"But English is something useful" – Typological differences in
spoken domains of English as a lingua franca

Saara Vasko, 274780

MA Thesis

October 2020

Department of Humanities

University of Eastern Finland

ITÄ-SUOMEN YLIOPISTO – UNIVERSITY OF EASTERN FINLAND

Faculty The Philosophical Faculty	**School** Department of Humanities			
Author Saara Vasko				
Title "*But English is something useful*" – Typological differences in spoken domains of English as a lingua franca				
Main subject English language and culture	**Level** Pro gradu -tutkielma Sivuainetutkielma Kandidaatin tutkielma Aineopintojen tutkielma	x	**Date** 23.10.2020	**Number of pages** 80

Abstract

This study set out to discover whether there is internal typological genre variation in spoken English as a lingua franca (ELF). From previous research it is known that spoken Standard British English (Szmrecsanyi 2009) and written ELF (Laitinen 2018) do exhibit this genre specific variation. In order to determine the typology of spoken ELF genres, the methodology by Szmrecsanyi (2009) was adopted and revised to suit this study. In this study, typology was defined to consist of analyticity and syntheticity; free standing and bound grammatical markers. These markers were searched for in the VOICE (*Vienna-Oxford International Corpus of English*) corpus, which provided the spoken ELF data.

The VOICE is divided into five domains: educational (ED), leisure (LE), and to three professional domains of research and science (PR), organisational (PO), and business (PB). These domains function as the different spoken ELF genres for this study. These domains were analysed for the analytic and synthetic markers using the AntConc corpus tool. The received marker frequencies inform the analyticity (AI) and syntheticity (SI) indices of each domain. These domain results were firstly examined statistically and in contrast to each other, after which a native comparison to a spoken Standard British English was executed. Szmrecsanyi's (2009) results on BNC (*British National Corpus*) genre variation were utilised as the comparison data.

The results derived from this study indicate that there is internal domain specific variation in spoken ELF. Nearly all loglikelihood statistical tests between the domains yielded significant results. The VOICE domains were recorded to vary more in analyticity but nonetheless, also the syntheticity varied significantly between majority of the domain comparisons. These results suggest that there is typological variation in ELF domains and thus, ELF users appear to accommodate their speech according to speech environment. The comparison with Standard British English spoken genres implies in turn that the spoken domains in ELF and native spoken genres share resemblances in how the domains and genres are distributed inside each variety. It was discovered in this study that for example, more formal and academic genres and domains portray increased syntheticity. While these similarities could be interpreted for some domain and genre comparisons it was not a valid result in all comparisons. For instance, the LE (leisure) domain exhibits high levels of syntheticity among the VOICE domains whereas in BNC the leisurely, informal speech genres are lower in syntheticity than the corpus mean.

These results endorse the conception that ELF users are aware of linguistic conventions and accommodate their speech accordingly. However, there are some limitations to this study and impetus for future research. Most potently, the five domains of VOICE are an inadequate representation of ELF. A more precise and wider selection of domains or genres would illustrate better the internal variation of ELF and enable a more precise juxtaposition with other varieties. However, despite these improvement suggestions, this study did discover support for the hypothesis that there is internal variation in spoken ELF and that in some respect the variation is similar to the variation of a native counterpart. In future research it would be interesting to see a similar study conducted on spoken ELF with a more recent corpus data and with a more comprehensive genre division. This way spoken ELF could be examined diachronically and compared more accurately with other World Englishes.

ITÄ-SUOMEN YLIOPISTO – UNIVERSITY OF EASTERN FINLAND

Tiedekunta	Osasto
Filosofinen tiedekunta	Humanistinen osasto

Tekijät				
Saara Vasko				

Työn nimi				
"*But English is something useful*" – Typological differences in spoken domains of English as a lingua franca				

Pääaine	Työn laji		Päivämäärä	Sivumäärä
Englannin kieli ja kulttuuri	Pro gradu -tutkielma	x	23.10.2020	80
	Sivuainetutkielma			
	Kandidaatin tutkielma			
	Aineopintojen tutkielma			

Tiivistelmä

Tämä tutkimus pyrki selvittämään, osoittaako puhuttu lingua franca -englanti (ELF) sisäistä genrekohtaista typologista vaihtelua. Aikaisemmasta tutkimuksesta tiedetään, että puhuttu brittienglannin yleiskieli (Szmrecsanyi 2009) ja kirjoitettu ELF (Laitinen 2018) ilmentävät tätä genretason typologista variaatiota. Puhutun ELF:n genrekohtaisen typologian selvittämiseksi Szmrecsanyin (2009) kehittämä metodologia sovitettiin tähän tutkimukseen sopivaksi. Tässä tutkimuksessa typologia määritellään koostuvaksi analyyttisyydestä ja synteettisyydestä eli itsenäisistä ja sidotuista kieliopillisista markkereista. Näitä kieliopillisia markkereita haettiin VOICE (*Vienna-Oxford International Corpus of English*) -korpuksesta, joka toimi datana puhutulle ELF:lle.

VOICE on jaettu viiteen osa-alueeseen: koulutukselliseen (ED) alueeseen, vapaa-ajan (LE) alueeseen, sekä kolmeen ammatilliseen osa-alueeseen, joita ovat tutkimus ja tiede (PR), järjestöllinen toiminta (PO) ja liiketoiminta (PB). Nämä osa-alueet edustavat puhutun ELF:n eri genrejä tässä tutkimuksessa. Osa-alueita analysoitiin etsien analyyttisiä ja synteettisiä markkereita AntConc -korpustyökalun avulla. Analyysista saadut markkerifrekvenssit kertovat analyyttisyys- (AI) ja synteettisyysindeksit (SI) jokaiselle korpuksen osa-alueelle erikseen. Tuloksia tarkasteltiin ensin tilastollisesti ja osa-alueita toisiinsa vertaillen sekä myös puhutun brittienglannin genreihin verrattuna. Szmrecsanyin (2009) tulokset BNC (*British National Corpus*) -korpuksen genrevariaatiosta toimivat natiiviverrokkina.

Tutkimuksesta saadut tulokset viittaavat siihen, että puhutussa ELF:ssä on sisäistä osa-aluekohtaista variaatiota. Lähes kaikki loglikelihood -tilastotestit antoivat merkittäviä lukemia. Tarkastelussa olleet VOICE-osa-alueet osoittivat enemmän variaatiota analyyttisyydessä, vaikka myös synteettisyys vaihteli merkittävästi enemmistössä osa-aluevertailuista. Nämä tulokset osoittavat, että myös puhuttu ELF vaihtelee typologisesti osa-alueittain. Siten voidaan myös päätellä, että ELF:n käyttäjät mukauttavat kieltään puhetilanteen mukaan. Vertailu brittienglannin kanssa puolestaan antaa ymmärtää, että ELF:n osa-alueet ja natiivigenret jakavat samankaltaisuuksia siinä, miten osa-alueet ja genret sijoittuvat typologisesti kummankin varieteetin sisällä. Tämä tutkimus paljasti esimerkiksi sen, että muodollisemmat ja akateemisemmat genret ja osa-alueet ovat synteettisempiä verrattuna muihin osa-alueisiin. Vaikka näitä samankaltaisuuksia oli tulkittavissa osassa osa-alue- ja genrevertailuista, typologinen vastaavuus ei ulottunut kaikkiin vertailupareihin. Esimerkiksi vapaa-ajan (LE) osa-alue osoittautui synteettisemmäksi kuin VOICE-keskiarvo, kun taas BNC:ssä vapaa-ajan genret olivat synteettisyydessä huomattavasti alle korpuksen keskiarvon.

Tästä tutkimuksesta saadut tulokset vahvistavat käsitystä, jonka mukaan ELF:n käyttäjät ovat tietoisia englannin kielellisistä konventioista ja he mukauttavat puhettaan vaaditun konvention mukaisesti. On kuitenkin huomioitava, että tutkimuksella oli joitakin rajoitteita, ja toisaalta jatkotutkimukselle voidaan esittää parannusehdotuksia. Merkittävimpänä rajoitteena voidaan pitää sitä, että viisi puheen osa-aluetta ei edusta ELF:iä tarpeeksi kattavasti. Tarkempi ja laajempi valikoima osa-alueita tai genrejä havainnollistaisi paremmin ELF:n sisäistä variaatiota ja mahdollistaisi myös täsmällisemmän vertailun muiden varieteettien kanssa. Näistä parannusehdotuksista huolimatta tämä tutkimus löysi tukea hypoteesille, jonka mukaan puhuttu ELF varioi sisäisesti ja että tämä variaatio on osittain samankaltaista kuin natiiviverrokilla. Tulevaisuudessa olisi mielenkiintoista nähdä samankaltainen tutkimus puhutusta ELF:stä uudemmalla korpusdatalla ja kattavammalla genrejaolla. Näin puhuttua ELF:iä kyettäisiin tarkastelemaan diakronisesti ja sitä voitaisiin vertailla täsmällisemmin suhteessa muihin maailman englanteihin.

Avainsanat
Lingua franca englanti, typologia, puhuttu englanti, tekstityyppivariaatio

CONTENTS

1.   INTRODUCTION

Like the quotation in the title of this paper shows, English is not merely a language but a useful communicative tool. During the 20th and 21st centuries the status of English has changed, and it has become the language of global business and communication. English is extensively relied on even among non-native speakers. This English as a Lingua Franca (hereon ELF) use of English connects people who do not share a common native language (Mauranen 2018, 7). In 2020, English is estimated to be the most spoken language in the world surpassing Mandarin Chinese (Eberhard et al. 2020). English is now so widely used by non-natives and throughout different areas of life that ELF cannot only be considered a contact language. It has acquired and developed distinct characteristics. So much so, that linguists are debating over ELF being acknowledged as its own distinguished variety of English, although this view is not generally acknowledged (e.g. Prodromou 2008; Laitinen 2018; Björkman 2018). Nevertheless, this changed status of ELF has attracted interest among linguists which has led to it being studied comprehensively in all areas of linguistics.

This research examines ELF, paying attention to its typological composition through grammatical markers. The aim is to determine how different spoken ELF domains, speech environments with specific modes of operation like business meetings for instance, differ from each other and from a native British English variety when compared typologically. Linguistic typology studies the differences and similarities of languages (Caffarel et al. 2004, 1). I will concentrate on grammatical typology, and more specifically, analyticity and syntheticity in ELF. The terms analytic and synthetic describe whether a language uses free standing grammatical markers (analytic; e.g. prepositions: *piece **of** cake*) or bound grammatical markers (synthetic; e.g. inflected verbs: *walk**ed***) (Kortmann and Szmrecsanyi 2011, 280). Linguistic typology attempts thus to describe the languages of the world according to, in the opposite ends

of the spectrum, analyticity and syntheticity. Most languages comprise of varying degrees of analyticity and syntheticity. For example, Turkish is highly synthetic with few analytic markers. (ibid., 264-286) Language typology in relation to English is discussed further below in the theory sections.

