 38). The idea is further echoed in Kohn (2015), where a reconciliation between ELF and ELT is called for. This underlying idea has triggered the present thesis as well: how can ELF be brought into the ESOL (English for speakers of other languages) classroom – not so much as a language to be taught but as a concept to be communicated to the learners and something that would help them become more confident speakers of the language outside the

classroom. The key to the matter probably lies in firstly raising the teachers' awareness of ELF. Being an English teacher myself, and seeing how, especially with adult learners (usually at higher levels, but not necessarily), communicative efficiency generally does not suffer due to lexicogrammatical innovations, I always try to draw the learners' attention to their communicative success and make them understand that there is no point in focusing all their energy on 'errors'. To be able to increase the credibility of that argument and show that the 'errors' can be considered innovations, as much data as possible should be gathered in the form of authentic speech events and compared across language users. Hence, my interest in collecting such data and exploring the ELF features in more detail.

Seidlhofer (2011: 187) argues that the learners of English as a foreign language (EFL) become ELF users as they move from the classroom setting into the actual communication settings. However, developing the understanding that ELF is an existing reality and that it is not simply imperfect English or an interlanguage (Jenkins 2006: 42-43), in learners, and firstly, in teachers, needs support from real language use. The 'interlanguage' issue has been explored often in ELF research, and continually, the conclusion is reached that ELF does not qualify as interlanguage, which is seen as a constantly changing concept that the *learners* navigate in, trying to rid themselves of the recurrent errors (Seidlhofer 2011: 89). While ELF emphasises the concept of *users*, the notion of interlanguage has *learner* rather than *user* embedded in itself. The question that ELF researchers pose as a counterargument is why, then, this defective language proves so effective (Seidlhofer 2011: 186). Also, the interlanguage concept leaves little room for these learners to construct their *own* L2, since the concept is firmly oriented towards the standard target language and learning is successful only if the result is a "clone" of the standard version of the target language (Kohn 2015: 61). Kohn's "my English" notion, on the other hand, is based on a social constructivist model

which sees acquiring English as “creatively constructing your own version of it in your mind, in your heart, and in your behaviour” (Kohn 2015: 56-57).

As indicated above, descriptions of authentic language use can add to the credibility of claims in favour of viewing ELF features as innovations. Spoken corpora provide invaluable insight into how a language is actually used for real-life communication. Among the corpora that focus on ELF are the Vienna-Oxford International Corpus of English (VOICE), the Corpus of English as a Lingua Franca in Academic Settings (ELFA), the Asian Corpus of English (ACE) and smaller-scale corpora compiled for specific (lexicogrammatical) research purposes (Dewey 2007b, Önen 2014). The growing amount of data allows researchers to determine any features that are reoccurring across different settings and L1 backgrounds. On the other hand, despite identifying features that are recurrent in spoken ELF data, some ELF researchers also emphasise the ‘fluidity’ of ELF (Seidlhofer 2009, Cogo & Dewey 2012, Dewey 2012), recognising that ELF communication contexts “are usually characterized by a high degree of linguacultural diversity, routinely resulting in highly variable and creative use of linguistic resources” (Dewey 2012: 163). This fluidity, nevertheless, might let us discover certain patterns “if we focus on ELF encounters in a certain community of practice, or groups of people who come together around a mutual aim” (Hynninen 2014: 298).

Although ELF has obtained world-wide recognition, no excessive research on its lexicogrammatical features has been carried out in Estonia. Soler-Carbonell (2014) has conducted research focusing on the sociolinguistic aspects of ELF in Estonia among L1 Estonian and L1 Russian speakers rather than looking at “what linguistic form it takes” (2014: 249). He has also explored English as an academic lingua franca in Estonia and the attitudes of students at one Estonian university towards it (Soler-Carbonell 2015). Advancing from the belief that any record and description of language as it is authentically

used adds to the credibility of the existence of a specific variety or phenomenon, this study attempts at making a small contribution towards establishing a description of ELF features or describing its fluidity. If ELF is to be described as a world-wide phenomenon, this account can undoubtedly benefit from any data collected from places where it has not been done previously.

The thesis has several aims. The first aim is to describe and discuss the compilation of a spoken mini-corpus that is hopefully the start of a much larger corpus that it could be developed into. Another aim is to explore the use of articles, prepositions and collocations with verbs of high semantic generality by Estonian ELF speakers and see whether there are features that coincide with those identified by Seidlhofer (2004), Dewey (2007b) and Önen (2014). Consequently, the following research question was formulated for the thesis: do Estonian ELF speakers follow similar patterns of article, collocation and preposition use as identified for English as a lingua franca? Furthermore, the thesis also hopes to contribute to the discussion of ELF and pedagogy. Nevertheless, it must be emphasised that the thesis does not wish to allude that ELF is the only possible way to go for when teaching, learning and using English, it simply explores the ways in which ELF has been addressed and researched and how it might be of use to our understanding of the changing role of English globally. I, as an English teacher, am in no way free from the identity and attitude issues that Jenkins (2007) discusses in length, and thus, while perhaps inclining towards preferring ELF to other treatments of English, I am certainly not its firm advocate and acknowledge the criticism directed at viewing ELF as an all-embracing neutral term (see section 1.4 below for some of the criticism raised against ELF).

The thesis is structurally divided into two core chapters. The first chapter provides an overview of relevant literature in the field. It starts with describing the development of ELF as a research discipline and its relevance. Next, the importance of using spoken corpora

in linguistic research is briefly discussed along with describing, more specifically, and in greater detail, the ELF corpora thus far compiled and used for research. The first chapter also introduces the studies relevant for comparison for this thesis. As this research is not only concerned with investigating the patterns of article, preposition and collocation use of Estonian ELF speakers, but also with the implications of these findings to teaching, the points of contact between ELF and ELT are taken up and examined in the first chapter. Finally, as ELF is a field that generates debate and heated discussion among both linguists and teaching professionals, the last part of the first chapter deals with criticism of ELF.

The second chapter focuses on the Estonian spoken ELF mini-corpus compiled for this thesis and the empirical findings of the study. It presents the methodological frame of reference for putting the corpus together and gives details on the data collection process and the participants involved. Next, the transcription and annotation processes are explained, followed by the presentation of the results of the study. In the Discussion section, the results of article, collocation and preposition use are examined further with a discussion of the possible underlying motives for the innovative features. The second chapter also elaborates on the pedagogical implications that the results of this study could have in the Estonian context and on a wider scale as well. In addition, the limitations of the study are presented along with proposals for future research both using the data already collected for this thesis and building upon the limitations to gather more data. The second chapter is followed by the Conclusion, which summarises the research findings. There are 5 Appendices at the end of the thesis, which include the participation information sheet and consent form, the participant information questionnaire, questions used for the interviews, transcription conventions and a transcribed example interview.

## **1. LITERATURE REVIEW**

The aim of this chapter is to present the theoretical background of the thesis and place it within the wider discourse of ELF studies, spoken corpus research and ELT. The first subsection of the chapter provides an account of the main principles guiding ELF research and how ELF has developed into the phenomenon that it is today. Next, the most important spoken ELF corpora are introduced along with reasons to opt for spoken corpora compilation and research when investigating ELF. The distinction between learner and ELF corpora is also explained. The points of contact between ELF and pedagogy are discussed further on in the chapter with an emphasis on the importance of raising awareness of ELF among teachers. Finally, the chapter ends with an overview of criticism directed at ELF research.

### **1.1 ENGLISH AS A LINGUA FRANCA: RELEVANCE AND DEVELOPMENT**

Nobody can deny the status of English as a global language in today's world. The reason why it is English and not some other widely spoken language in the world, does not lie in the number of English as a native language (ENL) speakers, the grammatical simplicity of English or its literary power, as has sometimes been assumed. Rather, it is the political, military and economic power of the speakers of the language that has made it a truly global language. (Crystal 2003: 3-8). Additionally, and perhaps also consequently, English is the most widely taught foreign language in the world (Crystal 2003: 5), and has become a 'basic skill' in globalised economies, next to literacy in the mother tongue and national language (in countries where they are different), numeracy and knowing how to use information technology (Graddol 2006: 72).

To ease communication in large international organisations such as UNESCO or the United Nations, one lingua franca very often prevails, and it tends to be English, even if, for

example, French and Spanish have also been set as official languages (Crystal 2003: 12). The same has been true for the European Union, however, it will be interesting to see whether the role of English also changes within the organisation once the United Kingdom has completed its withdrawal from the European Union. Whereas the fact that native speakers (NSs) have played an important part in the globalisation of English is inarguable, Haberland (2013: 195) draws our attention to the significance of non-native speakers (NNSs) in this process, by pointing out that “[O]nly the frequent choice of English for communication between non-first-language speakers of English makes English relevant globally”.

The pioneers in seeing the implications of what would in the future be called English as a lingua franca and in such, a separate research field, indicated their views in the 1980s. Brumfit (1982) in his treatment of English for international communication, argues for a need to establish an alternative normative model, which would not automatically be based on NSs. Furthermore, Smith (1983) called for the recognition of NNS-NNS communication in English and their significance to English as an international language (EIL) and pointed out the possible need for NSs to undergo training in order to be successful in NNS-NS situations where English is used. Dewey (2007b: 23) points out that Brumfit’s and Smith’s views remained rather unknown and obscure until fairly recently when ELF research started to gather more prominence. In addition to them, while Kachru (1985) is mostly known for his three concentric circles of English, the inner circle (English use norm-providing), outer circle (norm-developing) and expanding circle (norm-dependent), he also argued that native speakers had lost their prerogative to control the standardisation of English (for more discussion on this, see Dewey 2007b: 15-52).

Around the beginning of the new millennium, research into ELF started gaining momentum with seminal works by Jenkins (2000) and Seidlhofer (2001) published and creating debate in linguistics and ELT circles. Since then, ELF has established itself as a

distinguished area of research, one testimonial to this being the *Journal of English as a Lingua Franca* (edited by Barbara Seidlhofer) published by De Gruyter Mouton since 2012. The reason why ELF is so different from ENL as a concept, could be explained through the terms ‘language for identification’ and ‘language for communication’, with ELF falling into the latter category and ENL into the former (House 2003). Expressing the same idea, Widdowson (2012) explains the importance of researching ELF through its possibility of contributing to the reappraisal of how we understand what a language, in particular English, is. These ideas are greatly in line with reconsidering the ownership of English and the role of native speakers as models when learning and teaching English.

As for terminology, Jenkins emphasises the advantages of using English as a lingua franca over other terms such as English as an international language (EIL), International English or Global English, stressing the neutrality of ELF and the idea of community behind it (2007: 3-4). The formerly more widespread term, EIL is “regarded by ELF researchers as synonymous” with ELF (Jenkins 2012: 486). There have also been various definitions for ELF, but for the purpose of this study, the definition for English as a lingua franca proposed by Seidlhofer (2011: 7) is used: “ELF as any use of English among speakers of different first languages and for whom English is the communicative medium of choice, and often the only option”. This means that an ELF conversation can also include ENL speakers, who might have to accommodate themselves to the ELF situation and have indeed admitted to doing so, although not always happily (see Albl-Mikasa 2009). The above discussion can be summarised in the words of Jenkins:

To summarize then, ELF is the preferred term for a relatively new manifestation of English which is very different in concept from both English as a Second Language (ESL) – the label frequently given to outer circle Englishes, and English as a Foreign Language (EFL) – the traditional, if to a great extent anachronistic label for English in the expanding circle. (2007: 4)

## 1.2. SPOKEN CORPUS RESEARCH AND ELF CORPORA

Corpus linguistics “provides a means for the empirical analysis of language and in doing so adds to its definition and description” (O’Keeffe & McCarthy 2010: 7). A corpus is essentially a collection of texts, be it written or in the form of transcribed spoken language. The types of corpora are manifold and they vary in size, in purpose, and in the way the data has been collected and analysed. As one possible way to clarify corpora, they can first be divided into written and spoken corpora, with the former far outnumbering the latter. A newer additional category is audio-visual corpora, though these might also be considered a sub-category of the spoken corpora. Next, corpora can be categorised according to their application purposes, which, in addition to corpus linguistics as an end in itself, include language teaching and learning; discourse analysis; literary studies and translation studies; forensic linguistics; pragmatics; sociolinguistics, media discourse and political discourse (O’Keeffe & McCarthy 2010).

Tognini Bonelli (2010) proposes the following typology for categorising corpora: sample corpora; corpora for comparison (based on geographical, historical, topic and contrastive considerations); special corpora; corpora along the time dimension; bilingual and multilingual corpora; corpora concerned with normativeness; non-native speaker corpora; and spoken corpora. As can be concluded from this typology, all the other corpora except for spoken corpora in this categorisation dealt with written text. The decision of what kind of corpus to compile and whether there is a need to compile one in the first place should be based on the application purpose and on research on the already existing corpora. If the aim of the research is to focus on the features of spoken language in a specific context, this naturally means that a spoken corpus should be compiled. However, as a spoken corpus is much more complicated and time-consuming to build, the decision to do so should be well thought through. Nevertheless, once the decision has been taken, the results are likely to be

rewarding, as these corpora are then a source of valuable data for researchers and can “form the basis of new and emerging descriptions of naturally occurring discourse” (Adolphs & Knight 2010: 38).