In more detail, this thesis is interested in whether spoken ELF varies in its grammatical marker composition among the different domains. Furthermore, spoken ELF is compared to a spoken standard British English to see if the domains behave similarly in contrast to a native variety. This study is inspired by previous research (Szmrecsanyi 2009; Laitinen 2018) which is used as a comparison as well as a source. Szmrecsanyi (2009, 339) suggests that there is variation within native genres (or text types) of English. He (2009, 334) maps (Figure 1) multiple native genres from the BNC (British National Corpus) according to their analyticity and syntheticity. The scatterplot exhibits the significant variability of the native genres. These Szmrecsanyi's results add further interest on ELF. Furthermore, spoken ELF was chosen as the interest of this study, since Laitinen (2018) in his study establishes that in many written genres ELF does structurally follow native grammatic compositions. Thus, a similar assessment on spoken ELF is due. While the frequencies of analyticity and syntheticity vary between native genres and written ELF imitates the native genres, it proposes a question whether the same holds true for spoken domains of ELF.
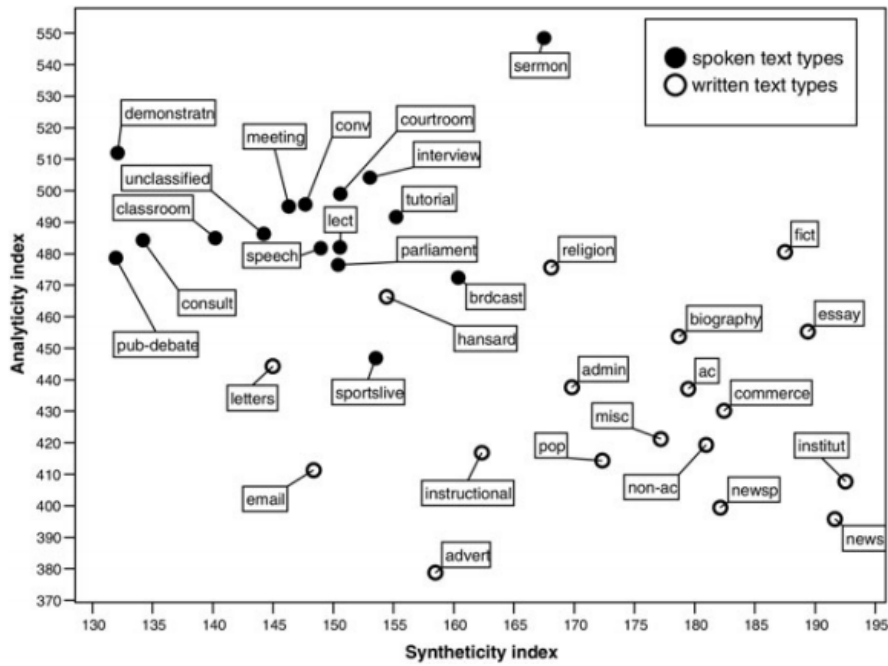
FIGURE 2. BNC macro registers: analyticity by syntheticity (in index points, ptw).

Figure 1. *Standard British English genres mapped according to grammatical typology by Szmrecsanyi (2009, 334).*

The previous research by Szmrecsanyi (2009) shows that a standard British English can be typologically distinguished between the written and spoken means. Furthermore, the Standard British English exhibits further internal variability in written and spoken modes on the level of genres. Similar genre variation is reported in written ELF by Laitinen (2018). This background knowledge on ELF typology gives impetus to this study. Spoken ELF is not yet examined according to different genres. What is more, spoken ELF is interesting in the respect that it is non-native speech. While Laitinen (2018) observed that written ELF conforms to native constructions, speech cannot be edited thus being more authentic language use. Despite native-like proficiency in written English, spoken ELF reveals the actual utterances delivered on the spot without editing. Therefore, any visible variability in spoken ELF genres is fascinating. In order to establish the variation in spoken ELF genres, the research questions formulated are as follows:

1. What are the frequencies of the analytic and synthetic markers in ELF in different spoken domains?
2. How do these frequencies of markers compare between the different spoken domains?
3. How do the spoken ELF domain marker frequencies compare with native spoken genres?

This study is conducted as corpus based and the corpus used is *the Vienna-Oxford International Corpus of English* (The VOICE). It offers spoken ELF data free and readily compiled online and as downloadable text files on *univie.ac.at/voice/*. The corpus is divided into five domains that represent different spoken genres: educational, leisure, professional business, professional organizational, and professional research and science (VOICE 2013). The VOICE uses the term *domain* which is adopted in this study as well to avoid confusion. However, in other instances this study favours the term *genre* to unify terminology, although different sources have opted for altering expressions when talking about the data gathered from varying language environments.

The results from this study could give further confirmation that ELF users are aware of structural linguistic customs of English also in speech. Theory behind this and relevant information relating to it is discussed in depth below. First however, English as a lingua franca is discussed in length, followed by a section on linguistic typology. Previous studies by Szmrecsanyi (2009), Szmrecsanyi and Kortmann (2011), and Laitinen (2018) are presented in more detail due to their relevance to this current thesis. The methodology and data are explained in a separate section. Finally, results and discussion sections reveal what was unearthed about spoken ELF from this typological point of view.

2.   ENGLISH AS A LINGUA FRANCA

2.1. What and whose is English today?

Before we can begin to understand ELF, a thorough understanding of what English language in itself is today is required. This is because English in the 21st century is not only a first language of some people anymore. The generally acknowledged fact that the native users have control and authorship of the language does not hold for English as pervasively. For instance, learners of German or Spanish are taught and encouraged to aim for a standard, like Swiss standard for German or Argentinian standard for Spanish. It is true that people's attitudes still favour British and American standards for English (see e.g. Leppänen et al. 2011, (Figure 20), for Finnish attitudes. For instance, Finns still find Standard British and American Englishes most appealing.). However, the global development of English becoming ever more widespread has led to its evolution as a language. A majority of the speakers are no longer native users of the language but speak English as a second, third or even subsequent language (Björkman 2013, 4). Indeed, English has acquired a status as no other language before, connecting speakers all around the world (Björkman 2018, 255). English is now used for communication in situations where all speakers could be non-natives. This means extensive English as a lingua franca interactions. In these circumstances, the language itself does not have significant value, it is merely "adapted to suit their [lingua franca user's] own communicative requirements" (Widdowson 2018, 101). Indeed, in the Finnish curriculum English is emphasised as a communication language with people from all over the world and aim for Standard English competence is not mentioned (Opetushallitus 2014, 242 – 247). Thus, it appears that English has as a global international language divided into two; there are the standard native and nativized varieties, and the communication English used increasingly by the non-natives.

This reduction of English to a communicative tool means that there are Englishes spoken around the world without actual speech communities (Widdowson 2018, 106). This in turn induces the

question whether the language used is in fact the same English the natives use. This is why the form World Englishes is essential as it allows us to comprehend the plurality of the English language (Widdowson 2018, 105). This plural suggests that English has been extended beyond its native varieties and the non-native users have been granted their own English. However, with the plurality come some difficulties. The World Englishes mentioned above are divided, as the name implies, into several Englishes, ELF being one of them. The definitions of these Englishes are quite subtle but nevertheless deal with their own domain of English use. Therefore, the next section defines ELF in more detail.

2.2. Defining ELF
2.2.1. What constitutes English as a lingua franca
The traditional definition of a lingua franca comprehends it as a contact language between speakers who come from different first language backgrounds (Dewey and Jenkins 2010, 72). However, as the previous section portrays, English does not really follow the traditional language practices like other languages anymore. Therefore, also English as a lingua franca requires an updated and specified definition. Since it is so widespread and so extensively used, ELF functions on a different level from other lingua francas. Mauranen (2018) demonstrates this in a comprehensive description of ELF, illustrating the many dimensions it extends to:

> ELF is not just a contact language where English is a domestic language or otherwise especially salient in a given community, but a non-local lingua franca, the means of communicating between people from anywhere in the world. Neither is its global weight restricted to elite usages in politics, international business or academia, but it is also employed by tourists, migrant workers, asylum seekers and just anyone in their daily lives over digital media. There is not even need to move around physically to be in contact with English.

<p style="text-align:right">(Mauranen 2018, 7)</p>

Furthermore, English as a lingua franca differs from other lingua francas "with one notable exception: the involvement of native speakers" (Dewey and Jenkins 2010, 72). Indeed, Mauranen (2018, 8) defines ELF as including the native speakers of English as well. The same view is adopted here. The author considers that in this global age the exclusion of natives, a large group of communicators, is not sensible or realistic. Since, it does not alter the role of ELF as a communicative tool, the inclusion of native speakers is reasonable. An interaction where English is spoken but it is not a shared first language between the speakers is therefore ELF usage, even if one or more of the speakers are native English users. However, this is not to be understood that all interactions where non-natives use English are ELF situations. For example, classroom English in Finland does not qualify as ELF, although it is English used by non-natives since, the purpose of the language in a classroom is not communication but learning. As a matter of fact, this communication aspect largely defines ELF.

For ELF scholars, a successful communication is the main interest, and all the varying ways of how the ELF speakers get there is what is under scrutiny. Jenkins (2006) aptly describes the aims of ELF and in addition, addresses variation in ELF, which will be further discussed in the following section:

> The existence of ELF is not intended to imply that learners should aim for an English that is identical in all respects. ELF researchers do not believe any such monolithic variety of English does or ever will exist. Rather, they believe that anyone participating in international communication needs to be familiar with, and have in their linguistic repertoire for use, as and when appropriate, certain forms (phonological, lexicogrammatical, etc.) that are widely used and widely intelligible across groups of English speakers from different first language backgrounds. That is why accommodation is so highly valued in ELF research. At the same time, ELF does not at all discourage speakers from learning and using their local variety in local

communicative contexts, regardless of whether this is an inner, outer, or expanding
circle English

(Jenkins 2006, 161).

In other words, ELF research attempts to discover, due to the wide variety of speakers, what kind of language is utilised in order for the speakers to achieve successful communication. Furthermore, Jenkins (ibid.) highlights the "linguistic repertoire" and "accommodation" of ELF users. This is a significant feature of ELF and deals with the inclusion of the user's background knowledge of any possible other languages and cultures. Linguistic repertoire is understood as covering the speaker's entire language capacity, including all languages, codes, dialects, and any other possible linguistic knowledge (Busch 2012). Accommodation, on the other hand, is a method in which a speaker attempts to, eponymously, accommodate their speech to match other speakers present in the interaction and to be more "intelligible" (Giles and Smith 1979, 54). As stated above, mutually understood communication is key in ELF, and sometimes that means resorting to other languages for a moment or extralinguistic means, like hand gestures or facial expressions. It could be stated that, translanguaging is present in ELF interactions by necessity since there is always more than one language closely present among the speakers. Translanguaging means that a speaker utilises all of their "multilingual repertoire in order to communicate successfully in intercultural interactions" (Wang 2018, 154). Indeed, especially in spoken ELF environments it is common to observe some translanguaging and code-switching (see for example, Cogo 2018, or Kimura and Canagarajah 2018). Kimura and Canagarajah (2018, 296, italics added for emphasis) further stress this link between ELF, communication, and translanguaging by remarking that "the type of competence translinguals have is a performative one, which *does not exist independently of communication*". Furthermore, Seidlhofer (2001, 143) describes that "ELF interactions often are consensus-oriented, cooperative and mutually supportive". Both quotations exemplify how ELF does not

9

concentrate on achieving grammatically correct English, but on reaching a mutually successful communicative situation.

2.2.2. Other Englishes and ELF

In addition to simply defining ELF on its own, the author feels it necessary to briefly address ELF in relation to other Englishes and what distinguishes them and justifies the separate titles. The distinctions are necessary to establish since the terminology can be confusing when this one language is spoken of in a variety of terms. There are for instance, the aforementioned World Englishes (WE), English as an international language (EIL), English as a second language (ESL), English as a foreign language (EFL), and of course English as a lingua franca (ELF), just to mention a few. Furthermore, as McKay (2018, 9) notes the terms World Englishes, ELF, and EIL are even sometimes used to mean the same thing. Indeed, this view can be seen in Jenkins' (2006, 160) article, in which she states that "EIL, itself an alternative term for ELF". Since Jenkins' 2006 view, researchers have come to distinguish the two acronyms (EIL and ELF) from each other, although, there are overlapping characteristics. Therefore, as this thesis is specifically interested in ELF, it is perhaps wise to explain what constitutes ELF in relation to these other branches of English and their research, although clear cut differences between some of the terms can be challenging to establish. Not all branches of English are going to be assessed here, but those that the author estimates are most associated and confused with ELF.

Firstly, while the World Englishes term can be understood as a kind of umbrella term for different Englishes, it also has its own definition and research aims. Most definitions of World Englishes lean on Kachru's (1985) Three Circles Model of Inner, Outer and Expanding varieties of English, and the study of WE concentrates on specific linguistic features in these varieties

(see Kachru 1985 for the Three Circles Model) (McKay 2018, 10). EIL, on the other hand, does comprehend English as language used by non-natives in international situations, as ELF often also is, but "EIL differs from both World Englishes and English as a lingua franca in insisting that the use of English for international communication must be based on a set of specific principles" (McKay 2018, 11). For example, emphasising the absence of culture in the English used and targeting pedagogical requirements according to the local situation are two of these listed principles according to McKay (for a more comprehensive definition on EIL and the principles list, see i.e. McKay 2018) (McKay 2018, 11). On the other hand, in contrast, culture could be considered important for ELF communication, since it is often relied on (for example, in instances of translanguaging) and used as help and reference. Actually, Baker (2015, 3) remarks that "it is (…) naïve to assume that ELF is a culturally and identity neutral form of communication".

Lastly, very briefly clarifications for the terms *English as a second language* (ESL) and *English as a foreign language* (EFL), as both of these overlap with the concept of ELF. Although ELF users often are second language speakers of English, "the term ESL is typically used to refer to English as studied, esp. by immigrants, in a country where it is the dominant language" (OED Online, 2020), while "EFL is typically used to refer to English as studied in a country where it is not the dominant language" (OED Online, 2020). Thus, even though there are shared characteristics among these Englishes, there is justification and place for each separate term.

To summarise, ELF is defined as interactions among users of varying backgrounds, be that native or non-native, for whom English is the only common language. Despite it having overlapping characteristics with other Englishes like EIL, what compresses ELF in particular is the communicative perspective into language. What separates ELF from the other Englishes

in this respect is that it is not a unified code, so researchers are interested how successful communication is achieved and what kind of language was used to get there. Considering this, ELF study does not limit itself to English but acknowledges that "ELF is situated within multilingualism, rather than replacing it" (Kimura and Canagarajah 2018, 305). The next section discusses in more detail the type of language found in ELF communication.

2.3. Characteristics of ELF

ELF's systematic and increased use has sparked a branch of research in English linguistics into it. This research on ELF has attracted increasing interest from the 1990's. For example, Jennifer Jenkins, Anna Mauranen and Barbara Seidlhofer are some of the most distinguished researchers on the topic. The research area in ELF is diverse covering grammar, business ELF (BELF), pronunciation, and humour to mention just a few. In effect, ELF can be examined like any other language. In this present section, the most substantial characteristics of ELF in relation to this thesis are briefly introduced and described. For example, grammar, morphological features, and complexity are given attention. Firstly however, the question of variety and variation in ELF is addressed.

2.3.1. Variation and the concept of variety in ELF

Seidlhofer (2018, 85), referencing Larsen-Freeman (2018), reminds that "it is a sociolinguistic commonplace that all natural languages are variable, continually in flux, complex and endlessly emergent", and this is true for ELF as well. Widdowson (2018, 101) expresses this: "the extension into lingua franca use quite naturally involves variation and change as the language is adapted to meet the needs of different communicative contexts and purposes". Thus, variation

in ELF is a major feature that constitutes to the kind of language the users produce. However, addressing ELF as a distinct variety is a linguistically loaded question, which is why it is expanded on here.

The knowledge of other languages factors into ELF variation significantly (Larsen-Freeman 2018, 53-54). This is because proficiency in other languages primes the learning and acquisition of other languages (Jessner et al. 2016, 166). Therefore, for ELF users this means that people have highly varying language inventory, every single user having their own unique disposition and arrangement of language knowledge.

Due to these diverse speaker backgrounds ELF expression is innovative. The language should not be expected to be systematic in its composition like standard varieties. However, as Björkman (2018, 260) notes, spoken ELF has morphosyntactic similarities with other World Englishes that rule out the possibility of randomly occurring features. This means that like World Englishes around the globe, ELF shares some of the recurring regular features, like double comparatives (e.g. *more nicer*) and superlatives (e.g. *most tallest)*, and zero marking in the plural (e.g. *two dog*) (Björkman 2013, 148-149). These features are discussed in more depth below (section 2.3.3).

Furthermore, in Laitinen (2018, 122), he concludes that when typological parameters are used, spoken and written ELF are different from learner English varieties and from British and American standards. Despite these somewhat systematic features and distinct language use, in the current view ELF scholars tend to not to consider ELF its own variety "because it simply lacks stability and a stable speech community" (Björkman 2018, 261). Björkman (2018, 263) further observes that the entire term *variety* is challenged, when discussing modern Englishes,

since the speech communities are "unstable". The instability in these circumstances means that ELF interactions can often be short, fleeing moments of English use, like service encounters abroad, after which the speakers might never meet again. The traditional notion of variety is discarded from the definition of ELF in this paper as well, as ELF speakers are diverse and many interactions sporadic, so although regular features have started to appear it cannot with good conscience be considered a separate variety. What is more, in ELF utterances the speakers' first languages can influence the language used (e.g. Larsen-Freeman 2018; Ellis 2008) and since the first languages change from interaction to interaction it adds to the unstable nature of ELF.

However, if ELF cannot be accepted as a variety, what is it then? In its somewhat fixedness, ELF is not adequately comprehended or described as merely a lingua franca. This has been argued by several researchers, like Mauranen (2018) who adopted the term *lect* and Mollin (2006) who suggested *register* as applicable for ELF. These changing definitions can be confusing. Therefore, perhaps the definition of *variety* should be expanded and modified, instead of disqualifying parts of language use from its range, like ELF. The author does not attempt to give a new definition here, for it is not the aim of this study, however, when necessary ELF is drawn parallel to a variety in comparisons with a Standard British English, due to this lack of suitable terminology. After all, terminology should help us determine concepts and if *variety* does not serve as needed anymore it could be redefined. As already quoted above, Björkman (2018, 263) considers the term *variety* challenged in the current global language world even when it comprehends the traditional definition. Therefore, even though the term is adopted here, it should be remembered that it is deployed in a wider sense than its usual acknowledged definition.

2.3.2. Complexity, grammar and morphological features in ELF

This section attempts to provide insights onto the structural and grammatical properties of ELF, and how ELF complexity, grammar, and morphology are pertinently connected. First, one of the interesting structural observations discovered about ELF concerns complexity. Although, L2 English and ELF are two distinct Englishes, there are similarities which the author deems permit comparison in this instance due to lack of sources on ELF. Kortmann and Szmrecsanyi (2011, 282) propose that L2 English speakers tend to favour zero marking over explicitness, meaning that, so to say, unnecessary grammatical markers are omitted. Kortmann and Szmrecsanyi (ibid.) are cautious in describing this phenomenon as contributing towards grammatically "easier" English for L2 speakers. More recently Mauranen (2018, 14) has pointed out that ELF "speakers tend to prefer structures that are easier to produce, and to avoid those that are hard to understand". However, both findings, by Kortmann and Szmrecsanyi (2011), and Mauranen (2018), highlight that non-native English speakers aim for effortless understandability. Therefore, it could be stated that it is apparent that at least semantically ELF and L2 users intend to produce less-complex utterances even if complexity of grammar is a multifaceted and difficult topic. What is more, according to the *accommodation theory,* speakers accommodate their speech to ensure mutual comprehension and this can happen for example, by elaborating descriptions and using more basic grammar formations (Giles and Smith 1979, 54). In addition, Seidlhofer (2001, 143) draws attention to ELF speakers' habit of "Let-it-Pass", meaning that utterances perceived difficult and incomprehensible are not dwelled upon and conversation continues as normal. This method might at first hand seem unproductive considering the objective of communication, but it has more of a social significance. Not acknowledging unsuccessful communication saves face, avoiding potentially mutual embarrassment.

The ELF speaker's emphasis is in getting the communicated message through, thus sacrificing language economy with explicitness. This suits ELF users, like the previous sections of this thesis demonstrate, the importance of successful communication is principal. Furthermore, the speakers in ELF interactions are often on different levels of language proficiency, which might affect the language used resulting in accommodation to ensure mutual understanding. In support of this view, Mauranen (2018, 13) refers to several previous studies (Seidlhofer 2004; Cogo and Dewey 2006; Mauranen 2007b, 2012) that have found that "tendencies of enhanced explicitness have been observed in ELF". Furthermore, what is interesting with ELF, is that this accommodation seems to be systematically developing into grammatical structures. For example, not marking the plural *-s* to nouns when there is a numeral determiner as in e.g. *two system*, or placing *not* right before what is being negated e.g. *This looks not sophisticated* (Björkman 2018, 257). Furthermore, as Björkman (ibid.) observes, these constructions lead to increased comprehensibility successfully.

Ranta (2018) explores ELF grammar within the range of research actually conducted in the area. She denotes that many of the ELF grammar studies available are merely observed feature lists and not so much an in-depth syntactic analyses (ibid., 248). Fortunately, these morphological feature lists are relevant and interesting considering the subject of this thesis, so I will not attempt to define ELF syntax profoundly either. To return to ELF grammar, Ranta (2018, 246) further reminds that ELF is "not a fixed code", so something observed in one ELF data might not be replicated in another. However, features that could be expected and in fact have been detected in ELF repeatedly are simplification (e.g. replacing bound morphology with analyticity) and leveling (unifying grammars in contact situations), emergence of language universals (using universal features of languages), and approximation (phrases that preserve the idea of a native utterance but uses divergent phrasing) (Ranta 2018, 247-248). According to

Ranta (ibid.), these characteristics are mainly due to ELF users' English as a second language (ESL) backgrounds and the nature of ELF being a contact language. Furthermore, all of these features quite distinctly also contribute towards ease of understanding, i.e. less complex structures. Let us, nevertheless, scrutinise more closely the morphological features already recognised in ELF.

With regard to explicitness mentioned above, Mauranen (2018, 13) notes that "discourse adaptations of this kind [enhanced explicitness] can also become drivers of grammar". This means that the explicit, more simple, and transparent forms could develop into stable structures in ELF grammar. Progression like this has been already observed. Björkman (2018, 257) claims that "these features could not have been caused simply by first-language (L1) transfer", referring to a list of features extracted from ELF speech situations. Some features on the list are for example:

(1)   Non-standard article usage, e.g. *Anyone can define the renewability?*

(2)   Double comparative/superlatives, e.g. *much more higher*

(3)   Not marking the plural on the noun in the presence of a quantifier, e.g. *two system, five reactor*

(Björkman 2018, 257)

These features induce language transparency, enhanced explicitness and they reduce redundancy (Björkman 2018, 257). Furthermore, according to her observations the speakers participating in her study were relatively proficient so that non-standard features were significantly less frequent compared to standard forms (Björkman 2018, 257). This indicates that speakers of ELF genuinely appear to have adopted some of these non-standard features.

What is more, these features and the simplification means that it directly affects the morphological, and therefore also the typological, composition of ELF. For instance, not marking the plural -*s* reduces syntheticity in ELF. This typology is expanded later below. Next, spoken language and ELF are briefly addressed.

2.4. Spoken ELF

The last ELF area augmented here concerns its spoken characteristics. It is imperative to draw attention to this distinction of written and spoken ELF, as it influences the language under scrutiny. Overall, spoken language represents more genuine language use than written language (Ranta 2018, 248). In reference to grammatical descriptions of languages by the now classical linguists Chomsky and de Saussure, Cienki (2017, 3) reminds that "they were describing our knowledge of language, not actual usage". Written language tends to be processed and modified several times before the final output, like for instance, in academic essays, but speech cannot be revised afterwards. Therefore, spoken interactions better illustrate how and what kind of language is used.

What is more, it has been observed that with ELF, its spoken and written forms differ significantly. Laitinen (2018, 122-124) detected that by typology, written ELF can be placed respectively among written British, American, and nativized varieties. However, spoken ELF situates quite independently from any other Englishes included in the study (ibid.). It could be suggested that the mentioned unrevised nature of spoken ELF is visible here. Björkman (2018, 257) and Ranta (2018, 250) both mention that non-standard language use is not excessively common in their ELF data, and that standard forms dominate. What could be deduced from this, is that ELF users are, in general, proficient which is reflected in the highly edited written

language. However, in spoken ELF the users have to think and speak on the spot. For many, English in ELF interactions is a second or subsequent language so, "as the less deeply entrenched language forms require more effort in retrieving and processing in L2 [second language] speech, L2 users may start resorting to processing shortcuts and thus approximating the ENL (English as a Native Language) forms" (Ranta 2018, 248). Thus, there is significant reasoning for differentiating ELF to its spoken and written types. Furthermore, to remind of the research aims of this thesis, this authenticity of spoken ELF is essential. It is interesting to discover whether differences between the spoken domains of ELF can be recorded when previous research suggests that the mode of instantaneous speech might affect ELF speaker output. The next chapter will see us turning to typology and analyticity and syntheticity.