Since the compilation of the first electronic corpus of spoken language at the University of Edinburgh in 1963-1965, spoken corpus research has come a long way and advances in technology have made it possible for researchers everywhere to compile and analyse their own corpora in addition to those made available on the Internet by others (Tognini Bonelli 2010). The majority of ELF corpora are also concerned with naturally occurring speech, while research and interest in written ELF is gradually growing as well (Cogo & Dewey 2012). There are three large ELF corpora currently freely available for researchers to use: the Vienna-Oxford International Corpus of English (VOICE), the Corpus of English as a Lingua Franca in Academic Settings (ELFA) and the Asian Corpus of English (ACE). Properties that distinguish ELF corpora from learner corpora can at times be difficult to justify, especially if the speakers in the corpora are in fact learners of English as well. However, for the purpose of this thesis, Mauranen’s (2011) approach is preferred, where a learner corpus is directly connected to the classroom setting and the ELF corpus features data from outside the classroom – from situations where ELF is chosen out of necessity or because it is the most comfortable for all interlocutors.

VOICE was the first of large ELF corpora, being launched in 2001 at the University of Vienna and made available for use online in 2009 (accessible at [www.univie.ac.at/voice](http://www.univie.ac.at/voice)) with Barbara Seidlhofer acting as project director. The corpus consists of 1 million words of transcribed spoken ELF collected from various domains (educational, leisure, professional) and comprising speech event types such as interviews, press conferences, working group discussions, meetings, conversations etc. The corpus includes approximately 50 first

languages, the majority of them European, including Estonian, represented by 3 individuals.

Seidlhofer (2004: 220) has identified the following features of spoken ELF based on VOICE:

- Dropping the third person present tense –s
- Confusing the relative pronouns *who* and *which*
- Omitting definite and indefinite articles where they are obligatory in ENL, and inserting them where they do not occur in ENL
- Failing to use correct forms in tag questions (e.g., *isn't it?* or *no?* instead of *shouldn't they?*)
- Inserting redundant prepositions, as in *We have to study about...*
- Overusing certain verbs of high semantic generality, such as *do, have, make, put, take*
- Replacing infinitive-constructions with *that*-clauses, as in *I want that*
- Overdoing explicitness (e.g. *black color* rather than just *black*)

VOICE has already served as a research tool for many scholars (see, for example, Breiteneder 2009, Hülmbauer 2009, Pitzl 2012, Seidlhofer 2004, Seidlhofer 2011) and will undoubtedly continue to serve as such.

The ELFA project, with the ELFA corpus (accessible at <http://www.helsinki.fi/englanti/elfa/elfacorus>) being part of it, started in 2003 and the corpus was completed in 2008 with Anna Mauranen acting as project leader. The corpus consists of 1 million words of transcribed spoken academic ELF, collected from four Finnish universities and disciplinary domains of social sciences, technology, humanities, natural sciences, medicine, behavioural sciences, economics and administration. The speech events have been divided into monologic events (lectures and presentations) and dialogic/polylogic events (seminars, thesis defences, conference discussions). The corpus features speakers of 51 different first languages, including Estonian (0.3% of all tokens). The ELFA project now also includes the Corpus of Written English as a Lingua Franca in Academic Settings (WrELFA), consisting of 1.5 million words of academic text collected from unedited research papers, PhD examiner reports and research blogs. The requirement for all the texts was that they should not have undergone any proofreading by an ENL speaker.

The latest addition to large ELF corpora is the ACE project which was started in 2009 and completed in 2014 (accessible at [corpus.ied.edu.hk/ace/](http://corpus.ied.edu.hk/ace/)) with Andy Kirkpatrick

acting as project leader. This corpus also consists of 1 million words of naturally occurring spoken ELF, compiled by teams across East and Southeast Asia in educational, leisure and professional settings. The speech events included in the corpus are also along the lines of those in VOICE, and it has been indicated that ACE could successfully be compared to VOICE to consider any similarities and differences between ELF in Asian and in mainly European settings (Kirkpatrick 2012: 132). In addition to comparing ELF in Asia and Europe, the ACE project lists other objectives such as exploring the role of L1 in negotiating meaning or comparing constructions, lexical items and sound patterns that are considered ungrammatical in standard L1 English to how these function in Asian ELF. Some of the results of the research that has evolved from ACE can be found in, for example, Kirkpatrick & Subhan (2014) and Walkinshaw & Kirkpatrick (2014).

Of the smaller corpora compiled by individual scholars for their own, more specific, research purposes, two are of particular interest in the context of this thesis – those compiled by Dewey (2007b) and Önen (2014). Dewey's research stems from his own experience as a language teacher and teacher trainer and the shift in his views on 'learner language' and the recurrent 'errors' to considering these as just examples of L2 language use (Dewey 2007b: 9-10). To explore these features in more detail in his PhD thesis, Dewey compiled a corpus of 61,234 words of spoken ELF in natural or 'quasi-natural' settings at International House, London and King's College London, with a few speech events also recorded elsewhere in the same city and also in China. The corpus consists of 42 different communicative events and includes 17 first languages (Estonian is not among them) (Dewey 2007b). Most of the speech events were discussions on given topics, without the presence of the researcher, and some were entirely natural conversations, the participants included teachers and students, with many of the latter at university level, but the contexts were rather informal (Dewey 2007b: 62-69). Dewey's (2007b) focus was on innovation in lexis and grammar and his

results, particularly the part pertaining to articles, prepositions and collocations with verbs of high semantic generality, will be discussed later in the second chapter of the thesis.

In a similar manner, Önen's 2014 PhD thesis focuses on the lexicogrammatical features of ELF and she also compiled a corpus, which she called Corpus IST-Erasmus. The name of the corpus refers to the participants, who were mainly ERASMUS students and to the place of the recordings – Istanbul. The number of L1s represented in the corpus is 24 (Estonian not being among them), and the speech events were interviews and focus group meetings, altogether 54 speech events, amounting to 93,913 words. The present thesis used a similar method of interviews with the interview questions also being formulated on the basis of Önen's (2014) questions. Additionally, the participant information questionnaire used for this study was modelled on Önen's (2014). Her results concerning prepositions, articles and collocation use will be given detailed attention in the second chapter of this thesis. Thus, this thesis is in a way a continuation of Dewey's (2007b) and Önen's (2014) research to see if some of their results can be reconfirmed in the Estonian setting or not.

### 1.3 ELF AND PEDAGOGY

Dewey's (2007b) and Önen's (2014) smaller corpora, described above, both had, in addition to the aim of exploring the lexicogrammatical features of ELF, the hope that these results can have pedagogical implications. ELF scholars are not advocating the teaching of ELF as a variety, but rather bringing the notion of ELF as a communication tool to the minds of teacher trainers, teachers, and finally, learners (Jenkins 2012, Kohn 2015, Seidlhofer 2011). This attempt is often understood by teachers as a suggestion to teach 'incorrect English' or favour deviations (Kohn 2015). Jenkins (2012), however, emphasises that ELF researchers have never wished to dictate to teachers how or what they should teach, but encouraged them to decide based on the given context and taking into consideration learner

choice. “All they ask is that learners are presented with the sociolinguistic facts of the spread of English around the world before they make their choice.” (Jenkins 2012: 492)

What ELF researchers are arguing *against* is the entrenched view in ELT that the only true English to be seen as a model by an L2 learner can be some ENL variety, which usually tends to be standard British or North American (Dewey 2007a: 346, Seidlhofer 2011: 41, 45, Jenkins 2012: 491). The suggested alternative model, more relevant to most L2 users of English, could be the competent non-native speaker (Ur 2010, Seidlhofer 2011, Dewey 2014, Hynninen 2014). Dewey (2012: 165) highlights the need for more empirical research on ELF communication which would help raise awareness among teachers and help them “move beyond normativity” (Dewey 2012: 166).

Kohn recognises the complexities that teachers face when considering the possibility of adopting an “ELF-informed ELT perspective” (2015: 55). Since Standard English (SE) is viewed as having high status in the socioeducational tradition, the strong orientation towards it when teaching is not easy to renounce. However, Kohn (2015) argues that this might be achievable when ELF and ELT were reconciled and he proposes a social constructivist model of “my English” to facilitate the reconciliation. This concept would ideally move the focus from a strong SE to a weak SE orientation (Kohn 2015: 62) in teaching and learning (therefore not abandoning the idea of SE altogether), because the strong one is seen by Kohn as a behaviourist cloning model which language learning cannot be (2015: 59). The weak SE orientation, on the other hand, would allow learners to use their own English for ELF communication purposes (Kohn 2015: 62).

The emphasis on native speaker-like proficiency as the ultimate goal of language learning, though in most cases a clearly unattainable one, was evident also in the original 2001 Common European Framework of Reference for Languages (CEFR) that is meant to cater for all languages and which has been seen as the yardstick to materials and test

developers, teachers and learners alike (Council of Europe 2001). In the context of a global language that English is, and bearing in mind the predominance of NNS-NNS communicative acts, this approach is problematic (Seidlhofer 2011: 185, Hynninen 2014). Fortunately, the CEFR is currently in the process of being reviewed and the pilot version of the extended CEFR Illustrative Descriptors allows us to hope for a new version that has shifted the focus to a ‘proficient speaker of the target language’ (Council of Europe 2016). Since in an authentic situation the aim is successful communication, it has been found that NNSs might in effect be more successful in negotiating meaning in a lingua franca context (Hülmbauer 2009). Thus, the fact that the Council of Europe has acknowledged the problematic aspect of using ‘native speaker’ in their previous descriptors, can be considered as one testament to the validity of the claim that NNSs can be as successful in communication as NSs and maybe even surpass NSs in certain contexts. Whether the revised version could then also fit the ELF paradigm, or would there still be a need to develop one specifically for ELF purposes, is yet to be seen.

Should the revised CEFR suit better the purposes of ELF, it would presumably significantly advance teachers’ and teacher educators’ understanding of ELF, give ELF more credibility and help bring it into the classroom. It has been claimed that in order for ELF to make its way into the classroom, the teachers need to start with developing a different mindset (Seidlhofer 2011). One that involves critical contemplation, rather than following some set rules that could be developed based on ELF research and would be imposed from ‘above’ – a “magic formula for teachers. [ELF research] rather should be a springboard for stimulating thought and generating discussion.” (Illés 2016: 142) Only then can ELF-awareness be also communicated to the learners (Sifakis 2014).

Since it is often exceedingly difficult to change how established practises have been thought of and put to use, Sifakis emphasises the importance of generating “a plan for ESOL

teacher education that both integrates such [ELF] research and responds to the rapidly changing needs and realizations of learners” (2014: 323). To this end, Sifakis (2014) puts forward the transformative framework to educate the ELF-aware teacher, as an alternative to the critical perspective. This framework would have two phases, firstly, reading selected ELF research and literature pertaining to critical pedagogy and post-modern applied linguistics; and secondly, action research of their own in their appropriate context. However, since autonomous training can prove challenging, teachers would need mentoring in the framework, to move from the more general to the more specific, and to what interests them. (Sifakis 2014: 328-330) Keeping in line with these ideas, my research could serve as the more specific reference material to ESOL teachers in Estonia.

#### 1.4 CRITICISM OF ELF

Phillipson (2008) argues that the dominance of English is directly linked to the UK’s and the USA’s policies during colonisation and post-colonially, this includes not only military or economic, but also cultural, religion, and language policies. He suggests that there is not one all-encompassing *lingua franca* in English, but that there are *lingua economica*, *lingua emotiva* (related to pop culture, consumerism and hedonism), *lingua academica*, *lingua cultura* (integrating literary text with language learning), *lingua bellica* (of wars) and *lingua americana* (Phillipson 2008: 250). He also implies that English might in fact be a *lingua frankensteinia* – having created the monster of English, or a *lingua cucula* – acting as a cuckoo who substitutes its eggs in place of other languages, especially in Europe and in the EU. Furthermore, he claims that ‘the variety and complexity of sociolinguistic realities’ prevent us from accurately establishing any standardisation of ‘world’ Englishes (Phillipson 2008: 261). This, in effect, would also mean that it is virtually impossible to show any tendencies in ELF usage or establish a Lingua Franca Core for pronunciation (see,

for example, Jenkins 2007: 22-28). According to Phillipson, “Global English is not a reality”, but a normative project like globalisation and the EU (2008: 260). Finally, although he credits Seidlhofer, Jenkins and other ELF advocates for trying to endorse equality in communication, Phillipson fears that they might “be doing themselves a disservice by using ELF for what they see as a distinct new variant of English, since the term *lingua franca* has so much cultural baggage and is open to so many interpretations” (Phillipson 2008: 262).

Sowden (2012) draws our attention to the complications of incorporating ELF to ELT, and claims that even if it would be possible to devise a lingua franca core and present it as teachable material, this would create a situation of choice for learners and teachers. He believes that the more ambitious would choose to learn with teachers of near-native competence and “the poorer sections of the community would be relegated to schools where ELF was the norm” (Sowden 2012: 93-94). In Sewell’s (2013) views, the main problem with ELF lies in the polarising of NS and NNS varieties of language, or ELF and non-ELF. He also argues that it is false to assume that only dominant groups produce ideologies that affect our attitudes towards language and that ELF also contains these (Sewell 2013: 7). Also, he asserts that we can only talk of an ELF community if the speakers themselves identify with ELF (Sewell 2013: 6). However, Sewell (2013: 8), in the same way as Phillipson (2008), acknowledges ELF’s role in becoming aware of the need to see concepts such as ‘community’ and ‘language’ as problematic.

O’Regan in his immanent critique has pointed out that “the ELF movement is ideologically conservative, is inconsistent in its arguments and is lacking in theorization” (2014: 534). He (O’Regan 2014) finds contradiction in how the ELF movement claims to have moved beyond the original focus on features and a possible codification while at the same time not abandoning hope for codification altogether (see Jenkins et al. 2011). Using terms such as ‘spoken ELF’, ‘written ELF’ or ‘ELF interaction’ in effect, he sees as the

hypostatisation of ELF where ELF as an abstraction “is made to *appear* already given, or ‘real’” (O’Regan 2014: 537, emphasis in original). O’Regan (2014) also describes what he calls ELF fetishism, in which ELF as a *thing-in-itself*, a ‘false consciousness’, is used to legitimize the project as a whole and he blames ELF for its weak philosophy in general.

To summarise, it could be said that ELF criticism concentrates a lot on the terminology and conceptual aspect of ELF, while acknowledging that ELF scholars deserve respect for bringing the question of the ownership of a language and what language is at all to the forefront in wider discussion. Naturally, ELF scholars are aware of the criticism and address it to some length in their writings (see, for example, Seidlhofer 2011 and Dewey 2013). In conclusion, this chapter aimed at establishing an understanding of the different conceptual aspects necessary to consider when dealing with ELF research, especially in the context of ELT, whilst also providing justification to the use of spoken corpora in exploring ELF communication. The next chapter will focus on the mini-corpus compiled for the present thesis and provide an analysis of some of its lexicogrammatical features.

## **2. THE COMPILATION AND LEXICOGRAMMATICAL ANALYSIS OF THE SPOKEN ELF MINI-CORPUS**

The purpose of this chapter is to describe the process of compiling the Estonian spoken ELF mini-corpus and present the results of the analysis of lexicogrammatical features in the corpus and compare them to previous research on lexicogrammatical features in ELF corpora. More specifically, the subject of study is the use of articles, prepositions and collocations with words of high semantic generality. The analysis and discussion reveal similar tendencies to those found in earlier studies and offer explanations as to the motives behind these innovations. Along with that, the limitations and pedagogical implications of the study are discussed with ideas for further research.

### **2.1 DATA COLLECTION**

The mini-corpus for this study is made up of 9 semi-structured interviews and comprises approximately 85 minutes of speech. The construction of the spoken corpus followed the basic procedure described by Adolphs & Knight (2010: 40): recording, transcribing and annotation, and analysis. To compile the corpus, 12 semi-structured interviews were recorded in November 2016. However, only 9 of the recorded interviews were later used for this study, the reasons for which are explained below. Before recording the interviews, a pilot interview was carried out to test the comprehensibility of the interview questions, the quality of the recording device and to determine the estimated length of an interview. Following the pilot, a few minor alterations were made to the wording of the interview questions, to make them easier to read and understand, but the content of the questions remained the same.

The interviewer for the Estonian native speakers was a Polish native speaker, thus making the interviews genuine ELF situations. At the same time, the interviews with the L1

Russian speakers and one bilingual speaker were conducted by an Estonian native speaker who shares a cultural and partially also a linguistic background with the interviewees (both L1 Russian speakers listed Estonian as a language they speak on the participant information questionnaire). It is therefore debatable whether the latter is ELF communication or not, since English would not have been the only possible communicative medium – the interviews could have been carried out in Estonian as well. This is the reason for excluding these three interviews from the corpus for the current study. Hence, only 9 of the original 12 recorded interviews were used for the study.

The interviews were semi-structured, meaning that although there were 13 questions that were expected to be asked by the interviewer, the interviewer was also instructed and encouraged to ask follow-up questions, add personal comments and skip questions in case they had already been addressed by the interviewee, all to make the interview resemble as much as possible natural conversation. The questions provided included three warm-up questions and 10 main questions. The general topic was studying at university and the English language. The questions were partly derived from Önen's (2014: 65) questions used for her study on the lexicogrammatical features of English as a lingua franca. Compared to Önen's questions, there were fewer in this study and they were slightly more general. Also, since her questions were specific to the context of international students in Turkey, they had to be modified to fit the Estonian context with Estonian interviewees. See Appendix 1 for the interview questions used for this study. Since Dewey's (2007b) data collection was more varied and included different speech events, the build-up of my data collection process did not draw much on Dewey's.

In the interview room, the only people present were the L1 Polish interviewer and the L1 Estonian interviewee. Although the interviews were recorded on campus, the students were all studying English and received course credit for participation, the corpus could not

be called a learner corpus, despite the initial plan to do so (as can also be seen from the participation information sheet and consent form), since the data was not recorded in a classroom environment. Whereas the students are language learners while they are in a classroom environment, they were users for the length of the interviews. English was used as an ‘instrument language’ rather than a ‘target language’ as it would be in a classroom (Mauranen 2011: 159). This principle was followed and explained by Dewey, who also deems his participants to be rightful users of the language, “regardless of whether they are or are not actively involved in English language study” (2007b: 56).

## 2.2. PARTICIPANTS

The participants interviewed were all students at the University of Tartu with English as their minor, some of them were also training to be teachers (of whom some stated a hope to become teachers of English). Participation in the interviews was voluntary, but upon participation, the students received extra course credit and gained valuable insight into how academic research is carried out. Some of them even expressed gratitude for the opportunity to be part of the research after the interview. Before recording the interviews, all participants were asked to read through the participation information sheet and to fill in the participant information questionnaire and consent form (see Appendices 2 and 3). The participants were given a copy of the information sheet and consent form to keep.

The participants were not asked to provide any kind of proof of their level of English, but the levels appear to range from B1 to C2 based on the current CEFR, with the majority of the students being at B2-C1 level. These judgements are based on my own opinion, as I am quite used to evaluating students based on the CEFR. Hence the interviewees qualify as independent/proficient, if not competent users. Neither Dewey (2007b) nor Önen (2014) specified the exact language level of their participants, however the general profile of their

participants is similar to those in this study. All of Önen's (2014) participants were also university students, while Dewey's (2007b) were a mix of students, trainee teachers, teachers and teacher trainers.

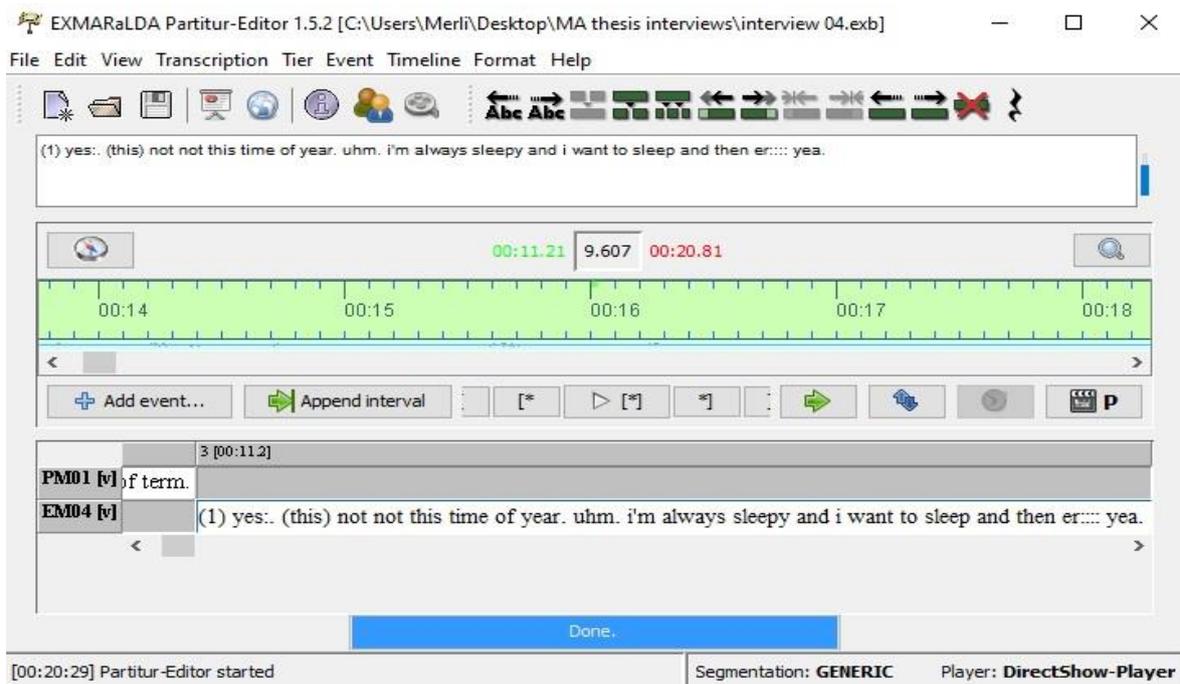
Nine of the 12 original interviewees listed Estonian as their mother tongue, two stated their mother tongue to be Russian and one student claimed to be bilingual and regarded both Estonian and Russian as their mother tongues. As explained above, during the preliminary analysis of the collected data, the decision was made to include only students who had stated Estonian to be their mother tongue in this study. Therefore, the interviews with the two Russian native speakers and the bilingual speaker were excluded for the time being from the corpus. This decision was taken based on the nature of the interviews and on the conceptual decision to consider this corpus an ELF rather than a learner corpus (see Mauranen 2011).

Besides Estonian, English and Russian, other languages that the participants noted on the participant information questionnaire as languages they have learnt or speak included Finnish, French, German, Korean, Latin, Norwegian, Old-Greek, and Swedish. Among the speakers included in this study, there was one male student and eight female students aged 19-27 years (average age 22.4 years). In comparison, Önen's participants were also aged 19-27 years. Dewey (2007b), by contrast, has only specified the nationality and first language(s) of his participants, since he did not consider the age, sex, socio-economic or educational background to be of relevance to his study, as his aim was to obtain as diverse a sample as possible (2007b: 66-67). The average age when the speakers in this corpus had started studying English was 8, with the earliest age being approximately 4 and the latest 11 years old. However, the average age of first exposure to English was 5.9 years. The average length of the interviews was 11.98 minutes, with the shortest lasting 8.10 minutes and the longest 17.30 minutes.

## 2.3 TRANSCRIPTION AND ANNOTATION

Following the recording sessions, the interviews were transcribed using the free software EXMARaLDA and the VOICE mark-up and spelling conventions for transcription (<https://www.univie.ac.at/voice/>). Considering the scope and focus of this study, some of the VOICE mark-up conventions, namely those indicating pronunciation variations, breathing, speaking modes and speaker noises were not made use of. Furthermore, as these interviews were not filmed, the conventions regarding non-verbal feedback and contextual events were also not relevant within the scope of this study. For a list of the conventions used for this study, see Appendix 4. The EXMARaLDA software allows the transcriptions to be saved in various formats which makes the data also compatible with other software. A detailed transcription of the data makes it possible to use the corpus for further research and using the VOICE conventions that other ELF studies (among them Dewey 2007b, Önen 2014) have also used, helps this study to fit into the larger paradigm of ELF research.

The interviewees were kept anonymous in the transcriptions by giving them speaker ID-s consisting of two letters – one to denote their L1 and one to denote their gender, followed by the number of the interview, resulting in, for example, the interviewer's speaker ID to be PM01 (Polish and male). This makes for a continuous system which is easy to use with new data when wishing to expand the corpus. Figure 1 shows how the transcribing process looked like in the EXMARaLDA Partitur-Editor tool. After transcription in the EXMARaLDA Partitur Editor tool, the output layout used for the interviews was that of a musical score, with the “speech arranged according to the time at which it occurred” (Adolphs & Knight 2010: 46). For an example of a fully transcribed interview, see Appendix 5.



**Figure 1. EXMARaLDA Partitur-Editor tool.**

After the interviews had been transcribed, they were annotated. To analyse article use in the corpus, nouns, noun phrases, quantifiers, ordinal numbers and superlative adjectives, for which article use was deemed innovative, were extracted as referents and the following categories were used for annotation: word class/referent type (noun, noun phrase, ordinal, quantifier, superlative, of-phrase); number (singular, plural, not applicable); countability (countable, uncountable, not applicable) and reference type (specific, generic, not applicable). In devising these categories, the example of Dewey (2007b) was broadly followed with reference to Quirk et al. (1985) and also Biber et al. (1999). Each article use was then also commented on from the perspective of typical SE article use, noting whether it was an omission or addition of an article in each case.

A similar approach was used with prepositions. All cases of innovative preposition use were extracted from the data and then commented on from the perspective of the innovation, whether it was substitution of another preposition typically used in ENL, omission, a redundant preposition, inconsistent use or self-correction. In addition to referring

to Quirk et al. (1985), Dewey (2007b) and Önen (2014), *the Oxford Collocations Dictionary for students of English* (OCD) was used for reference to check the ENL versions, as this dictionary is also corpus-based. The speaker and interview number were also indicated in the annotations with both article and preposition use.

As to the collocations with verbs of high semantic generality, all transcribed interviews were checked for the verbs *take*, *get* and *make* and their various forms and these were then extracted with enough data preceding and following the verb itself to provide context which could help determine the novelty of each particular collocation. Again, the OCD was used for reference in addition to the above-mentioned sources. In each individual interview transcript, all the verbs were marked, and so were the innovative uses of articles and prepositions so that when reading a transcript, the instances would all be instantly noticed.

## 2.4 RESULTS

The results pertaining to the data analysis of the use of articles, prepositions and collocations with verbs of high semantic generality are presented below. The data revealed the omission of definite articles in certain contexts and, on the other hand, substituting the zero article with the definite article in others. There were also cases of prepositions omitted, substituted or used redundantly. Finally, the verb *take* was used in novel collocations that seemed to be semantically motivated.