3.   TYPOLOGICAL VARIATION

3.1. Introduction to typology

This section discusses linguistic typology (or language typology), in the scope of this thesis. The research questions presented in the introduction heavily lean on typology therefore, also this area of linguistics is essential to grasp fully. Firstly, a definition for typology and associated terminology is provided after which, the topic is examined from the perspective of the English language. Lastly, complexity in typology is introduced and through existing studies, what it might impose on spoken ELF is evaluated.

Typology as a term indicates assessment of things, in this case languages, into types. Indeed, a definition by Daniel (2011, 2) perceives that:

> linguistic typology compares languages to learn how different languages are, to see how far these differences may go, and to find out what generalizations can be made regarding cross-linguistic variation. As languages vary at all levels, linguistic typology deals with all levels of language structure, including phonology, morphology, syntax, and semantics.
>
> (Daniel 2011, 2)

As the morphological view is the primary interest of this thesis, unfortunately the other aspects of typology are not investigated in depth here (for a more comprehensive look on language typology see e.g. Song 2011, *The Oxford Handbook of Linguistic Typology*). The comparison of languages is in one respect actualised in classifications. Generally, linguistic typology acknowledges three morphological types of languages; isolating (minimal morphology), agglutinating (words can consist of more than one morpheme), and fusional (boundaries between morphemes are undistinguishable) (Caffarel et al. 2004, 54). However, determining classes to languages is not straightforward due to the above-mentioned linguistic variation; languages apply several coding methods concurrently and therefore, instead of describing the

entire language, a more suitable approach in linguistic typology is "classifying individual morphological processes" (Velupillai 2012, 96). Once again, I will not go into all these known existing morphological processes but concentrate on what is integral to this study. Few of these fundamental concepts are analyticity, syntheticity, and grammaticity. These terms are used to describe how language users code grammatical information. Analyticity can be defined as a set of "coding strategies where grammatical information is conveyed by free grammatical markers, which we in turn define in a fairly standard way as closed-class word tokens that have no independent lexical meaning" (Kortmann and Szmrecsanyi 2011, 280). For instance, analytic markers include forms in English like the determiner *the* and the infinitive marker *to* (as in *to speak*). Syntheticity, on the other hand, is a set of "coding strategies where grammatical information is conveyed by bound grammatical markers" (Kortmann and Szmrecsanyi 2011, 280). Instances like the Anglo-Saxon genitive *'s* (*Matti's book*) and inflected verb forms (*walked*) exemplify synthetic coding. Grammaticity in turn, is generally understood as comprising of both free and bound grammatical markers in a text (Szmrecsanyi 2009, 322). It is necessary to note that these definitions and this study understand analyticity and syntheticity as coding *grammatical* information, and that *lexical* analyticity and syntheticity are not taken into account (e.g. constructions like colour-*less* portray lexical syntheticity).

Whether a language uses more analytic or synthetic markers significantly influences the type classification. For instance, isolating languages, like Vietnamese, allow minimal morphology so the analyticity frequency in these types of languages is high, as information is in free standing, separate words (Caffarel et al. 2004, 55). Most of the world's languages utilise both synthetic and analytic means to code grammatical information (ibid.) so it is not an either or situation where for example, it could be stated that since Finnish is highly synthetic there is no analyticity at all. This analyticity-syntheticity spectrum is merely one method of study in

linguistic typology but nevertheless, it offers a reasonably good way of comparing languages from the same family or different varieties of one language. As an example, study by Siegel (2012) studied and compared creoles through analyticity in order to determine possible simplicity of grammar in them. This complexity and simplicity aspect is lastly expanded here before English typology is examined.

Morphological complexity is one area of study enabled by analyticity and syntheticity. The whole concept of language complexity is a rather loaded question politically and philosophically (Szmrecsanyi and Kortmann 2012) however, here it is understood through successful communication and what has been observed in studies. Szmrecsanyi (2009) provides a useful tripartite distinction of analyticity, syntheticity, and grammaticity in relation to language complexity:

1. Increased analyticity increases explicitness and transparency and decreases hearer/reader comprehension complexity.
2. Increased syntheticity increases speaker/writer output economy vis-à-vis analytic marking, by virtue of being the more compact coding option.
3. Increased grammaticity (i) increases redundancy, thus (ii) decreasing overall speaker/writer output economy, because more grammatical information is subject to overt coding. Redundancies such as these, however, (iii) reduce hearer/reader pragmatic inference complexity.

(Szmrecsanyi 2009, 323)

To express in other words, higher analyticity increases simplicity while syntheticity tends to implicate complexity and impair understanding. This subject of complexity is continued below when the typology of ELF is discussed. Now, let us consider the English language and typology first more generally.

3.2. Quantitative English typology

Current 21st century English is considered to be a more analytic language rather than a synthetic one (Haselow 2011, 28; Szmrecsanyi 2012, 657). A diachronic look on English reveals that it evolved from greater syntheticity to the current more pronounced analyticity (Haselow 2011, 28). However, a shift back towards increasing syntheticity is visible: modern English in general still shows more analyticity, but synthetic markers are on the increase, especially in written English (Szmrecsanyi 2012, 657; Szmrecsanyi and Kortmann 2011, 168). Since English is highly diverse among even its own varieties, also typology and analyticity and syntheticity levels, differ according to the variety, and as this chapter will illustrate, according to genre.

Benedikt Szmrecsanyi (2009) has developed a quantitative corpus-based method for typological profiling of languages (inspired by Greenberg 1960). The method enables the assessment of grammaticity of a language by calculating frequencies of analytic and synthetic markers. This method has been adopted by other researchers and will be modified for this thesis as well. It is useful that the method has become widely used as it allows comparisons of results between studies. Szmrecsanyi's method and what has been discovered about English thus far using said method is introduced here. The method description here is referenced based on Szmrecsanyi's 2009 study.

Table 1. *Analytic and synthetic grammatical marker categories.*

	Analytic markers	Synthetic markers
(1)	Conjunctions, subjunctions, prepositions	*s*-genitive

(2)	Determiners, articles, *wh*-words	Comparative and superlative adjectives
(3)	Existential *there*	Plural nouns
(4)	Pronouns	Plural reflexive pronouns
(5)	*More, most*	Inflected verbs
(6)	Infinitive marker *to*	
(7)	Modals	
(8)	Negator *not, n't*	
(9)	Auxiliary *be*	
(10)	Auxiliary *do*	
(11)	Auxiliary *have*	

Central for this typological profiling method are the analytic and synthetic markers defined above in Table 1. Szmrecsanyi (ibid., 326 – 327) defines 11 categories for the analytic markers and the synthetic markers are categorised in five, respectively. Based on these marker categories, the grammaticity of a language can be determined by counting instances of these markers. Furthermore, these markers can be matched to parts-of-speech (henceforth POS) tags and Szmrecsanyi's method takes advantage of it. POS tags inform the function a word has in a sentence, whether it is a plural noun or a first-person present verb, for instance. These tags are relevant as for example, when searching for the infinitive marker *to*, the separate tag TO0 separates the *to* from a preposition. If the data is searched with only the word *to*, false positives, instances that are recorded although they are not legitimate results, will arise as the search will provide all instances of the word *to*. A conversion of the grammaticity marker categories into

POS tags is portrayed in the methodological section below (tables 5 & 6). Therefore, if the corpus of the language, variety, or genre studied is POS annotated, it can be searched for the analytic and synthetic marker tags. Counting the frequencies of the analytic markers together reveals the analyticity level. Similarly, the syntheticity level can be uncovered by counting together all the synthetic markers. As grammaticity comprises of analyticity and syntheticity, the two categories combined offer the overall grammaticity level. Szmrecsanyi normalises all frequencies per 1000 words and this enables comparable values between corpora of different sizes. These resulting values Szmrecsanyi addresses as analyticity index (AI), syntheticity index (SI) and grammaticity index (GI).

By utilising Szmrecsanyi's method, several studies have already been conducted that have discovered differences in the typologies of Englishes (e.g. Szmrecsanyi himself 2009, 2012; Laitinen 2018; Szmrecsanyi and Kortmann 2011). Frequencies of these analytic and synthetic markers vary in different Englishes and even according to genre (See e.g. Szmrecsanyi 2009, and Kortmann and Szmrecsanyi 2011). Majority of native varieties of English, whether spoken or written, tend to exhibit higher grammaticity compared to non-native varieties (e.g. Laitinen 2018). This means that in native Englishes there are more analytic and synthetic markers than in, for example, learner Englishes. There are notable differences among the standard varieties as well (Szmrecsanyi 2009, 328). For instance, according to Szmrecsanyi (ibid.) spoken New Zealand English and spoken Standard British can be typologically differentiated as grammaticity in New Zealand English is lower. It is relevant to note that distinctions can also be detected between spoken and written English even inside a single variety (Szmrecsanyi 2009). In fact, in the already cited study by Szmrecsanyi (2009, 328–337) spoken texts in the British National Corpus (BNC) are "significantly more analytic than written texts", "written texts are significantly more synthetic than spoken", and overall "spoken texts exhibit

significantly more grammaticity than written texts". The BNC contains circa 100 million words from British English texts in different genres, it can be accessed on the internet for free (corpus.byu.edu). These are interesting observations to consider and remember below when ELF and typology is assessed.

The diversity of English typology is not limited to native versus non-native and spoken versus written dichotomies. Szmrecsanyi's (2009) study discovered frequency differences between genres of English as well. He mapped the different genres, including both spoken and written modes, according to their analyticity and syntheticity. Summarising his results, whether the register was spoken or written, the more formal registers, like essays or broadcasts, were found more synthetic compared to the less formal ones (ibid., 334). Additionally, spoken registers are systematically more analytic compared to the written registers (ibid.). These Szmrecsanyi's genre specific results are returned to in more detail later on. Next, what is currently known about ELF typology is presented.

3.3. ELF and typology

When attempting to profile ELF typologically, the non-nativeness is a significant feature. Since ELF is not a stable established variety, its typological profiling might seem futile however, it provides interesting insights into how ELF users utilise the language. Not many studies have been conducted in this area, mainly Laitinen in 2018. Before examining his results, relevant findings by Szmrecsanyi and Kortmann (2011) portray what they discovered when analyticity and syntheticity levels in learner Englishes and indigenised Englishes were compared.

Szmrecsanyi and Kortmann (2011) aimed at figuring out whether *English as a Foreign Language* (EFL) and *English as a Second Language* (ESL) can be distinguished in typology. Their results indicate that indeed the EFL frequencies of analyticity are higher and syntheticity frequencies lower than in the L2 Englishes (ESL) (ibid., 175). Furthermore, majority of the EFL varieties included in the study situated in the proximity of conversational Standard British English. This suggests that EFL might resemble speech like language. (ibid., 173) Of course, as previous sections in this paper display the definitions of EFL and ELF are not equivalent, but these results give information on non-native English typology and what kind of behaviour could be expected about ELF. Various sources report similar findings, how non-native and expanding circle Englishes comprise of higher degrees of analyticity, lower degrees of syntheticity and overall lower grammaticity (Szmrecsanyi 2009; Kortmann and Szmrecsanyi 2011; Siegel 2012). What could be deduced from these results is that the speech like language of EFL also signifies lower complexity and that in turn again can be associated with enhanced explicitness. Therefore, the typology of EFL appears to support the view that non-native Englishes strive for successful communication.