### 2.4.1 ARTICLES

All but one speaker exhibited use of articles in novel ways in the mini-corpus. There were 55 instances of article use that could be deemed either innovative or the innovativeness of which was exceedingly difficult to judge. The latter was mainly the case with ordinal

numbers and there were 18 such cases where it was either problematic to judge whether the definite article should have been employed or not or it had been omitted and substituted by the zero article. In SE, it is a general rule to use the definite article before ordinals (Quirk et al. 1985: 270, Biber et.al 1999: 90), however when the reference seems to be rather generic than specific, it seems reasonable to omit the article, for example when talking about *first grade* and referring to the first year of school in all schools not the first grade of a specific school or a person. The definite article has been indicated as optional in phrases such as (*the*) *second grade* in the OCD as well (2009: 367). However, even when the reference is more specific and the definite article most probably would be used in ENL, it is most likely substituted by the zero article, although there were also instances of the same individual speaker varying between the definite and the zero article for the same referent, which is in line with Dewey's findings (2007b: 159-161). All in all, there is a strong tendency to use the zero article with ordinal numbers in the Estonian ELF mini-corpus. The randomly chosen examples from the data are the following:

**Example 1: interview01, line 53**

EF01 [v] uhm. well. i started in fourth grade? i was: (1) my math is <@>very

**Example 2: interview04, line 61**

EM04 [v] i was in third grade and i ithink (.) but they're sometime uh basically they're

**Example 3: interview07, line 44**

EF07 [v] when i was:: in like sixth grade? we er:: in maths we had a really er really difficult

Other cases of using the zero article to substitute the definite article included inherently unique referents such as *the truth*, *the Internet* or *the USA* and fixed phrases such as *in the morning* or in the phrase *the west side of Estonia* where we have both the cardinal point and the of-phrase which in ENL would be used with the definite article. These findings

are in line with Önen's (2014: 93), who also argued that ELF speakers prefer to omit the definite article *the* when compared to standard ENL use. Furthermore, as in Önen (2014: 92), the indefinite article *a* was also substituted by the zero article in a number of cases, for example with the quantifiers *a lot of* and *a bit*:

**Example 4: interview04, line 44**

EM04 [v] =it's u:.....h (2) m maybe uh i'm. bit ros-rusty or something and i haven't (1) |

**Example 5: interview10, line 42**

EF10 [v] know</9> lot of people also start in the first grade? (. ) s but i did (on the |

Similar examples were reported by Dewey (2007b: 111-113) with possible explanations for this innovation that will be discussed further below.

Looking at the use of the definite article, the results in my data are completely in line with Dewey's (2007b: 116) who noted that "the strongest trend though is the use of *the* for generic reference, in conjunction with both singular as well as plural referents". Due to the topic of the interview questions, the recurrent tokens which were used in the generic sense but were preceded by *the* were connected to education, namely *middle school*, *university* or *basic school*. Önen (2014: 94) also reports the "tendency to overuse *the* in  $\emptyset$  contexts".

The indefinite article *a* featured the least in novel positions (4 tokens, for example in *take a subway*) in the data and *an* did not feature at all. Two of the cases were with generic and two with specific referents, two countable and two uncountable, so no generalisations can be made on the basis of these criteria. Önen's (2014) data also allows us to suggest that the indefinite article is used much less for substitution than the definite article.

### 2.4.2 PREPOSITIONS

Dewey (2007b: 97) states that his data provides plenty of evidence “to suggest that prepositions constitute an area where language change will be very productive in ELF settings”. There were 47 cases of innovative preposition use in the mini-corpus and each speaker exhibited at least two (though in most cases more) innovative uses. Out of these 47 cases, 14 were clear instances of omission, the most omitted preposition being *of*, for example in:

**Example 6: interview07, lines 154-155**

EF07 [v]	(2) um:: (.) i've studied a bit russian and a
EF07 [v]	bit french in school= =so i:: er i u-understand a bit of russian but i i can't like

However, this is an interesting example, since the speaker later in the same utterance uses the same construction with the preposition *of*, which signifies inconsistent use and could mean that the construction is in a ‘fluid’ state or undergoing change in ELF usage, as there were other similar examples by different speakers too:

**Example 7: interview12, lines 115-116**

EF12 [v]	i speak swedish as well? a::nd a little bit
EF12 [v]	norwegian (.) i've also studied russian for like (2) six or seven years? in middle and

Other examples of preposition omission included omission in utterance final position:

**Example 8: interview06, line 79**

EF06 [v]	because (.) there are a lot of places i would want to go (2) and (.) the
----------	--

or omission of *with* with the verb *help*:

**Example 9: interview10, line 17**

EF10 [v]	really slow but (.) i think my estonian teacher =really helped me that. i didn't
----------	--

The example of omitting the preposition *to* where it should be present in ENL with the verb *listen* that has been discussed and illustrated with examples by both Dewey (2007b: 96-98) and Önen (2014: 102) appears once in my data:

**Example 10: interview02, lines 60-61**

EF02 [v]	then we are constantly in the en-english e-environment.	and it er helps us to
EF02 [v]	listen the teacher. or (2) we a:::re (1) e:::r w studying during	listening to the teacher.

What is striking in this example is that, once again, the same verb appears with *to* in the same utterance later, but it does not appear to be a case of conscious self-repair, rather another example of inconsistent use. This could constitute as evidence of Dewey's claim that "this item is unstable either in this speaker's language use or perhaps more probably, that it is unstable more generally and therefore currently in a period of transition" (2007b: 97-98).

Moving on to redundant prepositions, all in all there were 8 such cases in the mini-corpus. There are two instances of using *of* after *because* where the preposition is actually not used in ENL:

**Example 11: interview07, line 51**

EF07 [v]	rewarding because of you can see the results	=even when you or w mostly you
----------	--	--------------------------------

**Example 12: interview10, lines 5-6**

EF10 [v]	(1) i think it's going to be be busy soon because of the exams
EF10 [v]	are coming <2>soon</2> but (.) right now i'm kind of free.

The explanation to these examples probably lies in the very widespread use of *because of* where *because* is a reason adjunct, whereas in the above examples, the function is that of a conjunction. Similar instances can be noted where *of* is used redundantly after *a lot*:

**Example 13: interview01, lines 62-63**

EF01 [v]	wel:::: i think that i'm not. i'm not (very) sufficient like i think i have a lot
EF01 [v]	of. a lot of to improve? still because i know that i have m: some problems with

**Example 14: interview08, line 51**

EF08 [v] could do a lot of there. for example in our cell biology lab we only looked at

The most frequent type among all the innovative uses of prepositions was substitution for another preposition, which amounted to 21 cases. The most common preposition to be used in place of another was *in* (7 tokens) and the prepositions it substituted were *on* and *at*, for instance:

**Example 15: interview02, lines 120-121**

EF02 [v] norwegian. (.) since i lived there? i studied it a little (.) an:::d i do speak a  
EF02 [v] little. but only:: in a basic level. again. (.)

**Example 16: interview04, line 85**

EM04 [v] (2) in daily basis i'm. not. er (1) need. (.) to. use english

**Example 17: interview10, lines 50-51**

EF10 [v] (1) i think i need english (.) english (1)  
EF10 [v] pretty much? in the internet? = i chat with my friends? (1) from england too in

On the other hand, there was one instance of *at* being used in place of *in*:

**Example 18: interview07, line 37**

EF07 [v] e::r and then i er at eleventh grade i also wanted to have (.) a::: a side option in case

and also one instance of *on* being used instead of *in* (or *during*):

**Example 19: interview07, lines 70-71**

EF07 [v] in the first year maybe i'll like (.) it'll be a bit easier than just like taking it on  
EF07 [v] second year and= =and then having a lot of stuff to do so (.) <12>i'm

Önen (2014: 102-103) also found that the prepositions *in*, *on* and *at* are often used interchangeably in ELF communication.

Another case of substitution which provides an interesting example is the use of the phrase *communicate with* in one speaker's speech. Namely, in lines 95-96, she uses the preposition *to* instead of *with*:

**Example 20: interview02, lines 95-96**

EF02 [v]	(.) so. (.) e::r what was that i DO rely on english. i think it's e::r good enough for
EF02 [v]	communicating to:: different people? i. yea i can make myself

however, a little later, in lines 109-110, she firstly employs the same substitution, but then self-repairs the preposition:

**Example 21: interview02, lines 109-110**

EF02 [v]	everybody knows if you know english you can communicate to (1) a lot of people
EF02 [v]	in the world. communicate WITH (.) but:: (1) er (.) WHY do i think that?

This might again serve as implication of a phrase that is unstable in the speaker's repertoire or in ELF communication in general.

#### 2.4.3 COLLOCATIONS WITH WORDS OF HIGH SEMANTIC GENERALITY

Seidlhofer (2004: 220) identified the verbs *have*, *do*, *take*, *make* and *put* as verbs of high semantic generality, but as Dewey (2007b) and following up on his study, Önen (2014) both looked at *get* rather than *put*, the same is done in this study for ease of comparison. Due to the scope of this study, *have* and *do*, which have a wider variety of functions than *make* and *take*, were not investigated. Cogo and Dewey (2012: 70) argue that verbs of high semantic generality tend to combine into innovative collocations and fixed expressions in ELF language use; this topic is explored in the present corpus through the verbs *take*, *get* and *make*.

In my data, there were 40 instances of the verb *take* used in different forms (*take*, *took*, *taken*, *taking*). Most of the collocations were variations on *take a class*, *take a course*

or *take (some) language* (meaning a language course), which is in line with Önen's findings (2014: 113). However, in one case, this meaning had also been transferred to *summer school* in the following utterance:

**Example 22: interview08, lines 73-74**

EF08 [v]	(.) yes i actually wen::t two years ago i went to oxford?= =i took this summer
EF08 [v]	school there? and i was for there for two weeks where i studied (.) english literature

This could be explained as a semantically motivated innovation (Dewey 2007b: 150), because the meaning of *summer school* is very close to *course* or *class* in that it is usually comprised of these. ENL speakers would most likely combine *summer school* with *take part in* or *participate in*, or even *go to*.

In a similar way, the same speaker uses the phrase *take a laboratory*, which is certainly also a semantically motivated innovation, since it also has a direct connection with *course*, *class* or *subject*, as can be seen from the utterance itself, where the laboratory is part of a biochemistry course:

**Example 23: interview08, line 50**

EF08 [v]	sugars and (1) and the:: (.) the laboratory we took in biochemistry was very we
----------	---

More interesting and somewhat more difficult to explain is the following example (which, however, is probable, since the phrase was not entirely audible in the recording):

**Example 24: interview12, lines 6-7**

EF12 [v]	(.) yeah it is i had a i had the flu in september? so i <1>(took)
EF12 [v]	out</1> of school and i'm left behind in some subjects so i had to do some <2>extra

*to take out of school* – the possible explanation for this phrase could be the mixing of the phrases *to drop out of school* and *to take time off* – which, combined, is more or less the idea that the speaker is trying to convey. In this case, the instance is also semantically explainable – in order to express an idea, two phrases have been blended into one.

There were 37 instances of the verb *get* (in the forms *get*, *gets*, *getting*, *got* and *gotten*) in my data. No particularly innovative collocations emerged from the data, on the contrary, it seems that in my data the use of the word *get* is more similar to its use in ENL, where *get* is regarded as a ‘key word’ that improves spoken fluency and ‘naturalness’ (Dewey 2007b: 121). The speakers use expressions such as *got to give a class* and *didn’t get to come* instead of expressing the same thoughts with the word *could*. In ELT this would be a clear indication of NS-like spoken language usage.

There were 16 instances of using the verb *make* in the corpus (*make*, *makes*, *making* and *made*), however, none of them were innovative uses, they all followed the patterns of ENL use of the verb.

## 2.5 DISCUSSION

One of the first thing to note regarding the innovations reported in the Results section above is that the innovative use of articles, prepositions and verbs of high semantic generality did not result in miscommunication in this mini-corpus. The results are therefore in line with Dewey’s findings (2007b: 132). Thus, it is a case in point that innovations in lexis and grammar do not generally lead to a loss of quality regarding effective interaction (Dewey 2007b: 131). It can also be said that several features found during this study coincided with those identified by Seidlhofer (2004), Dewey (2007b) and Önen (2014). Dewey (2007b: 137-138) explores a variety of underlying motivations that could be the cause of innovations in

lexicogrammar. Some of these will now be explored in relation to the mini-corpus compiled for this study.

There seem to be two emerging patterns in the Estonian ELF speakers' use of articles. The first is exploiting redundancy – the omission of the definite article with words that are inherently unique, and this includes ordinal numbers, cardinal points, abbreviations of country names and so on. This tendency has been observed also by Dewey (2007b) and Önen (2014), which might serve as proof for its generality, regardless of the speakers mother tongue. Of course, the fact that Estonian does not have an article system, might play its role, but this influence should be explored further before making any final claims.

The second emerging pattern is substituting the zero article with the definite article before generic referents that in ENL usually do not require a definite article, such as institutions in general reference. This pattern was also observed by Dewey (2007b: 155) and Önen (2014: 93-94). Another possible motivation behind using the definite article in place of the zero article could be added prominence – the referents are for some reason deemed more important than they would generally be in ENL. The present mini-corpus therefore also confirms the claim that “it is not the case that the indefinite and definite article is used less in ELF, but that the article system is being employed differently” (Cogo & Dewey 2012: 62).

As to the prepositions, the omission, substitution and insertion of (redundant) prepositions were observed in the data. The most frequent of these was substitution, with the prepositions *in*, *on* and *at* being used interchangeably the most – an observation also made by Önen (2014). This could indicate that the three prepositions in ELF are undergoing a shift in semantic terms, where they could in fact be used interchangeably, since they are very often semantically ‘empty’ and entirely dependent on the preceding or following words (cf. Dewey 2007b: 143). Omission was the second most widespread innovation, with the

preposition *of* being omitted the most. The examples included omission in the phrase *a bit of* and while it was also the case that the same speaker later used the preposition in the same phrase, it was not an instance of self-correction. This might imply that these constructions are also in a state of transition or in a ‘fluid’ state. There were not many cases of redundant prepositions, but again, the preposition *of* was the one inserted in phrases *because of* and *a lot of* in contexts where it usually does not appear in ENL. Since *a lot (of)* is similar to the phrase *a bit (of)* discussed above, this could again serve as evidence for the inconsistency of these structurally and semantically related phrases.

The verbs of high semantic generality explored in the mini-corpus for this study were *take*, *get* and *make*. While *take* was used innovatively, there were no instances of the verb *make* being used in novel ways in this corpus and the use of *get* rather revealed similarities with colloquial ENL. The innovations regarding *take* were semantic generalisations based on educational concepts such as *take a class* or *take a course* which had been applied to the phrases *take a summer school* and *take a laboratory*. Dewey (2007b: 150) also argues that “The innovations in the way these verbs collocate in the ELF corpus seem largely to be semantically motivated.” This should, of course, additionally be explored with the verbs *do*, *have* and *put* too. However, the assumption that can be made based on the data analysed is that the innovations in collocations with verbs of high semantic generality tend to be semantic as well.

Whereas the focus of this study was on using articles, prepositions and collocations with (some) verbs of high semantic generality by Estonian ELF speakers, the implications for language teaching can also only be drawn for this area of the discourse. The corpus compiled for this research can, however, be used to investigate other areas such as the rest of the collocations with verbs of high semantic generality, word order and transitivity, which have been addressed by Dewey (2007b), or the use of or use of the 3<sup>rd</sup> person *-s* that have

also been dealt with by Breiteneder (2009) and Önen (2014). Another aspect that could be given attention is the use of tenses and the self-repairs speakers demonstrate, since Dewey (2007: 133) has pointed this out as something that should be given more priority in language teaching, as compared to, for example, prepositions and articles which do not seem to be ‘repaired’ as much. Nevertheless, there were clear instances of self-repair in the use of prepositions in the data of this study. Self-repairs have also been the subject of research for Kaur (2011). To make any conclusions as to the self-repairs of Estonian ELF speakers, more data should certainly be gathered.

If more of the current data were analysed, the possible outcome that could be of relevance for teachers, would be the compiling of a kind of a reference material – not teaching materials as such – but something that would help the (also would-be) teachers understand the patterns that seem to be emerging in (Estonian and other) ELF speakers’ spoken language use. I believe it would be especially useful for teacher students so that they would already at the beginning of their careers become aware of ELF and then be able to make up their minds as to what they want to focus on while teaching, as suggested by Jenkins (2012). This research can then indeed serve as the more specific reference material in the framework proposed by Sifakis (2014) and I am positive that the matter itself will also generate discussion, as Illés (2016) indicated that ELF research should do. In an ideal scenario, from ELF-aware teachers, the awareness of ELF could also reach the classroom of (more higher-level) learners. I believe that when these learners who actually already use English in their daily lives outside the classroom become aware of ELF and identify themselves as ELF speakers, the criticism voiced by Sewell (2013) towards the existence of an ELF community would be easier to refute.

Another potential future study that exploits the already gathered data would be to transcribe the recorded interviews again with the focus on the pronunciation of Estonian ELF

speakers – the VOICE mark-up conventions provide a framework for that as well. The data could also be used to analyse the speakers' accommodation practises to see the extent to which the L1 Polish interviewer influenced the L1 Estonian interviewees' language use, since the influence is evident in several instances. Furthermore, this study only focused on the language use of the Estonian participants and consequently, the L1 Polish interviewer's language use was not analysed at all. This could be done during further research. The corpus could be developed into a more comprehensive ELF corpus if more interviews or discussions were carried out with people of other L1 backgrounds either as interviewers or interviewees. Alternatively, if more interviews were conducted with only L1 Estonian speakers, the current corpus as it is could be expanded and used to draw more definitive conclusions on the use of ELF by Estonian native speakers with perhaps more emphasis on the possible influence of L1 on the Estonian speakers' ELF use.

Continuing on the topic of influence, the choice of the interview as the means of gathering data was certainly a limitation in itself – the questions and topics discussed act as a trigger for certain language – specific vocabulary and grammatical constructions. Thus, in future studies, it would be a good idea to also include group discussions or try to record completely naturally occurring language too. However, for a study of this scale, the interviews proved both time-efficient and not too complex to set up. Also, if the desired outcome of a study should in effect be to obtain language use of a certain kind, the interview can help with guiding the speakers in the wished direction to then hopefully get that data. Another possible follow-up would be to conduct very similar interviews (if possible, with the same interviewer) with NSs and then compare this data to the given mini-corpus. In addition to comparing Estonian ELF speakers' and ENL speakers' language use, we would gain valuable insight into the accommodation practices of both groups to see how valid in this case would be the claims sometimes made in ELF literature that in an ELF situation, the

NNSs do most of the work while NSs do not accommodate so much (Jenkins 2007: 136) or if they do, are not very happy about it (Albl-Mikasa 2009).

## CONCLUSION

Communication using English as a lingua franca is generally seen as communication between people who do not share a common native language and for whom English is the chosen medium for communication (Seidlhofer 2011: 7). ELF is an area of research that has gradually been making its way into the awareness of linguists and English language teaching professionals over the past few decades. However, ELF scholars (see, for example, Jenkins 2007, Seidlhofer 2011, Cogo & Dewey 2012, Sifakis 2014, Kohn 2015) still call for more awareness-raising with the help of empirical research on ELF communication. This thesis hopes to contribute to the field by drawing together the theoretical frameworks of spoken corpus research and pedagogy and using these as a basis for analysing the spoken ELF mini-corpus specifically compiled for the thesis.

The concept of ELF has its beginnings in the 1980s when Brumfit (1982), Smith (1983) and Kachru (1985) all acknowledged the increasing role of non-native speakers in the globalisation of English and in how the English language should be viewed – not as belonging to native speakers anymore, since the number of non-native speakers (NNSs) using English had started to outnumber native speakers (NSs). Their views, however, did not attract much attention until the beginning of the 2000s, when research into English as a lingua franca started to establish itself as a separate field of study. Jenkins (2000) started by focusing on the phonological features of ELF speakers and Seidlhofer (2001) laid the foundation for establishing a description of ELF. Throughout the years, ELF research has depended greatly on the corpora that have been compiled about ELF communication. The majority of these are spoken corpora and the three largest and best-known (VOICE, ELFA, ACE) are described in the first chapter of the thesis. In addition to the major corpora that have already provided researchers with ample data for analysis, ELF researchers have

sometimes opted for compiling their own corpora – which was the case with Dewey (2007b) and Önen (2014), who used the corpora for lexicogrammatical analyses.

In addition to looking at lexicogrammatical features, both Dewey (2007b) and Önen (2014) were concerned with the possible pedagogical implications of their research. Kohn (2015) emphasises the need for reconciliation between ELF and ELT, since the two paradigms have often been seen as being at odds with each other. One of the differences between ELT and ELF is that what in ELT is usually considered erroneous language use, can be seen in ELF as innovative. ELF research in general has a rather strong focus on English language teaching, albeit not on *teaching* ELF but on raising the teachers' awareness of ELF and hoping that with the help of teachers the awareness of learners would also eventually increase. If learners became aware of ELF, they might start to consider themselves as part of the ELF community, as *users* rather than *learners*, which is how ELF scholars prefer to see the learners – only then could we talk of an ELF community – according to Sewell (2013), who directed his criticism at the non-existence of an ELF community. Other critical treatments of ELF (Phillipson 2008, O'Regan 2014) mostly focus on the conceptual aspect and terminology in ELF research and condemn the polarisation of ELF and non-ELF (Sewell 2013). The present thesis acknowledges the criticism directed at ELF and does not wish to advocate ELF as the only alternative in the treatment of English as a global language or in ELT. However, it does agree with ELF scholars on the need to raise awareness of ELF in everyone dealing with English in any way. Also, the thesis agrees with the fact that the awareness can only be raised if analyses of ELF communication demonstrate that the phenomenon of ELF exists.

Therefore, a mini-corpus of spoken ELF communication was compiled for this thesis to investigate the existence of previously identified lexicogrammatical ELF features in the speech of Estonian ELF speakers. The corpus currently consists of 9 semi-structured

interviews featuring an L1 Polish interviewer and L1 Estonian interviewees. The corpus comprises approximately 85 minutes of speech. For the purpose of this study, only the language use of the L1 Estonian speakers was explored since Estonian was missing from the first languages of speakers in the corpora of Dewey (2007b) and Önen (2014), which have served as the primary models and basis for comparison for the present thesis. ELF had previously not been researched in the Estonian context from the lexicogrammatical perspective; the previous studies dealing with ELF in Estonia have focused on the sociolinguistic aspects and attitudes towards ELF (Soler-Carbonell 2014, 2015). With the compilation and analysis of a spoken mini-corpus, the thesis aimed at pioneering the investigation into the linguistic form of ELF in Estonia and comparing the results to those of Seidlhofer (2004), Dewey (2007b) and Önen (2014). The lexicogrammatical features examined were articles, prepositions and collocations with verbs of high semantic generality (*take, get, make*). To this end, the interviews were first transcribed and annotated with the given features in mind.

The analysis of the language use of Estonian ELF speakers revealed some tendencies similar to the lexicogrammatical innovations observed by Dewey (2007b) and Önen (2014) in their respective corpora. There were two emerging patterns regarding article use. The first was exploiting redundancy by omitting the definite article with words exhibiting inherent uniqueness (for example ordinal numbers and cardinal points). The second tendency showed the addition of the definite article before generic referents. Both patterns were also observed by Seidlhofer (2004), Dewey (2007b) and Önen (2014), which allows us to speculate on their generality in ELF communication. Furthermore, the mini-corpus revealed the omission, substitution and insertion of (redundant) prepositions. Substitution was the most frequent innovation and the prepositions *in, on* and *at* were used interchangeably the most. Omissions and insertions were mostly connected to the use of the preposition *of* which might be an

indication of the transitive or ‘fluid’ state of that preposition in ELF language use (in Estonia) – a similar observation was made by Dewey (2007b) in relation to the preposition *to*. Moving on to collocations with verbs of high semantic generality, the only verb that had generated innovative use in the mini-corpus was *take*. These innovations were semantically motivated generalisations, which, again, corresponds to the findings of Dewey (2007b: 150).

To conclude, this thesis serves as a starting point to the research into lexicogrammatical innovations in ELF communication in the Estonian context. It has demonstrated the relevance of ELF research in the treatment of English in linguistics and in English language teaching. The empirical part of the study revealed emerging patterns in the language use of Estonian ELF speakers that coincide with earlier findings by Seidlhofer (2004), Dewey (2007b) and Önen (2014). Whether these are ELF features or *errors* made by learners remains a question of belief and theoretical assumptions of the researcher. In the future, other lexicogrammatical features identified by earlier research can and should also be studied using the existing corpus. Additionally, the corpus can be used for pronunciation and pragmatics research and the language use of the L1 Polish interviewer should also be analysed in future studies. The present thesis can serve as a reference material for teachers and teacher students in Estonia who wish to increase their ELF-awareness. Furthermore, the existing corpus can easily be enlarged following the principles that have guided the compilation process thus far with an addition of other types of speech events besides interviews. To make further generalisations as to the lexicogrammatical features in ELF communication in the Estonian context, more research is undoubtedly needed. However, it is hoped that the present thesis illustrates that an ELF-based analysis can be a plausible alternative to the more traditional ways of considering EFL data.

## REFERENCES

- ACE project. n.d. Available at [corpus.ied.edu.hk/ace/](http://corpus.ied.edu.hk/ace/), accessed April 2017.
- Adolphs, Svenja & Dawn Knight. 2010. Building a spoken corpus: what are the basics? In Anne O’Keeffe & Michael McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*, 38-52. London, New York: Routledge.
- Albl-Mikasa, Michaela. 2009. Who’s afraid of ELF: ‘failed’ natives or non-native speakers struggling to express themselves? In Michaela Albl-Mikasa, Sabine Braun & Sylvia Kalina (eds.). *Dimensionen der Zweitsprachenforschung - Dimensions of Second Language Research. Festschrift für Kurt Kohn*. Tübingen: Narr. 109-129.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Breiteneder, Angelika. 2009. English as a lingua franca in Europe: an empirical perspective. *World Englishes* 28 (2): 256-269.
- Brumfit, C. J. (ed). 1982. *English for International Communication*. Oxford: Pergamon Press.
- Cogo, Alessia & Martin Dewey. 2006. Efficiency in ELF communication: From pragmatic motives to lexico-grammatical innovation. *Nordic Journal of English Studies* 5 (2): 59-93.
- Cogo, Alessia & Martin Dewey. 2012. *Analysing English as a Lingua Franca: a Corpus driven investigation*. London: Continuum.
- Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe. 2016. CEFR Illustrative Descriptors. Extended version 2016. Pilot version for consultation. Available at: [http://www.coe.int/t/dg4/linguistic/cadre1\\_en.asp](http://www.coe.int/t/dg4/linguistic/cadre1_en.asp), accessed 23 April 2017.
- Crystal, David. 2003. *English as a Global Language* (2<sup>nd</sup> edition). Cambridge: Cambridge University Press.
- Dewey, Martin. 2007a. English as a lingua franca and globalization: an interconnected perspective. *International Journal of Applied Linguistics* 17 (3): 332-354.
- Dewey, Martin. 2007b. *English as a lingua franca: an empirical study of innovation in lexis and grammar*. PhD dissertation, King’s College London.
- Dewey, Martin. 2012. Towards a *post-normative* approach: learning the pedagogy of ELF. *Journal of English as a Lingua Franca*. 1 (1): 141-170.
- Dewey, Martin 2013. The distinctiveness of English as a Lingua Franca. *ELT Journal* 67 (3): 346-349.