However, more recent findings by Laitinen (2018) challenge the hypotheses about ELF resembling other non-native Englishes depicted above. Laitinen discovered that written ELF does not significantly differ from a written native variety in terms of grammaticity (ibid., 124). One explanation for this native-like writing in ELF could be the overall proficiency of ELF users, as Björkman (2018, 257) in her study suggests. Written text can be modified unlike speech, so any non-standard features can be edited to resemble a Standard. However, spoken ELF compared to other varieties, written or spoken, quite distinctly differentiates according to grammaticity; Laitinen (2018, 122–124) found that the levels of analyticity and syntheticity are

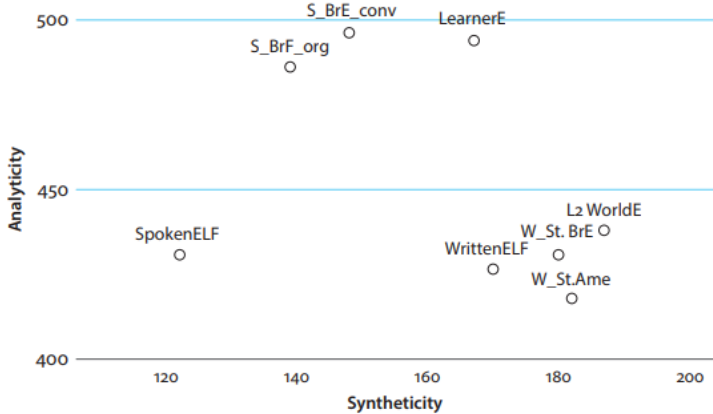both remarkably lower than in any of the compared varieties. Below Figure 2 demonstrates his findings (ibid.).

Figure 2. Written and spoken ELF compared with select variety types in Szmrecsanyi (2009) and Szmrecsanyi & Kortmann (2011)

Figure 2. *ELF compared to native varieties by Laitinen (2018, 122).*

The typological composition of spoken ELF appears to vary significantly from the other varieties of English included in the analysis. Based on previous results of varieties of Englishes, it would appear justifiable to assume that ELF, being a non-native variety, would exhibit lower levels of syntheticity, higher levels of analyticity, and altogether less grammatical markers than a native variety like for example, a Standard British English. Indeed, it does exhibit lower analyticity and syntheticity than native varieties however, ELF appears to portray even lower levels of grammaticity than learner Englishes. This is why the dissimilarity between Learner English and ELF in Laitinen's results is notable; spoken ELF stands quite independently from any other variety, even from the variety expected most to resemble it.

Since spoken ELF is a non-native English and low in grammaticity, what does it impose on ELF complexity? The low syntheticity indicates less complex languages but unlike other

spoken varieties of English, in ELF also analyticity is low. In fact, Laitinen (2018, 122) observed that the written and spoken ELF do not differ statistically in analyticity. As established already above, higher analyticity is expected in spoken language as it increases hearer comprehension (Szmrecsanyi 2009, 323). However, spoken ELF does not appear to follow this convention. Thus, it can be deduced that ELF users preserve the output economy but in order to facilitate their speech syntheticity is compromised as a result. Björkman (2018, 260), referring to previous studies, suggests that actually in non-native Englishes the decreased redundancy promotes comprehension. So, ELF could well be an extreme example of this phenomenon.

This background knowledge on ELF and typology provides the foundation for the research at hand. Before turning to methodology and the actual study, a description of few previous studies more exhaustively. Firstly, Szmrecsanyi (2012) established typology in the history of English. Szmrecsanyi's 2009 study provides the methodological procedure example, as well as a native variety comparison point with the spoken genres he analysed typologically. Finally, Laitinen's (2018) research is the other methodological model and it provides background knowledge on typology of ELF as demonstrated already shortly above.

4. PREVIOUS STUDIES

This section introduces previous research in typology of English and ELF. Primarily, I will describe in more depth the aforementioned study by Szmrecanyi (2009). All of the studies accounted here rely on Szmrecsanyi's methodology. For instance, the study typologically profiling ELF by Laitinen (2018) and another study by Szmrecsanyi (2012) both utilise the grammatical marker methodology. First, however, a historical perspective on how English typology has evolved.

4.1. Szmrecsanyi 2012

Using his typological profiling method, Szmrecsanyi (2012) established the development of grammaticity in the history of the English language. The pervasive notion in English typology has been that Old English was highly synthetic and that the language since then has continually evolved towards increasing analyticity (ibid., 655–656). Szmrecsanyi remarks that these typological generalisations have been drawn on the basis of studies conducted on few isolate features (ibid.). In addition, research has already begun to undermine the accuracy of this ever increasing analyticity. Szmrecsanyi (2009) showed that at least present-day written British and American Englishes portray increasing syntheticity. Indeed, his method enabled a more reliable quantitative approach to this diachronic investigation.

The corpora Szmrecsanyi used were the Penn Parsed Corpora of Historical English series, which entails three corpora dating from the 12th century to the early 20th century. The corpora are POS-tagged so Szmrecsanyi was able to make the queries for the analytic and synthetic word tokens using the tags. These frequencies were also normalised per one thousand words from which values he calculated the analyticity (AI) and syntheticity indices (SI). The results

Szmrecsanyi derived from the study show high diversity according to century in analyticity as well as in syntheticity. Firstly, the syntheticity in the 12th century started off as high and decreased until the 15th century, after which it increased again century by century so much that by the 20th century the syntheticity level is closest to the 12th century frequency (ibid., 658). Secondly, the analyticity levels, Szmrecsanyi (ibid.) found, spiked from the 13th century to the 14th century by nearly fifty index points. So, the assumption that English developed from prominent syntheticity to increasing analyticity does appear genuine between the 12th and 14th centuries. However, after the 14th century Szmrecsanyi (ibid.) records a somewhat regular decrease in analyticity. By the 20th century the continuous decrease in analyticity means that the frequency is lower than in the 12th century (ibid.). Szmrecsanyi (ibid., 659) suggests thus, that since the 17th century the trend has actually been increasing syntheticity and decreasing analyticity. This is an opposite finding in comparison to the previous conception of typology in the history of English.

Szmrecsanyi (ibid., 660) also discovered that the historical variance of English is significant in analyticity but not in syntheticity when the language is observed in a crosslinguistic manner. Furthermore, this examination shows that even in its most synthetic form in the 12th century, English was still less synthetic than Italian or German (ibid.). Szmrecsanyi (ibid., 661) reports some intertextual variability, especially in the Early Modern English, but attributes this to the diverse selection of texts in the corpora.

Finally, Szmrecsanyi (2012, 661 – 664) addresses grammatical markers and their effect on English analyticity and syntheticity throughout history. He uses one-way ANOVA to determine the most varied analytic features (ibid., 661 – 662). These are determiners, pronouns, infinitive markers, and auxiliary verbs (ibid.). Szmrecsanyi (ibid.) suggests that from these four features,

determiners and auxiliary verbs are increasing, while pronouns and infinitive markers have been decreasing since the 17th century. A similar investigation on the synthetic features discovered the following four markers: the possessive marker, inflected adjectives, plural nouns and nominals, and inflected verbs (ibid., 663). Szmrecsanyi (ibid., 663 – 664) reports that the possessive marker significantly decreased after the 12th century, after which the feature shows mainly moderate fluctuation. On the other hand, inflected adjectives are steadily increasing (ibid.). Inflected verbs have evolved in a U-shaped fashion, meaning that their frequency currently is close to their level in the 13th century (ibid.). Plural nouns and nominals, Szmrecsanyi (ibid.) describes, have increased steadily. Szmrecsanyi (2012, 664 – 665) concludes the study by reviewing these most significant results. In addition, he remarks that similar studies should be conducted on Old English and in other languages so that crosslinguistic research in diachronic typology would become possible (ibid.).

4.2. Szmrecsanyi 2009

In 2009, Szmrecsanyi set out to discover variability in English. The methodology of his study was already described above (section 3.2.) in this study. Szmrecsanyi approaches this mission through morphological typology, and included in the analysis geographical, text type, and real time dimensions (2009, 319). Morphological typology is realised as grammaticity, and analytic and synthetic markers as the means to study it (ibid.). The research paper accounts for several native Standards like British, Irish, and American, indigenised varieties like Indian and Jamaican Englishes for World Englishes comparison, and what is more, Szmrecsanyi even evaluates written and spoken language differences in British English (ibid.). The data for this corpus based study is from five existing corpora; *The British National Corpus* (BNC) representing British variety, as well as providing the text type and spoken-written mode

comparisons, *The Brown family of corpora* (Brown, LOB, Frown, F-LOB) including American and British text corpora from the 1960's and 1990's thus enabling diachronic analysis, *Switchboard* characterising spoken American, *The Freiburg Corpus of English Dialects* (FRED) for differences inside the UK due to geography, and *The International Corpus of English* (ICE) from which the spoken parts of certain sub-corpora are included for the World Englishes perspective (Szmrecsanyi 2009, 323-324). From these corpora, results concerning the text type analysis of the BNC are the most intriguing and will also be referenced to later in the results of this paper. Furthermore, as mentioned earlier the methodology of Szmrecsanyi's study is of relevance as it functions as a base for the method in this research as well.

Szmrecsanyi utilises the part-of-speech (POS) tagging in the corpora to search for the analytic and synthetic markers that he has assessed representative of morphological typology. These markers he describes as "component categories" (2009, 325). The five corpora chosen use different tagging systems so altogether four sets of POS-tags are included: *CLAWS5e*, *CLAWS5*, *CLAWS8*, and *Hepple tags* (ibid., 326-327). Simply counting the frequencies of these marker tags reveals the grammaticity level of each variety analysed.

The results Szmrecsanyi (2009) presents in the three dimensions "geographic variability in World Englishes", "text type variability", and "short-term diachronic variability". Firstly, the World Englishes analysis reveals that there is significant typological diversity among all the different geographical locations. The other significant notion to be derived from these results is that low-contact varieties (i.e. rural dialects in the UK) are more synthetic and thus more economic than high-contact varieties that appear to emphasise transparency and which in consequence are higher in analyticity (ibid., 331). Furthermore, native L1 varieties portray overall more grammaticity than the indigenised L2 varieties (ibid.). The text type results further

reveal that the variability concerns even individual varieties (ibid., 339). According to the spoken or written mode the typological composition can vary significantly, primarily in the fashion that spoken language is more analytic and overall more grammatic than written texts (ibid., 340). As this part of the results is imperial for the current paper, differences Szmrecsanyi discovered in the spoken text types are extended beyond the focal findings about the spoken-written dichotomy. In Szmrecsanyi's (332–334) observations the spoken types do not drastically vary in analyticity but, the syntheticity frequencies set the types apart. The least synthetic one is 'pub-debate' and the most synthetic 'sermon', which also happens to be the most analytic speech event type (ibid.). However, as Szmrecsanyi points out sermons are 'scripted speech', since many beforehand write it down, explaining the higher syntheticity, he further notes that sermons require explicitness and clarity hence the high analyticity (ibid.). The implications to be driven from these results are that it could be expected other varieties behave similarly; the more formal speech events show more syntheticity than the informal ones, which in turn could be anticipated to represent higher analyticity in contrast.

Lastly, the results on the short-term diachronic analysis uncover that at least the two Standards, American and British English, have developed towards higher syntheticity (Szmrecsanyi 2009, 348). This terminates the trend and view of increasing analyticity in English that has persisted for a millennium (ibid.). Szmrecsanyi's study gives a rather comprehensive look into typological variation in Englishes. Furthermore, the genre specific results will be returned to later in the discussion of this thesis.

4.3. Laitinen 2018

Laitinen (2018) in his study aims at describing ELF typologically in order for it to be compared to native varieties. He utilises the methodology developed by Szmrecsanyi (2009) (above) so the results are largely comparable. Like Szmrecsanyi (ibid.), Laitinen uses the POS-tagging to determine the grammaticity markers, with only few adjustments to the tags. The modals *gonna, wanna, hafta* and *gotta* are included which differ from Szmrecsanyi's categorisation (Laitinen 2018, 116). The ELF data comprises of the *Vienna-Oxford International Corpus of English* (VOICE) corpus for spoken ELF and the *Written English as a Lingua Franca* (WrELFA) for written mode (ibid., 114).

The main results Laitinen (2018) reports fall into three sections. First, overall grammaticity frequencies of ELF (excluding tweets) show that it exhibits low levels of grammaticity (ibid., 118). Comparing ELF to Szmrecsanyi's World Englishes frequencies it situates third lowest according to overall grammaticity, only Hong Kong English and Singapore English are showing lower levels (ibid.). According to Laitinen (ibid., 119) this supports previous interpretations that "ELF speakers often avoid overt grammatical marking". What is more, when ELF is broken down to its different genres, ELF tweets and spoken ELF are even lower in grammaticity than the Hong Kong and Singapore Englishes (ibid., 120). Turning to the second area of the results, it deals with the written-spoken dichotomy inside ELF. While written ELF is similar to other written varieties of English, spoken ELF stands quite independently on the analyticity-syntheticity diagram (ibid., 122). Laitinen discovered that spoken ELF is statistically significantly less synthetic than its written counterpart (ibid.). Furthermore, these results establish that ELF differs from Learner Englishes thus, giving support to the view that foreign language acquisition and second language use are not equivalent (ibid., 123).