- Dewey, Martin. 2014. Pedagogic Criticality and English as a Lingua Franca. *Journal of the Spanish Association of Anglo-American Studies*. 36 (2): 11-30
- The ELFA project. 2014. Available at <http://www.helsinki.fi/englanti/elfa/elfacorpus>, accessed April 2017.
- Graddol, David. 2006. *English next*. London: The British Council.
- Haberland, Hartmut. 2013. ELF and the bigger picture. *Journal of English as a Lingua Franca*. 2 (1): 195-198.
- House, Juliane. 2003. English as a lingua franca: A threat to multilingualism? *Journal of Sociolinguistics* 7 (4): 556-578.
- Hülmbauer, Cornelia. 2009. 'We don't take the right way. We just take the way that we think you will understand' – The Shifting Relationship between Correctness and Effectiveness in ELF. In Anna Mauranen & Elina Ranta (eds.). *English as a Lingua Franca: Studies and findings*. Newcastle upon Tyne: Cambridge Scholars Publishing, 323-347.
- Hynninen, Niina. 2014. The Common European Framework of Reference from the perspective of English as a lingua franca: what we can learn from a focus on language regulation. *Journal of English as a Lingua Franca*. 3 (2): 293-316.
- Illés, Éva. 2016. Issues in ELF-aware teacher education. *Journal of English as Lingua Franca*. 5 (1): 135-145.
- Jenkins, Jennifer. 2000. *The Phonology of English as an International Language*. Oxford: Oxford University Press.
- Jenkins, Jennifer. 2006. The spread of EIL: a testing time for testers. *ELT Journal*. 60 (1): 42-50.
- Jenkins, Jennifer. 2007. *English as a lingua franca: attitude and identity*. Oxford: Oxford University Press.
- Jenkins, Jennifer. 2012. English as a Lingua Franca from the classroom to the classroom. *ELT Journal*, 66 (4): 486-494.
- Jenkins, Jennifer, Alessia Cogo & Martin Dewey. 2011. State-of-the-art article: Review of developments in research into English as a lingua franca. *Language Teaching* 44 (3): 281–315.
- Kachru, B. B. 1985. Standards, codification and sociolinguistic realism: the English language in the outer circle. In R. Quirk and H.G. Widdowson (eds.). *English in the world: Teaching and learning the language and literatures*. Cambridge: Cambridge University Press. 11-30.

- Kaur, Jagdish. 2011. Raising explicitness through self-repair in English as a lingua franca. *Journal of Pragmatics* 43: 2704-2715.
- Kirkpatrick, Andy. 2012. English as an Asian Lingua Franca: the ‘Lingua Franca Approach and implications for language education policy. *Journal of English as a Lingua franca* 1 (1): 121-140.
- Kirkpatrick, Andy & Sophiaan Subhan. 2014. Non-standard or new standards or errors? The use of inflectional marking for present and past tenses in English as an Asian lingua franca. In Sarah Buschfeld, Thomas Hoffmann, Magnus Huber & Alexander Kautzsch (eds.). *The Evolution of Englishes*. Amsterdam: John Benjamins. 386-400.
- Kohn, Kurt. 2015. A pedagogical space for ELF in the English classroom. In Bayyurt, Yasemin & Sumru Akcan (eds.). *Current Perspectives on Pedagogy for English as a Lingua Franca*. 51-67.
- Mauranen, Anna. 2011. Learners and users - Who do we want corpus data from? In Fanny Meunier, Sylvie De Cook & Gaëtanelle Gilquin (eds.) *A Taste for Corpora: In honour of Sylviane Granger*. Amsterdam: John Benjamins Publishing Company. 155-171.
- Oxford Collocations Dictionary for students of English*. 2<sup>nd</sup> ed. Oxford: Oxford University Press.
- O’Keeffe, Anne & Michael McCarthy (eds). 2010. *The Routledge Handbook of Corpus Linguistics*. London; New York: Routledge.
- Önen, Serap. 2014. *Lexico-grammatical features of English as a lingua franca: A corpus based study of spoken interactions*. PhD dissertation. Istanbul University.
- O’Regan, John P. 2014. English as a Lingua Franca: An Immanent Critique. *Applied Linguistics* 35 (5): 533–552
- Phillipson, Robert. (2008), Lingua franca or lingua frankensteinia? English in European Integration and Globalization. *World Englishes* 27 (2): 250-267.
- Pitzl, Marie-Luise. 2012. Creativity meets convention: idiom variation and re-metaphorization in ELF. *Journal of English as a Lingua Franca* 1 (1): 27–55.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.
- Seidlhofer, Barbara. 2001. Closing a conceptual gap: The case for a description of English as a Lingua Franca. *International Journal of Applied Linguistics* 15 (2): 326-345.
- Seidlhofer, Barbara. 2004. Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics* 24: 209-239.

- Seidlhofer, Barbara. 2009. Common ground and different realities: World Englishes and English as a lingua franca. *World Englishes* 28 (2): 236-245.
- Seidlhofer, Barbara. 2011. *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.
- Sewell, Andrew. 2013. English as a lingua franca: ontology and ideology. *ELT Journal*. 67 (1): 3-10.
- Sifakis, Nicos C. 2014. ELF awareness as an opportunity for change: a transformative perspective for ESOL teacher education. *Journal of English as a Lingua Franca*. 3 (2): 317-335.
- Smith, L.E. (ed). 1983. *Readings in English as an International Language*. Oxford: Oxford University Press.
- Soler-Carbonell, Josep. 2014. Emerging ELF as an intercultural resource: Language attitude and ideologies. *Journal of English as a Lingua Franca*. 3 (2): 243-268.
- Soler-Carbonell, Josep with Hakan Karaoglu. 2015. English as an academic lingua franca in Estonia: students' attitudes and ideologies. In Andrew Linn, Neil Bermel & Gibson Ferguson (eds.) *Attitudes towards English in Europe*. Berlin, Boston: De Gruyter Mouton, 213-238.
- Sowden, Colin. 2012. ELF on a mushroom: the overnight growth of English as a Lingua Franca. *ELT Journal* 66 (1): 89-96.
- Tognini Bonelli, Elena. 2010. Theoretical overview of the evolution of corpus linguistics. In Anne O'Keeffe & Michael McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*. London, New York: Routledge. 14-27.
- Ur, Penny. 2010. English as a Lingua Franca: A Teacher's Perspective. *Cadernos de Letras (UFRJ)* 27: 85-92
- VOICE project. n.d. Available at [www.univie.ac.at/voice](http://www.univie.ac.at/voice), accessed April 2017.
- VOICE Project. 2007. "Mark-up conventions". VOICE Transcription Conventions [2.1]. Available at [http://www.univie.ac.at/voice/documents/VOICE\\_mark\\_up\\_conventions\\_v2-1.pdf](http://www.univie.ac.at/voice/documents/VOICE_mark_up_conventions_v2-1.pdf), accessed March 2017.
- VOICE Project. 2007. "Spelling conventions". VOICE Transcription Conventions [2.1]. Available at [http://www.univie.ac.at/voice/documents/VOICE\\_spelling\\_conventions\\_v2-1.pdf](http://www.univie.ac.at/voice/documents/VOICE_spelling_conventions_v2-1.pdf), accessed March 2017.
- Walker, Robin. 2010. *Teaching the Pronunciation of English as a Lingua Franca*. Oxford, New York: Oxford University Press.
- Walkinshaw, Ian & Andy Kirkpatrick. 2014. Mutual face preservation among Asian speakers of English a Lingua Franca. *Journal of English as a Lingua Franca*. 3 (2): 269-291.

Widdowson, H.G. 2012. ELF and the inconvenience of established concepts. *Journal of English as a Lingua Franca*. 1 (1): 5-26.

## **APPENDIX 1: INTERVIEW QUESTIONS**

### **WARM-UP:**

1. How are you today?
2. Have you taken part in or conducted any academic research before?
3. Is it a busy time for you right now at university?

### **MAIN QUESTIONS:**

1. When and how did you decide you would want to major in what you are studying at the moment?
2. What were your expectations before starting university and was the reality very different from them?
3. Of all the courses you are taking this term, which interests you the most, and why?
4. How many of the courses you have taken during your studies here have been in English? Does it make studying easier or more difficult for you?
5. At what age did you start learning English? Is it the usual age for starting in Estonia?
6. Are you satisfied with your English language proficiency (in terms of grammar, writing, vocabulary, speaking)? Please explain.
7. How much English do you need in your daily life? Do you rely on your English when you are in another country?
8. Do you agree with the idea that English is the language of communication in the world? Why/Why not?
9. Do you speak any other foreign languages besides English? Which other languages would you like to learn?
10. Have you considered moving to another country? What would be the reasons for such a change?

## **APPENDIX 2: PARTICIPATION INFORMATION SHEET AND CONSENT FORM**

The aim of this Master's thesis is to explore the lexico-grammatical features of English as a lingua franca in the Estonian context. To do this, a small corpus of spoken language will be collected and analysed to find features of English as a lingua franca that have been identified in already existing corpora.

Your participation in this study is voluntary, however, you will receive course credit for taking part and by agreeing to participate you will contribute greatly to the compiling of a larger corpus at the Department of English Studies, University of Tartu. Participation means taking part in an interview that lasts approximately 10 minutes. The interview will be recorded and transcribed afterwards. There are no right or wrong answers at the interview.

You will remain anonymous in the study and all the information that you provide will be used for academic purposes only. Upon your agreement to take part in this study, you will be given this information sheet to keep, asked to fill in the participant information questionnaire and sign the consent form.

If you have any questions, please contact me.

## CONSENT FORM

**Title of Project:** Lexico-grammatical features in the Estonian learner corpus of spoken English as compared to the lexico-grammatical features of English as a lingua franca.

**Name of Researcher:** Merli Kirsimäe

**Contact Details:**

**E-mail:** merli.kirsimae@gmail.com

**Tel:** +372 53463541

**Participant's Statement:**

I, \_\_\_\_\_  
confirm that I have received satisfactory explanation on the research project described above and have had the possibility to ask further questions. I understand that any data that I provide will be used for academic purposes only and I agree to take part in the study.

**Signed** \_\_\_\_\_ **Date** \_\_\_\_\_

**Researcher's Statement:**

I, \_\_\_\_\_  
confirm that I have explained the project carefully to the participant.

**Signed** \_\_\_\_\_ **Date** \_\_\_\_\_

**Interviewer's Statement:**

I, \_\_\_\_\_  
confirm that I am aware of the nature of this study and will keep the interviewee anonymous.

**Signed** \_\_\_\_\_ **Date** \_\_\_\_\_

### APPENDIX 3: PARTICIPANT INFORMATION QUESTIONNAIRE

Interview Number: \_\_\_\_\_

Speaker ID: \_\_\_\_\_

Date: \_\_\_\_ / \_\_\_\_ / 2016

*Your personal information will be treated as completely confidential.*

#### PERSONAL INFORMATION:

Name: \_\_\_\_\_

Surname: \_\_\_\_\_

Nationality: \_\_\_\_\_

Male / Female (please underline)

Age: \_\_\_\_\_

University: \_\_\_\_\_

Department: \_\_\_\_\_

Year: \_\_\_\_\_

E-mail address: \_\_\_\_\_

#### LINGUISTIC INFORMATION:

Mother Tongue: \_\_\_\_\_

Medium (Language) of Education in:

Kindergarten: \_\_\_\_\_

Primary School: \_\_\_\_\_

Basic School: \_\_\_\_\_

Secondary School: \_\_\_\_\_

University: \_\_\_\_\_

Age of first exposure to English: \_\_\_\_\_

Age when started to learn English: \_\_\_\_\_

Other languages you speak:

\_\_\_\_\_  
\_\_\_\_\_

## APPENDIX 4: TRANSCRIPTION CONVENTIONS



### Spelling conventions

Version 2.1 June 2007

<p style="text-align: center;"><b>1. CHARACTERS</b></p> <p>a b c d e f g h i j k l m n o p q r s t u v w x y z</p>	<p>Only alphabetic <b>Roman characters</b> are used in the transcript. No diacritics, umlauts or non-roman characters are permitted in the running text.</p>
<p style="text-align: center;"><b>2. DECAPITALIZATION</b></p> <p><u>Example:</u> S8: so you really can &lt;@&gt; control my english &lt;/@&gt;</p>	<p><b>No capital letters</b> are used except for marking emphasis (cf. mark-up conventions).</p>
<p style="text-align: center;"><b>3. BRITISH SPELLING</b></p> <p>British spelling</p>	<p>British English spelling is used to represent naturally occurring ELF speech. The Oxford Advanced Learner's Dictionary (<b>OALD</b>), <b>7<sup>th</sup> edition</b>, is used as the primary source of reference. If an entry gives more than one spelling variant of a word, the first variant is chosen. If there are two separate entries for British and American spelling, the British entry is selected.</p>
<p style="text-align: center;"><b>4. SPELLING EXCEPTIONS</b></p> <p style="text-align: center;">center, theater behavior, color, favor, labor, neighbor defense, offense disk program travel (-l-: traveled, traveler, traveling)</p>	<p>The 12 words listed on the left <b>and all their derivatives</b> are spelled according to American English conventions (e.g. colors, colorful, colored, to color, favorite, favorable, to favor, in favor of, etc.).</p>
<p><u>Example:</u> S2: we are NOT quite sure if it will REALLY be (.) privatized next year</p>	<p>In addition, <b>all words</b> which can be spelled using either an <i>-is</i> or an <i>-iz</i> morpheme are spelled with <b>-iz</b> (e.g. to emphasize, organizations, realization, recognized, etc.).</p>
<p style="text-align: center;"><b>5. NON-ENGLISH WORDS</b></p> <p><u>Example:</u> S1: &lt;L1de&gt; wieso <b>o</b>esterreich? {why austria} &lt;/L1de&gt;</p> <p><u>Example:</u> S3: &lt;LNfr&gt; c'est ferme? {is it closed} &lt;/LNfr&gt;</p>	<p>Non-English words are rendered in the standard variant of the original language (i.e. no non-standard dialect). The roman alphabet is always used, also in the case of languages like Arabic or Japanese. <b>No umlauts</b> (e.g. NOT österreich), <b>no diacritics</b> (e.g. NOT fermé) and <b>no non-roman characters</b> are permitted.</p>

<b>6. FULL REPRESENTATION OF WORDS</b>	
<p>Although words may not be fully pronounced or may be pronounced with a foreign accent, they are generally represented in standard orthographic form.</p> <p><u>Example:</u> S7: the students that (.) decide freely to enter (.) this kind of master knows (.) for example that <b>he</b> can (.) at the end achieve (.) sixty credits</p>	<p>Explanation: S7 is Italian and pronounces the <i>he</i> in <i>he can</i> as /ɪ/, swallowing the initial h. Nevertheless, this is regarded as a minor instance of L1 accent and therefore represented in standard orthography (<i>he</i>).</p>
<b>7. FULL REPRESENTATION OF NUMBERS, TITLES &amp; ABBREVIATIONS</b>	
oh/zero, two, three, ... one hundred, nineteen ten, eighteen twenty-seven, ...	Numbers are fully spelled out as whole words. British English hyphenation rules apply.
missis (for <i>Mrs</i> ), mister, miss, mis (for <i>Ms</i> ), doctor, professor, ...	Titles and terms of address are fully spelled out.
et cetera, saint thomas, okay, ...	Forms that are usually abbreviated in writing, but spoken as complete words are fully spelled out.
<b>8. LEXICALIZED REDUCED FORMS</b>	
cos gonna, gotta, wanna	Lexicalized phonological reductions are limited to the four on the left. All other non-standard forms are fully spelled out (e.g. /hæftə/ = <i>have to</i> ).
<b>9. CONTRACTIONS</b>	
<p>i'm, there're, how's peter, running's fun, ... i've, they've, it's got, we'd been, ... tom'll be there, he'd go for the first, ... we aren't, i won't, he doesn't, ...</p> <p>what's it mean, where's she live, how's that sound ... let's</p>	<p>Whenever they are uttered, all standard contractions are rendered. This refers to verb contractions with <i>be</i> (<i>am, is are</i>), <i>have</i> (<i>have, has, had</i>), <i>will</i> and <i>would</i> as well as <i>not</i>-contractions.</p> <p>Additionally, 's is used to represent <i>does</i> when reduced and attached to a <i>wh</i>-word. It is also used to represent the pronoun <i>us</i> in the contracted form <i>let's</i>.</p>



## **Mark-up conventions**

Version 2.1 June 2007

<b>1. SPEAKER IDS</b>	
S1: S2: ...	Speakers are generally numbered in the order they first speak. The speaker ID is given at the beginning of each turn.
SS:	Utterances assigned to more than one speaker (e.g. an audience), spoken either in unison or staggered, are marked with a collective speaker ID <b>SS</b> .
SX:	Utterances that cannot be assigned to a particular speaker are marked <b>SX</b> .
SX-f: SX-m:	Utterances that cannot be assigned to a particular speaker, but where the gender can be identified, are marked <b>SX-f</b> or <b>SX-m</b> .
SX-1: SX-2: ...	If it is likely but not certain that a particular speaker produced the utterance in question, this is marked <b>SX-1</b> , <b>SX-2</b> , etc.
<b>2. INTONATION</b>	
<u>Example:</u> S1: that's what my next er slide? does	Words spoken with rising intonation are followed by a question mark "?".
<u>Example:</u> S7: that's point two. absolutely yes.	Words spoken with falling intonation are followed by a full stop ".".
<b>3. EMPHASIS</b>	
<u>Example:</u> S7: er internationalization is a very IMPORTANT issue	If a speaker gives a syllable, word or phrase particular prominence, this is written in capital letters.
<u>Example:</u> S3: toMORrow we have to work on the presentation already	
<b>4. PAUSES</b>	
<u>Example:</u> SX-f: because they all give me different (.) different (.) points of view	Every brief pause in speech (up to a good half second) is marked with a full stop in parentheses.
<u>Example:</u> S1: aha (2) so finally arrival on monday evening is still valid	Longer pauses are timed to the nearest second and marked with the number of seconds in parentheses, e.g. (1) = 1 second, (3) = 3 seconds.

<b>5. OVERLAPS</b>	
<p><u>Example:</u> S1: it is your best &lt;1&gt; <b>case</b> &lt;/1&gt; scenario (.) S2: &lt;1&gt; <b>yeah</b> &lt;/1&gt; S1: okay</p>	<p>Whenever two or more utterances happen at the same time, the overlaps are marked with numbered tags: &lt;1&gt; &lt;/1&gt;, &lt;2&gt; &lt;/2&gt;,...</p> <p>Everything that is simultaneous gets the same number. All overlaps are marked in <b>blue</b>.</p>
<p><u>Example:</u> S9: it is (. ) to identify <b>some</b>&lt;1&gt;<b>thing</b> &lt;/1&gt; where (. ) S3: &lt;1&gt; <b>mhm</b> &lt;/1&gt;</p>	<p>All overlaps are approximate and words may be split up if appropriate. In this case, the tag is placed within the split-up word.</p>
<b>6. OTHER-CONTINUATION</b>	
<p><u>Example:</u> S1: what up till (. ) till twelve? S2: <b>yes=</b> S1: <b>=really</b>. so it's it's quite a lot of time.</p>	<p>Whenever a speaker continues, completes or supports another speaker's turn immediately (i.e. without a pause), this is marked by "=".</p>
<b>7. LENGTHENING</b>	
<p><u>Example:</u> S1: you can run faster but they have much <b>mo:re</b> technique with the ball</p>	<p>Lengthened sounds are marked with a colon ":".</p>
<p><u>Example:</u> S5: personally that's my opinion the: <b>er::m</b></p>	<p>Exceptionally long sounds (i.e. approximating 2 seconds or more) are marked with a double colon "::".</p>
<b>8. REPETITION</b>	
<p><u>Example:</u> S11: e:r i'd like to go t- t- <b>to to</b> this type of course</p>	<p>All repetitions of words and phrases (including self-interruptions and false starts) are transcribed.</p>
<b>9. WORD FRAGMENTS</b>	
<p><u>Example:</u> S6: with a minimum of (. ) of <b>participa-</b> S1: mhm S6: <b>-pation</b> from french universities to say we have er (. ) a joint doctorate or a <b>joi-</b> joint master</p>	<p>With word fragments, a hyphen marks where a part of the word is missing.</p>
<b>10. LAUGHTER</b>	
<p><u>Example:</u> S1: in denmark well who knows. <b>@@</b> S2: &lt;@&gt; yeah &lt;/@&gt; <b>@@</b> that's right</p>	<p>All laughter and laughter-like sounds are transcribed with the @ symbol, approximating syllable number (e.g. ha ha ha = @@@). Utterances spoken laughingly are put between &lt;@&gt; &lt;/@&gt; tags.</p>
<b>11. UNCERTAIN TRANSCRIPTION</b>	
<p><u>Example:</u> S3: i've a lot of very (<b>generous</b>) friends</p> <p><u>Example:</u> SX-4: they will do whatever they want because they are a compan(<b>ies</b>)</p>	<p>Word fragments, words or phrases which cannot be reliably identified are put in parentheses ( ).</p>

<b>14. NON-ENGLISH SPEECH</b>	
<u>Example:</u> S5: <L1de> <b>bei firmen</b> </L1de> or wherever	Utterances in a participant's first language (L1) are put between tags indicating the speaker's <b>L1</b> .
<u>Example:</u> S7: er this is <LNde> <b>die seite? (welche)</b> </LNde> is	Utterances in languages which are neither English nor the speaker's first language are marked <b>LN</b> with the language indicated.
<u>Example:</u> S4: it depends in in in <LQit> <b>roma</b> </LQit>	Non-English utterances where it cannot be ascertained whether the language is the speaker's first language or a foreign language are marked <b>LQ</b> with the language indicated.
<u>Example:</u> S2: erm we want to go t- to <LNvi> <b>xx xxx</b> </LNvi> island first of all	Unintelligible utterances in a participant's L1, LN or in an LQ are represented by <b>x</b> 's approximating syllable number.
<u>Example:</u> S4: and now we do the boat trip (1) <L1xx> <b>xxxxx</b> </L1xx> S3: mhm	Utterances in a language one cannot recognize are marked <b>L1xx</b> , <b>LNxx</b> or <b>LQxx</b> .
<u>Example:</u> S3: <L1fr> oui un grand carre { <b>yes like a big square</b> } </L1fr> (.) i <fast> think it would </fast> be better if we put the tables a <soft> different way </soft>	If possible, translations into English are provided between curly brackets { } immediately after the non-English speech.
<b>15. SPELLING OUT</b>	
<u>Example:</u> S1: and they (3) created some (1) some er (2) JARGON. do you know? the word JARGON? (.) <spel> <b>j a r-</b> </spel> <spel> <b>j a r g o n?</b> </spel> jargon	The <spel> </spel> tag is used to mark words or abbreviations which are spelled out by the speaker, i.e. words whose constituents are pronounced as individual letters.
<b>20. ANONYMIZATION</b>	
	A guiding principle of VOICE is sensitivity to the appropriate extent of anonymization. As a general rule, names of people, companies, organizations, institutions, locations, etc. are replaced by aliases and these aliases are put into square brackets [ ]. The aliases are numbered consecutively, starting with 1.
<u>Example:</u> S9: that's one of the things (.) that i (1) just wanted to clear out. (2) [S13]?	Whenever speakers who are involved in the interaction are addressed or referred to, their names are replaced by their respective speaker IDs.  A speaker's first name is represented by the plain speaker ID in square brackets [S1], etc.
<u>Example:</u> S6: so: (1) ei:ther MYself or mister [S2/last] or even boss (.) should be there every year	A speaker's last name is marked [S1/last], etc.
<u>Example:</u> S8: so my name is [S8] [S8/last] from vienna	If a speaker's full name is pronounced, the two tags are combined to [S1] [S1/last], etc.
<u>Example:</u> S2: that division is headed by (1) [first name3] [last name3] (1)	Names of people who are not part of the ongoing interaction are substituted by [first name1], etc. or [last name1], etc. or a combination of both.
<u>Example:</u> S5: erm she is currently head of marketing (and) with the [org2] (1)	Companies and other organizations need to be anonymized as well. Their names are replaced by [org1], etc.
<u>Example:</u> S1: i: i really don't wanna have a: a joint degree e:r with the university of [place12] (.)	Names of places, cities, countries, etc. are anonymized when this is deemed relevant in order to protect the speakers' identities and their environment. They are replaced by [place1], etc.

## APPENDIX 5: EXAMPLE INTERVIEW

### Speakertable

#### PM01

Sex: m  
 Languages used: eng  
 L1: pol

#### EF10

Sex: f  
 Languages used: eng  
 L1: est

[1]

	0 [00:00.0]	1 [00:02.0]	2 [00:03.1]	3 [00:04.0]	4 [00:04.7]
PM01 [v]	so hello my name is [PM01] [EF10] [EF10] <1>right</1> okay so				
EF10 [v]	hi. [EF10] [EF10] <1>mhmh</1>				

[2]

	5 [00:07.4]	6 [00:10.1]
PM01 [v]	how are you today [EF10] mhmh (.) okay. so: er is it	
EF10 [v]	(1) i'm quite fine. thank you	

[3]

	7 [00:16.8]
PM01 [v]	your first time? have you taken part in any academic research before?
EF10 [v]	no i haven't. it

[4]

	8 [00:18.7]
PM01 [v]	(.) alright so:: by the way er er is it a busy time for you at the
EF10 [v]	's my first time.

[5]

	9 [00:25.4]
PM01 [v]	university right now
EF10 [v]	(1) i think it's going to be be busy soon because of the exams

[6]

	10 [00:31.4]	11 [00:35.0]
PM01 [v]	<2>mhmh</2>	mhmh (.) alright (.) so::
EF10 [v]	are coming <2>soon</2> but (.) right now i'm kind of free.	

[7]

..

<b>PM01 [v]</b>	er it was the warm up part of the interview? = (let) just meet each other make
<b>EF10 [v]</b>	=oh=

[8]

..

<b>PM01 [v]</b>	<3>yours</3> make ourselves comfortable and now. so. when and how did you
<b>EF10 [v]</b>	<3>mhmh</3>

[9]

.. 12 [00:51.3]

<b>PM01 [v]</b>	decide that you want to major? in what you are studying right now.
<b>EF10 [v]</b>	(2) it wa:::s <spel>a</spel>

[10]

.. 13 [00:56.8]

<b>PM01 [v]</b>	mhm=
<b>EF10 [v]</b>	hard (.) challenge for me? =to figure it out? (.) i think i figure it out (.) last

[11]

.. 14 [01:01.9] 15 [01:02.2] 16 [01:04.7] 17 [01:07.4]

<b>PM01 [v]</b>	mhmh? so really recently you have a good memories of it
<b>EF10 [v]</b>	year? (.) because (1) yeah (.) i

[12]

.. 18 [01:11.3]

<b>PM01 [v]</b>	=mhmh=
<b>EF10 [v]</b>	didn't know what to do with my future? =so i just kinda. winged it? and

[13]

.. 19 [01:17.0] 20 [01:19.6]

<b>PM01 [v]</b>	mhmh=
<b>EF10 [v]</b>	went (.) but i think i did a right choice =i'm studying to be a teacher. (.)