The last results section discusses "genre differences in ELF" with the intention that it reveals possible systematicity in ELF (Laitinen 2018, 124). In essence, systematicity and resemblance of native genres would indicate ELF users' "awareness of genre characteristics in terms of structural features" (ibid.). Indeed, Laitinen discovered that at least the genres of spoken language, news, academic texts, and fiction show no statistically significant differences in their analytic-synthetic constructions when ELF and the BNC data is compared (ibid., 126). Which suggests that there is some systematicity in ELF. However, although both low in grammaticity spoken ELF and ELF tweets are not constructed similarly. Tweets are more synthetic while spoken language is more analytic (ibid., 125).

Laitinen (2018, 127) concludes that his study provides new and interesting insights as "on purely structural grounds ELF is another distinct variety type". His results are relevant considering the objects of this present study. Since, Laitinen established that typologically ELF can be distinguished from other Englishes, and ELF users appear to be aware of structural differences of different genres, it encourages this research with impetus to investigate whether the structural differences reach different domains of spoken ELF. The *VOICE* is divided into five domains which enables this research. The next section will introduce *VOICE* more profoundly and the methodology of this thesis study.

5. METHODS AND MATERIALS

Here, the corpus, POS-tag conversion, digital tools, statistical testing, and other methodological features utilised are established. The methodology of this study relies heavily on those utilised by Szmrecsanyi (2009) depicted above in detail. First, an insight onto the data of this study.

5.1. Materials

Grammatical features are something people do not usually register using while speaking, which is why a corpus-based methodology was chosen for this study. A corpus is a collection of texts from a particular language environment, like literary texts that comprise a text corpus or recorded and transcribed interviews for example, that can comprise a spoken language corpus (McEnery 2012, 1–2). A corpus of spoken language represents more accurately how features and language are used in reality. Considering this study examines spoken ELF, the corpus chosen is the VOICE (*The Vienna-Oxford International Corpus of English*). VOICE (2013) is a spoken corpus of ELF and includes texts transcribed from situations like interviews, seminar discussions and press conferences. It can be accessed and downloaded from the internet for free at www.univie.ac.at/voice/. The recordings were gathered from July 2001 to November 2007 and encompass altogether 1,023,082 words of transcribed text (ibid.). All the recordings for VOICE were un-scripted and the speech events happened face-to-face between speakers (ibid.). Barbara Seidlhofer acts as the project director for VOICE (ibid.).

The texts in VOICE are categorised in different domains according to the environments the texts were recorded in. The three main domains are ED (educational), LE (leisure), and P (professional), which is further divided into three; PB (professional business), PO (professional organizational), and PR (professional research and science) (VOICE 2013). These five domains

are the main interest of this study. The domains are somewhat unambiguous; the educational domain texts are gathered from educational settings, leisure domain texts from recreational discussion situations, and the three professional domain texts are from varying specialist or academic situations (ibid.) These domains are also divided into "Speech Event Types", separating for example, interviews, conversations, and meetings (VOICE 2013). This study concentrates on the domain specific results. However, it is also relevant to know the speech event types the domains consist of. Below in Table 2 the corpus is broken down to demonstrate the relative sizes of the domains.

Table 2. *The VOICE divided into the separate domains. (Table from univie.ac.at)*

Domain	Speech Events	Words	Relative in 1/100
ED (educational)	35	261003	25.51
LE (leisure)	26	101216	9.89
PB (professional business)	23	203421	19.88
PO (professional organizational)	41	354602	34.66
PR (professional research and science)	26	102945	10.06

An examination of Table 2 reveals that the domain PO is the most extensive, covering more than 30 per cent of the word count. Therefore, the figures derived from the analysis are normalised, calculated per one thousand words, to ensure comparability between the domains. The subcategorization according to the speech event types, which inform the type of situation the recording was derived from, is presented in Table 3 below. The reason, why the attention in this study is on the domain specific rather than speech event specific results, is that the domain could be considered affecting the language more strongly. A working group discussion can hold very different type of language whether it is held in an educational setting or in a business setting, despite the speech event being the same.

Table 3. *The VOICE speech event types. (Table from univie.ac.at)*

Spet	Speach Events	Words	Relative in 1/100
con (conversation)	36	158075	15.45
int (interview)	16	36362	3.55
mtg (meeting)	20	273458	26.73
pan (panel)	10	92719	9.06
prc (press conference)	5	17588	1.72
qas (question-answer session)	10	27541	2.69
sed (seminar discussion)	6	63625	6.22
sve (service encounter)	11	14894	1.46
wgd (working group discussion)	19	181055	17.7
wsd (workshop discussion)	18	157870	15.43

Nevertheless, the nature of these subcategories is relevant when the results are presented and discussed. Three speech event types cover over fifty per cent of the corpus. Working group discussions (wgd), meetings (mtg), and conversations (con) add up to 59.88 per cent (table 3). It is worth noting these three types dominate the data, especially when the results are discussed and compared with Szmrecsanyi's (2009, 333) native spoken language genres. The domains vary according to the speech event types, ED covering six while LE only consists of two speech event types (VOICE 2013). The largest speech event type of each domain will become relevant later, in the results section, so they are here introduced simply in text. The largest speech event type in ED domain is the 'workshop discussion', 'conversation' dominates the LE domain, 'meeting' is largest in PB, 'working group discussion' in PO, and 'panel' speech event type in PR (ibid.).

A brief introduction onto the BNC spoken texts Szmrecsanyi (2009) utilised is relevant as well. The BNC spoken genres are later compared to the VOICE domain results of this study. The spoken BNC is categorised into 24 different genres as the Table 4 by Kettemann and Marko (2002, 279) below demonstrates. An examination of the genre table reveals that the BNC genre

'lectures' consists of five separate lecture sub-types, for instance. This might prove essential during the comparison of VOICE and BNC grammaticity results, since based on these sub-types the 'lectures' genre can be assessed as representing academic speech. The BNC is not further expanded here since beyond these spoken genres only Szmrecsanyi's (2009) results are of importance.

Table 4. *The genres of spoken BNC*

BNC SPOKEN	Super Genre
S_brdcast_discussn	Broadcast
S_brdcast_documentary	
S_brdcast_news	
S_classroom	
S_consult	
S_conv	
S_courtroom	
S_demonstratn	
S_interview	Interviews
S_interview_oral_history	
S_lect_commerce	Lectures
S_lect_humanities_arts	
S_lect_nat_science	
S_lect_polit_law_edu	
S_lect_soc_science	
S_meeting	
S_parliament	
S_pub_debate	
S_sermon	
S_speech_scripted	Speeches
S_speech_unscripted	
S_sportslive	
S_tutorial	
S_unclassified	

(Table derived from Kettemann and Marko 2002, p.279.)

The BNC genres and the VOICE domains are not directly comparable as an examination of tables 2 and 4 shows. There are nevertheless similarities, so the results are observed, with some reservations, in comparison with Szmrecsanyi's (2009) BNC results. However, few of the BNC genres are entirely discarded since they are not applicable for a rational comparison. For example, the 'sermon' genre is such a specific type of spoken language, and there is no equivalent domain in VOICE, that including it in the comparison would not be sensible. The discarded genres are the aforementioned 'sermon', 'broadcast', 'courtroom', 'sportslive', 'parliament', and the 'unclassified'. The rest are examined in contrast to the five VOICE domains ED, PB, PR, PO, and LE.

5.2. Method

A factor that also favours the selection of VOICE for this research is that it is Parts-of-Speech (POS-) tagged. Thus, Szmrecsanyi's (2009) method, introduced previously, of determining grammaticity by counting markers, is transferrable to this study. This means that the VOICE is searched for the grammatical markers with their corresponding POS-tags. In other words, for example, all prepositions, one of the analytic marker categories, can be searched simultaneously with the POS-tag 'IN' (explanation below in Table 5) instead of searching for *under*, *from*, or *in* each separately, for instance. However, the 11 categories for analytic markers and five for synthetic markers that Szmrecsanyi (2009) defines need to be converted to match the VOICE POS-tags.

The VOICE uses its own tagging based on the Penn Treebank Project, which is a rather rough categorisation of word classes (VOICE 2014, 6). Szmrecsanyi (2009) used several different

tagsets in his study, since the several corpora he worked with utilised different tagging systems. However, only one tagset is needed to convert the marker tags applicable for VOICE, as only this one corpus is under scrutiny. The conversions are shown below in tables, the analytic markers and tags in Table 5 and the synthetic ones in Table 6. The conversion by Laitinen (2018, 117), who used the CLAWS7 as tagset reference, is used as example model since he has already converted the same categories from Szmrecsanyi (2009) to correspond to the VOICE tagging. The tagging on VOICE is not comprehensive or entirely reliable so manual checking will be done in addition to simple tag searches during the analysis. With the established correct POS-tags the VOICE is analysed for the grammatical markers and their frequencies recorded.

Table 5. *Analytic component categories and their grammatical tags*

Categories	CLAWS7	VOICE
1.Conjunctions, subjunctions, prepositions	CC*, CS*, I*	CC, IN
2. Determiners, articles, *wh-*words	APPGE, AT*, D*, RGQ*, RRQ*	WRB, WDT, WP, DT, PRE, PDT
3. Existential *there*	EX	EX
4. Pronouns	P*	PP
5. *more*, *most*	RGR, RGT	*more, most*
6. Infinitive marker *to*	TO	TO
7. Modals	VM*	MD, *gonna, wanna, gotta, hafta*
8. Negator *not*, *n't*	XX	No separate tag but both forms under lemma *not*
9. Auxiliary *be*	VBD*|VBG|VBM|VBN|VBR|VBZ* + (*)? + V*	VB* + Manual search
10. Auxiliary *do*	VD* + (*)? + V*, VD* + XX	Lemma *do* + Manual search
11. Auxiliary *have*	VH* + (*)? + V*, VD* + XX	VH* + Manual search

Table 6. *Synthetic component categories and their grammatical tags*

Categories	CLAWS7	VOICE
1. *s*-genitive	GE, MCGE	POS
2. Comparative/superlative adjectives and adverbs*	JJR, JJT	JJR, JJS, RBR*, RBS*
3. Plural nouns	NN2, NNL2, NNO2, NNT2, NNU2, NP2, NPD2, NPM2	NPS*, NNS
4. Plural reflexive pronouns	PPX2	*selves* Needs manual checking
5. Inflected verbs	VBDR, VBDZ, VBG, VBM, VBN, VBR, VBZ, VDD, VDG, VDN, VDZ, VHD, VHG, VHN, VHZ, VVD, VVG, VVGK, VVN, VVNK, VVZ	VVD, VBD, VHD, VVG, VBG, VHG, VVN, VBN, VHN, VVZ, VBZ, VHZ, VHS, DOS, VBS, VBP
6. *Gonna, wanna, hafta, gotta*		Tagged in VOICE following format 'gon_VVG na_TO' etc.

* - Marks the tags added by the author of this study, with the exception of *selves*.

Laitinen (2018) included also the informal forms of *gotta, hafta, wanna* and *gonna* in the analytic modal category, however, the author deems it appropriate that the forms load both categories, since these modals do apply synthetic constructions as well in merging the modal and the infinitive marker together. However, it is necessary to note that the frequencies of *gotta, hafta, wanna,* and *gonna* are so low that they do not affect the normalised frequencies in either marker category. Other deviations to the previous studies' categorisations include, the added synthetic tag NPS standing for plural proper nouns, and the RBR and RBS synthetic tags for comparative and superlative adverbs. The comparative and superlative adverbs portray similar constructions to the already included markers of comparative and superlative adjectives, which prompts inclusion for the adverbs as well. The frequencies of these markers are not high but deemed important to include, nevertheless. As observable in the tables above, the VOICE

tagging lacks certain categories altogether. For example, the auxiliary verbs are not tagged in VOICE. This means that manual searching and checking is required during the analysis for these markers. The auxiliaries *be* and *have* are searched with the respective tags VB* and VH*, and the auxiliary functions recorded. *Grammar of Spoken and Written English* (Biber et al., 2000) is consulted in order to ensure correct instances of auxiliary use with these verbs. Instances of *be* and *have* include formations where a second verb follows directly after the auxiliary, with one word between the auxiliary and verb, and negations of these constructions. *Do* has no individual tagging on VOICE like that of *be* and *have,* so it is searched for its lemma *do* and then for the auxiliary forms. Lemma is a base form of a word, and if a corpus is lemma tagged inflected word forms can be searched with the lemma for example, *destroying*, *destroyed¸* and *destroys* have all the same base, or lemma, *destroy* (McEnery 2012, 245). Auxiliary functions of *do* that are included in this analysis are negations (e.g. *doesn't smoke*), inversions (e.g. *does she like*), interrogatives (e.g. *do you like*), *Wh*-interrogatives (e.g. *what did you bring*), tags (e.g. *didn't she*), and emphasis (e.g. *you did promise*).

During the data gathering, using these converted tags, the five domains of VOICE are searched separately for the analytic and synthetic markers. The search tool used in this study is AntConc, and it was chosen since the author is familiar with its functions and since the tool is freely available on the internet for download. The results are primarily analysed quantitatively. The frequency of each tag is recorded on Excel, where the sum of all analytic and all synthetic markers is calculated separately within the domain in question. This means that the analysis returns analyticity and syntheticity values for each separate domain. This allows comparison between the domains. The actual frequency levels are calculated per 1,000 words to ensure comparability with Szmrecsanyi's (2009) results. This normalized frequency is commonly used when analyticity and syntheticity levels are discussed, thus presenting the findings in this form

allows comparability with previous and future research. The normalized frequency formula used is:

$$\frac{N}{Words\ in\ corpus} \times 1000$$

Where, N is the number of analytic or synthetic markers found in the texts within the domain in question. The received values form the analytic and synthetic indices (AI and SI).

Once the corpus analysis is complete, a log likelihood calculator provided by Paul Rayson from the University of Lancaster (available online at http://ucrel.lancs.ac.uk/llwizard.html) is utilised to calculate statistical significances between the five domains. Log likelihood compares two corpora and the frequencies of a word, or in this case the number of analytic or synthetic markers, in those corpora and reports whether the frequency difference is significant. Log likelihood is convenient because it allows comparisons between corpora that are of different sizes. For example, on VOICE the different domains vary considerably in word count. The log likelihood calculations are done on the raw data and not on the normalised frequencies. The log likelihood value received can be then compared to the critical values that inform relative p-values. For example, for the p-value to be on the significant level of 0.05 or less, the respective log likelihood value needs to be higher than, or at least, 3.84. Log likelihood values of 10.83 or more equate to a p-value of 0.001. Naturally, even higher log likelihood values mean higher significance. Log likelihood is also referred to below as an acronym LL. A p-value is a statistical probability that informs a null hypothesis to be true. The significance for a p-value is usually set at 0.05, or in other words, there is five per cent chance of the null hypothesis not being true (Mäkisalo 2009, 58). The smaller the p-value the higher significance it equates to.

Table 7. *Log-likelihood and p-value correspondence.*

P-values	Log-likelihood significance values
p< 0.05	3.84
p< 0.01	6.63
p< 0.001	10.83
p< 0.0001	15.13

Furthermore, the domain results are compared to Szmrecsanyi's (2009) BNC results of spoken genres of Standard British. An overall comparison of the domains and genres is conducted as well as, individual contrasts between domains and genres.

Next, in the results section, the frequencies from the analysis are compared to each other within the different domains, and possible observations reported in relation to the research questions. The results and frequencies are also compared to Szmrecsanyi's (2009) spoken genre results in order to contrast with a native variety. Since the genres in Szmrecsanyi's study and the domains in VOICE do not correspond perfectly the results are not directly comparable as indicated in this chapter. However, they both provide information about spoken texts in different speech environments, and a native comparison is interesting as it enables to set these ELF results into perspective.

6. RESULTS

6.1. Quantitative results

The results are presented in this section. However, first a recap on the research questions. Firstly, the focus is on the overall frequencies of the analytic and synthetic markers across the five domains. Furthermore, some frequency comparison between these domains is of interest as well as, a comparison of the ELF results with a standard British English and its various spoken genres. The results are presented below first in plain frequencies and then in relation to these research questions.

Table 8. *VOICE domain specific analyticity, syntheticity and grammaticity indices*

Domain	AI	SI	GI
Educational (ED)	431	120	551
Professional Business (PB)	410	122	532
Professional Research and Science (PR)	418	132	550
Professional Organisational (PO)	439	129	568
Leisure (LE)	409	130	539

Table 8 establishes the domain specific results. The initial examination of the results reveals that there is notable variation among the domains. The results portray that PO (professional organisational) is the most grammatic of the domains with a grammaticity index (GI) value of 568. PO is also the most analytic domain with analyticity index (AI) of 439. The syntheticity index (SI) for PO domain is 129. The least grammatic domain is PB (professional business) with a GI of 532, it comprises of AI of 410 and SI of 122. Rest of the domains fall between these two domains. Least analytic is the LE (leisure) domain with 409 analytic markers per thousand words. However, with a SI of 130 LE is the second most synthetic domain but still

penultimate in overall grammaticity with a GI of 539. PR (professional research and science), on the other hand, portrays highest syntheticity with an SI of 132. With an AI of 428 and therefore with a GI of 550, PR situates in the middle among the other domains in overall grammaticity. Finally, the educational domain ED is lowest in syntheticity with an SI of 120. AI is in contrast second highest with 431 markers, and overall GI is thus 551, which is also the second highest. A further observation on the GI reveals that analyticity appears to affect grammaticity more than syntheticity. The most synthetic domains are not reflected as the most grammatic, while this is true with analyticity. The two most analytic domains PO and ED are also the most grammatic of the domains. Figure 3 better visualises what all these differences mean. Also included in the scatterplot is the mean value of the domains.
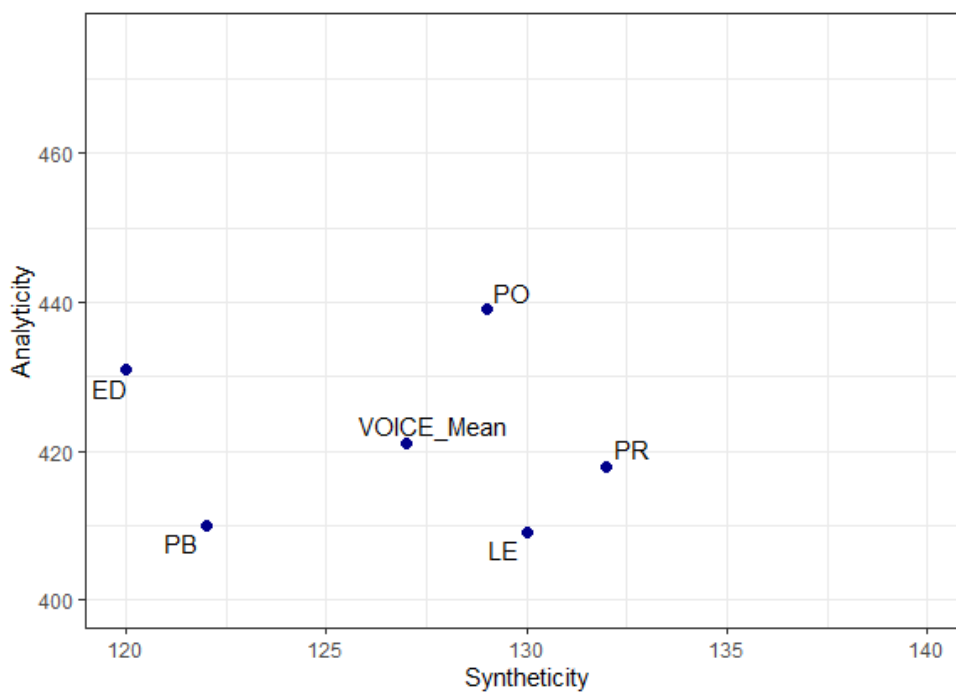

Figure 3. *Scatterplot of VOICE domain grammaticities and a mean value.*

The domain mean values are 421 analytic markers per thousand words and 127 synthetic ones. Hence, the GI for VOICE is 548. These normalised frequencies do not reveal much on their own. Therefore, statistical testing is in place to determine any significant results.

Log likelihoods in Table 9 tell whether the differences observed between the domains are noteworthy in a statistical respect, and fortunately there are significant p-values to report. The most significant difference in analyticity is between the domains PO and PB, with a LL value of 251.38. This indicates a very highly significant difference in the frequencies with a p-value smaller than 0.0001. Simply observing the normalised frequencies and Figure 3, the most noticeable difference in analyticity seems to be between PO and LE, which is why log likelihood is useful as it takes into account the relative sizes of the corpora. The only not significant domains in analyticity are LE and PB with an LL value of 0.22. The analytic log likelihoods between rest of the domains are either highly significant or very highly significant. It can therefore be stated that spoken ELF portrays variation in analyticity in most domains.

Table 9. *Log likelihood values between the five VOICE domains. The * -asterisk marks a statistically significant p-value.*

DOMAINS COMPARED	ANALYTIC LL	P-VALUES	SYNTHETIC LL	P-VALUES
PO AND PB	251.38	p<0.0001***	38.19	p<0.0001***
LE AND PB	0.22	p>0.05	33.03	p<0.0001***
PR AND PB	10.24	p<0.01**	48.25	p<0.0001***
LE AND PR	10.09	p<0.01**	1.01	p>0.05
PO AND LE	165.59	p<0.0001***	1.87	p>0.05
PB AND ED	122.34	p<0.0001***	6.06	p<0.05*
PR AND ED	30.94	p<0.0001***	84.89	p<0.0001***
ED AND PO	19.66	p<0.0001***	89.97	p<0.0001***
LE AND ED	86.81	p<0.0001***	63.21	p<0.0001***
PR AND PO	81.02	p<0.0001***	6.93	p<0.01**

The synthetic frequencies also exhibit variation, although with more inconsistent fashion and with smaller log likelihood values. However, there are still six very highly significant p-values

smaller than 0.0001. These significances are between PO and PB, LE and PB, PR and PB, PR and ED, ED and PO, and LE and ED. Syntheticity between PB and ED and between PR and PO is significant but with more moderate LL and p-values ($p<0.05$ and $p<0.01$). On the other hand, LL produced statistically not significant results for the comparisons between LE and PR and PO and LE. This is intriguing, for an observation of Figure 3 shows a distinguishable gap in syntheticity between the domains. ED and PB appear to situate more similarly while PO, PR, and LE are grouped closer together. This observation is supported by the fact that the statistical differences inside these two groups are either not significant (LE and PR, PO and LE) or significant but not highly significant (ED and PB, PR and PO). A similar distinct division between the domains cannot be observed in analyticity. Possible reasons behind this dichotomy of domains are addressed in the discussion section below. Regardless, also these synthetic results suggest that there is variation and significant differences in the compositions of the different spoken ELF domains. None of the comparisons produced insignificant results so that both analyticity and syntheticity between the same domains would be affected. In other words, although the syntheticity between LE and PR is not statistically significant, analyticity is. So, all domains in VOICE are statistically distinguishable from each other either by analyticity, syntheticity or by both.

To further demonstrate the differences between the domains, below are examples from PO and PR. Both extracts discuss language in professional settings, but a clear distinction can be made between the two domains. PO speaker uses more vague colloquial terminology while the PR speaker applies linguistic jargon.

    (1)   hh but it will still be a bit difficult because I have to transcribe the whole thing
                    hh and I mean we use er **specific language and specific terms** and

                                                (VOICE, POcon543)

(2)   I do see two more topics er [speaker3] and [speaker2] raised I mean I just mention it maybe we are just too tired by now erm one is erm **the semiotics the very specific semiotics of translation** the the shifts of of concepts if they are translated into another culture

<div align="right">(VOICE PRpan225)</div>

After providing answers to the two first research questions, let us concentrate on answering the third, concerning a comparison with a native variety. In addition, the domain variance is further explored and expanded on.