[14]

.. 21 [01:21.4] 22 [01:22.2]

<b>PM01 [v]</b>	teacher. teacher of what exactly? teacher of english? yeah. it's really nice (.)
<b>EF10 [v]</b>	english?

[15]

..

**PM01 [v]** so (1) e::rm any particular event was connected? to the your decision? or just e::rm

[16]

.. 23 [01:37.2]

**PM01 [v]** it was slow and m: m: complex process of you figuring it out

**EF10 [v]** (2) well (1) it wa::s

[17]

.. 24 [01:45.9]

**PM01 [v]** mhmh=

**EF10 [v]** really slow but (.) i think my estonian teacher =really helped me that. i didn't

[18]

..

**PM01 [v]**

**EF10 [v]** realise it at first? but now if i think about it then (.) i think my estonian teacher

[19]

.. 25 [01:55.5] 26 [01:57.9]

**PM01 [v]** mhmh= =because she was supportive as a

**EF10 [v]** really helped me. =with this decision yea=

[20]

.. 27 [02:02.0]28 [02:03.8]

**PM01 [v]** language techer herself? maybe? no-i she was a friend of yours (maybe)

**EF10 [v]** (2)

[21]

.. 29 [02:05.8]

**EF10 [v]** she was just very kind and she really wanted to know where we go and what do we

[22]

.. 30 [02:13.7] 31 [02:16.2]32 [02:16.7]

**PM01 [v]** <4>mhmh</4> mhmh

**EF10 [v]** do with our futures <4>and</4>she really wanted the best for us so we had a

[23]

	..	33 [02:20.7]	
<b>PM01 [v]</b>			okay that's really nice. (.) so er what were your
<b>EF10 [v]</b>			conversation and (.) yeah (.)

[24]

	..		
<b>PM01 [v]</b>			expectations before going to the university and the reality was it very different from

[25]

	..	34 [02:31.1]	
<b>PM01 [v]</b>			them
<b>EF10 [v]</b>			(1) i expected to it to be (.) pretty much the same as it wa::s the other schools

[26]

	..	35 [02:42.1]36 [02:42.6]	
<b>PM01 [v]</b>			mhmh
<b>EF10 [v]</b>			i was in but (.) it's really different because the classes are (.) like mondays

[27]

	..	37 [02:52.2]	
<b>PM01 [v]</b>			mhmh
<b>EF10 [v]</b>			are totally free? (.) er but tuesdays are full with classes like until to-i eight

[28]

	..	38 [02:52.7]	
<b>EF10 [v]</b>			so it is really different. plus i have to do my own (1) what is it called (1) student

[29]

	..	39 [03:00.0]	40 [03:02.6]
<b>PM01 [v]</b>			mhmh <5>okay?</5> sure. (.) so er of all courses? you
<b>EF10 [v]</b>			plan? <5>what lessons</5> do i attend (.)

[30]

	..	41 [03:09.4]	42 [03:10.6]
<b>PM01 [v]</b>			are taking this term? er which interests you the most. psychology. (1)
<b>EF10 [v]</b>			psychology?

[31]

	..	43 [03:13.1]	44 [03:20.8]
<b>PM01 [v]</b>	okay. why.		the empathy? is the
<b>EF10 [v]</b>		(1) i like to:: think (2) i don't know why hm:.	(1)

[32]

	..	45 [03:22.5]	46 [03:23.6]47 [03:24.1]
<b>PM01 [v]</b>	word?	okay	
<b>EF10 [v]</b>	mhmh (.)	plus i like if i see a person? i like to think what they are t	

[33]

	..		
<b>PM01 [v]</b>		=mhmh=	
<b>EF10 [v]</b>	thinking why they are doing what they are doing=	=and (.) i like to help	

[34]

	..	48 [03:37.3]	
<b>PM01 [v]</b>		okay. (1) okay i think that i see	
<b>EF10 [v]</b>	people so psychology really helps in that area? (1)		

[35]

	..		
<b>PM01 [v]</b>	that (.) so ehm (1) how many of the courses (.) y-you've taken in the university		

[36]

	..		
<b>PM01 [v]</b>	were in english and er did it make studying them easier? or more difficult for you.		

[37]

	..	49 [03:54.9]	
<b>PM01 [v]</b>		<6>mhmh</6>	
<b>EF10 [v]</b>	(1) i have two lessons in english <6>right now?</6> (2) (but) they're both english		

[38]

	..	50 [04:04.8]	
<b>PM01 [v]</b>	<7>oh okay</7>	mhmh? (.) because well yeah the	
<b>EF10 [v]</b>	<7><@>so</@></2>@@ so it makes way easier.		

[39]

	..	51 [04:11.0]
<b>PM01 [v]</b>	subject matter is english	<8>itself okay @@</8> (1) okay. so at what age did you
<b>EF10 [v]</b>		<8>mhmh exactly</8>

[40]

	..	52 [04:19.0]
<b>PM01 [v]</b>	start learning english and is it the usual of starting in estonia?	
<b>EF10 [v]</b>		(1) i started in third

[41]

	..	53 [04:22.3]
<b>PM01 [v]</b>	mhmh=	
<b>EF10 [v]</b>	grade? =i don't know what age i was (then)? (.) but it's (.) pretty typical. <9>i	

[42]

	..	55 [04:31.2]56 [04:31.6]
<b>PM01 [v]</b>	<9>mhmh</9>	mhmh
<b>EF10 [v]</b>	know</9> lot of people also start in the first grade?	(.) s but i did (on the

[43]

	..	57 [04:33.8]	58 [04:36.0]
<b>PM01 [v]</b>	(.) yeah but=	(1) e something around the beginning of eme-	
<b>EF10 [v]</b>	third)? =eleven maybe?		

[44]

	..	59 [04:40.6]
<b>PM01 [v]</b>	elementary school. <10>right? okay.</10> (.) so er (.) are you satisfied? with your	
<b>EF10 [v]</b>	<10>mhmh mhmh</10>	

[45]

	..	
<b>PM01 [v]</b>	english language proficiency? in terms of vocabulary grammar speaking (.) and er	

[46]

	..	60 [04:51.4]	61 [04:54.6]62 [04:55.3]
<b>PM01 [v]</b>	why why not	mhmh	
<b>EF10 [v]</b>	(.) i'm not very satisfied (with my) speaking?	(.) because i (.)	

[47]

.. 63 [04:58.1] 64 [04:58.6]

PM01 [v] =mhmh  
 EF10 [v] can't speak? @@= but i like to i like my grammar? i like to write in english?

[48]

.. 65 [05:05.6]

PM01 [v] mhm. (1) alright (.) so (.) er how much  
 EF10 [v] (.) so i think i'm pretty good at that.

[49]

PM01 [v] english. do you need in your daily life. (.) and would you say that you rely on

[50]

.. 66 [05:18.2]

PM01 [v] english while being in another countries.  
 EF10 [v] (1) i think i need english (.) english (1)

[51]

.. 67 [05:25.7]

PM01 [v] mhm=  
 EF10 [v] pretty much? in the internet? = i chat with my friends? (1) from england too in

[52]

.. 68 [05:31.5] 69 [05:31.9]

PM01 [v] mhmh  
 EF10 [v] english? (2) i would say (1) (i mean) i need it (.) i couldn't (.) live without it

[53]

.. 70 [05:40.3]

PM01 [v] alright (1) so er about the other part of the question er while being  
 EF10 [v] right now? (.)

[54]

.. 71 [05:51.1]

PM01 [v] abroad do you rely on english just you know for your communication.  
 EF10 [v] (2) i wouldn't



[63]

	..	83 [06:49.4]	84 [06:55.6]
<b>PM01 [v]</b>	learn in the future?		mhmh=
<b>EF10 [v]</b>	(1) u:::hm:: (1) i want to learn more france?		=and korean.

[64]

	..	85 [06:57.3]
<b>PM01 [v]</b>	oh. it's really nice? e::r you're from here? right? okay. and e::r you do not have	
<b>EF10 [v]</b>	(.)	

[65]

	..	86 [07:08.0]
<b>PM01 [v]</b>	any (wh) asian roots? whatsoever? <11>no. okay</11> er i'm really sorry you	
<b>EF10 [v]</b>	no?<11>@@@</11>	

[66]

	..	87 [07:12.7]	88 [07:15.2]
<b>PM01 [v]</b>	kind of look like that.		so have you considered? moving to
<b>EF10 [v]</b>	oAH <@>okay</@> (1)		

[67]

	..	89 [07:18.1]	90 [07:21.2]
<b>PM01 [v]</b>	another country?		mhmh=
<b>EF10 [v]</b>	(2) not permanently?		=i('d) like to go to teach english in

[68]

	..	91 [07:28.0]	92 [07:28.6]	93 [07:30.6]
<b>PM01 [v]</b>			mhmh	okay.
<b>EF10 [v]</b>	other countries but i wish to come back here		i want to live in estonia.	

[69]

	..
<b>PM01 [v]</b>	and er (1) let's just say if you would yeah? what would be the reasons for such a

[70]

	..	94 [07:39.9]
<b>PM01 [v]</b>	change. (.) teaching english yeah that would be one	
<b>EF10 [v]</b>	mhmh. teaching english (.)	

[71]

.. 95 [07:46.7]

**PM01 [v]** live o:r just travel maybe.  
**EF10 [v]** uhm (.) why i want to. like. live in another country?

[72]

96 [07:50.7]

97 [07:55.8]98 [07:56.8]

**PM01 [v]** mhmh  
**EF10 [v]** (1) to:: (1) see the world wider? to see different people? their cultures (.) to

[73]

.. 99 [08:04.7]

**PM01 [v]** mhmh? okay. that's really fun. so thank you for  
**EF10 [v]** get out of my (.) little estonian box?

[74]

.. 100 [08:10.7]

**PM01 [v]** this interview? that was the last question

## RESÜMEE

TARTU ÜLIKOOL  
ANGLISTIKA OSAKOND

**Merli Kirsimäe**

**The compilation and lexicogrammatical analysis of an Estonian spoken mini-corpus of English as a lingua franca / Inglise keele kui *lingua franca* suulise kõne mini-korpuse koostamine Eesti kontekstis ning selles esineva sõnavara ja grammatika analüüs**

Magistritöö

2017

Lehekülgede arv: 71

Käesoleval magistritööl on mitu eesmärki. Esimeseks neist on anda lühiülevaade uurimisvaldkonnast, milleks on inglise keel kui *lingua franca* (*English as a lingua franca – ELF*), ning tuua välja selle valdkonna seosed inglise keele õpetamisega. Teiseks eesmärgiks on kirjeldada *ELFi* suulise kõne mini-korpuse loomist Eesti kontekstis ning analüüsida korpuses esinevaid sõnavaralisi ja grammatilisi erisusi, võrreldes neid varasemates teadustöodes välja toodud uuenduslike tunnustega.

Töö esimeses peatükis keskendutakse teooriale ning selgitatakse mõistet ‘inglise keel kui *lingua franca*’ täpsemalt, andes ülevaate valdkonna kujunemisloost ning tähtsusest tänapäeva globaalse inglise keele kontekstis. Esimesest peatükist saab lugeja teada veel tuntumatest suulise kõne korpustest, mis on valdkonnas koostatud, ning ka kahest väiksemast korpusest, mis on antud magistritöö jaoks oluliseks võrdlusmaterjaliks. Lisaks sellele tuuakse välja puutepunktid inglise keele õpetamisega ja ideed, kuidas neid kaht valdkonda – *ELFi* ja inglise keele õpetamist teineteisele lähemale tuua. Peatüki viimases osas võib lugeda kriitikast, mis *ELF*-koolkonnale osaks on saanud.

Töö teises, empiirilises peatükis tegeletakse mini-korpusega, mis töö jaoks koostati. Kirjeldatakse detailselt korpuse koostamist intervjuude läbiviimise teel, kuidas intervjuusid transkribeeriti ning sõnavaralisi ja grammatilisi uuendusi silmas pidades märgendati. Täpsemalt vaadatakse, kuidas on korpuses kasutatud artikleid, eessõnu ja verbe (*take, get, make*), millel on semantiliselt palju tähendusi ja kombineerimisvõimalusi. Tulemuste analüüs ja arutelu kinnitavad sarnaste joonte olemasolu Eesti *ELF* suulise kõne mini-korpuses ja varasemates korpustes, mis on samuti keskendunud artiklite, eessõnade ja mainitud verbide kasutusele. Inglise keele õpetamise seisukohast võiks magistritöö olla üheks algmaterjaliks inglise keele õpetajatele ja õpetajaks õppijaile Eestis, kes sooviksid tõsta oma teadlikkust inglise keelest kui *lingua franca*’st ning seda pikkamööda ka oma õpilastele edastada. Töö kokkuvõttes jõutakse seisukohale, et antud mini-korpust tuleks analüüsida *ELFi* tunnuseid silmas pidades veel põhjalikumalt ning korpust ka suurendada, ent käesolev magistritöö on juba teinud suure panuse inglise keele kui *lingua franca* uurimisse Eestis ning tõestanud mõningate varasematest uurimustest leitud tunnuste esinemist Eesti *ELF* kõneleajate keeles.

Märksõnad: inglise keel kui *lingua franca*, suulise kõne korpus, sõnavaralis-grammatilised uuendused, inglise keele õpetamine

**Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Merli Kirsimäe,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

**The compilation and lexicogrammatical analysis of an Estonian spoken mini-corpus of English as a lingua franca,**

mille juhendaja on Jane Klavan,

- 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
  3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 16.05.2017

Merli Kirsimäe