6.2. Variance

The native variety chosen for this typological comparison is the British Standard and its spoken genres provided by Szmrecsanyi (2009) in his study. First, an overall review on the grammaticity differences between spoken ELF and spoken Standard British English. As already reported, the mean grammaticity of the VOICE domains is 548, while the spoken genres in BNC exhibit a mean grammaticity index of 638 (calculated on the basis of numbers given by Szmrecsanyi 2009, 333). The BNC portrays 90 index points higher grammaticity than VOICE. However, this result was expected and confirms previous research since Laitinen (2018, 125) already observed this substantive difference in the two spoken varieties.
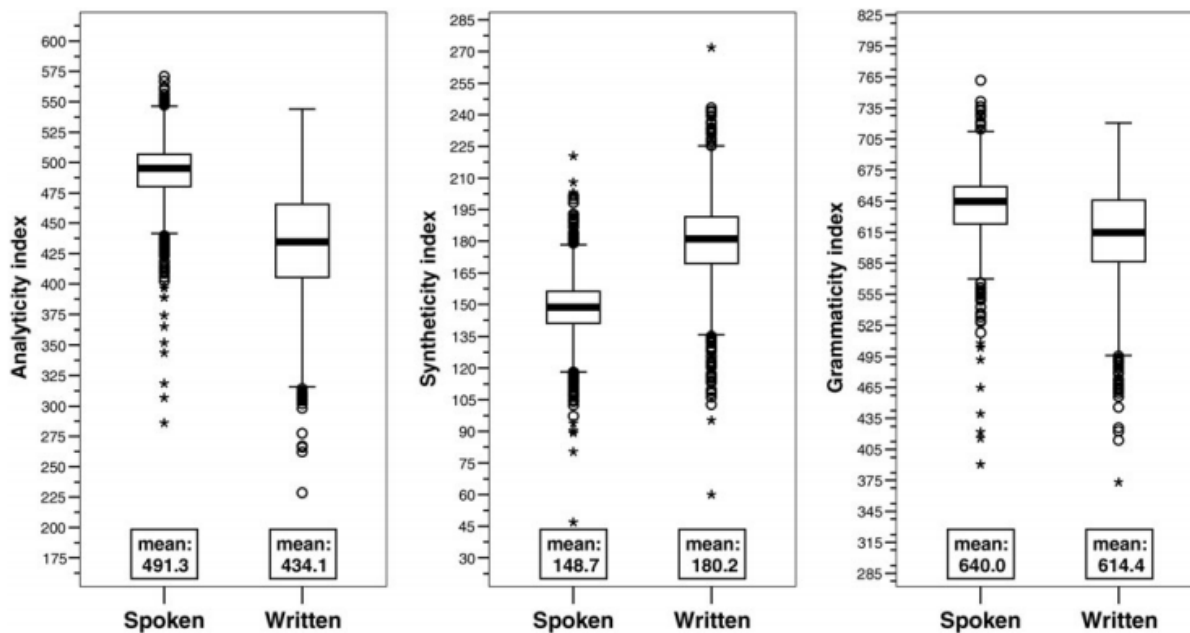
FIGURE 3. Spoken vs. written text types (variance in index points, ptw).

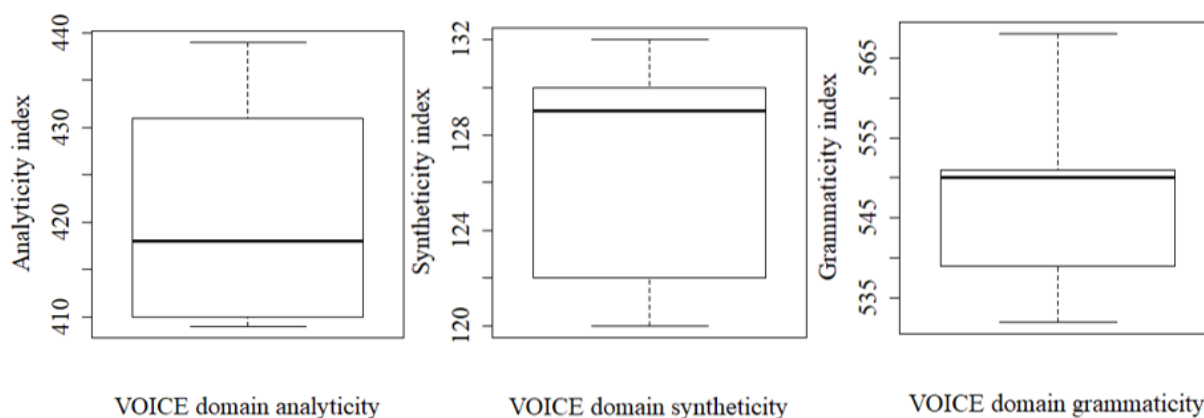**Figure 4**. *Box plots of spoken and written text types by Szmrecsanyi (2009, 336).*

Figure 5. *Box plots of VOICE domain analyticity, syntheticity, and grammaticity variance.*

Figure 4 demonstrates the variance in BNC that Szmrecsanyi (2009) discovered. He (ibid., 337) further elaborates that the "interquartile range for spoken texts spans only about 25 index points". Compared to the written texts in the BNC the spoken texts are less varied, however, a similar examination of VOICE reveals that the domain grammaticity variance is around 12 index points (Figure 5). Thus, it can be supposed that spoken ELF shows even less variance among the domains than the compared spoken Standard British genres. Of course, the scarcity of data, there only being five domains in VOICE, affects this variance. Comparing the two sets

of box plots shows that indeed, overall VOICE is substantially less varied; there are no outlier points and the boxes are short and compact. The BNC box plots on the other hand, are more spread out hence more varied. Naturally, again, what is important to remember is that these spoken BNC values are calculated from 16 genres, while the VOICE only covers five domains. Despite this data size difference, a Chi squared statistical test confirms that the difference of the two variance indices is statistically significant with a *p*-value < 0.05 (chisq 6.83, df=1). A visual representation of the variance difference between VOICE and BNC is provided in a density map in Figure 6. The density map shows how the two varieties form two distinct circles, BNC being the more concentrated since there are more genres. Furthermore, Table 10 lists the analytic and synthetic index values of all spoken BNC genres next to the VOICE domain values. This numeric comparison further demonstrates how low the VOICE domain AI and SI are in contrast to the BNC genres.
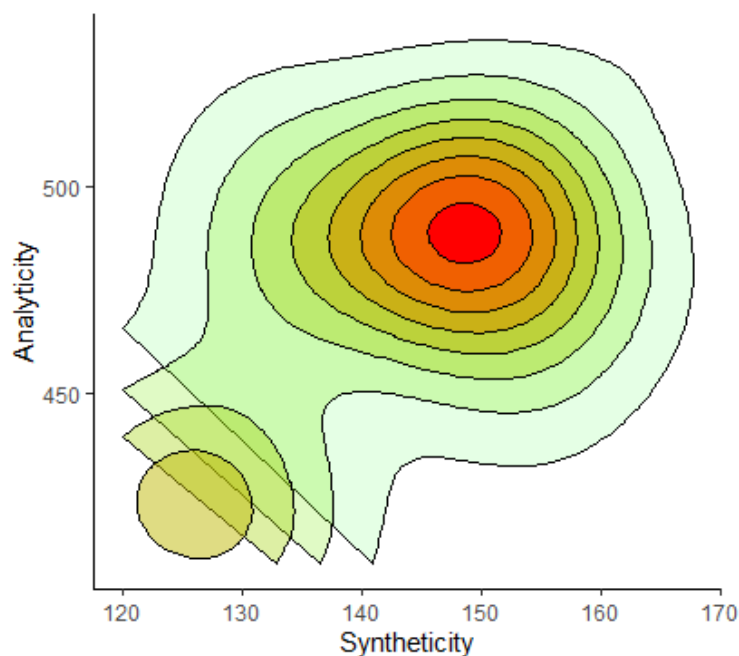

Figure 6. *Density map of VOICE domains and BNC genres.*

Table 10. *Analyticity (AI) and syntheticity (SI) indices of spoken BNC genres (from Szmrecsanyi 2009. Values rounded by the author) and VOICE domains.*

BNC genre	AI	SI	VOICE domain	AI	SI
Sermon	548	168	ED (educational)	431	120
Demonstration	512	132	PO (professional organisational)	439	129
Consultation	484	134	PB (professional business)	410	122
Conversation	496	148	PR (professional research and science)	418	132
Meeting	495	146	LE (leisure)	409	130
Interview	504	153			
Courtroom	499	150			
Tutorial	492	155			
Lecture	482	151			
Broadcast	472	160			
Sports live	447	154			
Parliament	477	150			
Speech	482	149			
Unclassified	486	144			
Classroom	485	140			
Pub-debate	479	132			

The mean values of the two varieties under examination are also worth some further investigation. VOICE mean grammaticity being 548, and respectively 638 for the spoken BNC denotes that there is a 90 index point difference, like observed above. Szmrecsanyi (2009, 330) discovered a similar pattern, British varieties' mean grammaticity being higher than average and non-native indigenised L2 varieties' being lower than average, among 16 varieties of Englishes. Meaning that according to these VOICE results ELF appears to act like a L2 variety, at least based on low grammaticity and in comparison to Standard British English results. What needs to be remembered is that Szmrecsanyi's frequencies include written and spoken data. However, it is interesting that the spoken ELF and spoken Standard British English data mimic previously observed results so closely. Based on Szmrecsanyi's (ibid., 329) values spoken ELF resembles most closely Singapore English which has a GI of 549. Szmrecsanyi (ibid.) includes

Singapore English as a non-native indigenised L2 variety. Furthermore, a comparison to Szmrecsanyi and Kortmann's (2011, 174) results of indigenised varieties and learner Englishes reveals that the analyticity levels of ELF (421) are closer to those of indigenised varieties of English (AI 438) than to those of learner Englishes (AI 494). Therefore, this supports previous research findings that ELF is typologically not like learner Englishes but more similar to the non-native L2 Englishes (e.g. Ranta 2018).

Szmrecsanyi (2009, 337) also discovered the spoken genres in BNC adhere to the notion that "increased analyticity incurs reduced syntheticity, and vice versa". In other words, he found that analyticity and syntheticity correlate weakly in the spoken genres ($r = .13$, $p < .001$) (ibid.). So, when a spoken genre is high in analyticity it is more likely to portray lower levels of syntheticity and vice versa. This tendency was not observed in the ELF data. The VOICE domains show no correlation whatsoever ($r = -.105$, $p > .87$). This result suggests that, although log likelihood reveals statistically significant differences between the domains in analyticity and syntheticity in spoken ELF, the variance is not systematic.

Lastly, before turning to domain and genre specific comparisons, a final angle of view on general variance between these two varieties. Among these chosen BNC genres the most substantial difference is between 'pub-debate' (AI 479) and 'demonstration' (AI 512) with 33 index points, while in VOICE the analyticity difference between PO (AI 439) and LE (AI 409) is 30 index points. So, the range of variation in analyticity between the two varieties is very similar. Whereas, syntheticity fluctuates more. The most synthetic BNC genre is 'tutorial' (SI 155) and the index point difference to least synthetic genre 'pub-debate' (SI 132) is 23. This is nearly double to VOICE values, where the index point difference is 12 between PR (SI 132) and ED (SI 120). In this respect, the two varieties seem to portray variation differently. The

spoken BNC genres vary more according to syntheticity while in VOICE the domains are more distinguished in analyticity. Whether this distinction is due to ELF speakers avoiding syntheticity and thus complexity, or because of their general aim towards reduction of redundancy is not apparent based on these results but is nevertheless something worth investigating in future research.

6.3. Comparison of VOICE domains and BNC genres

Next, some comparison of the individual BNC genres and VOICE domains. Not all of the BNC genres are included in this comparison as explained in the methodology of this study. Figure 7 exhibits the genres under scrutiny.

A direct normalised frequency comparison with the BNC genres and VOICE domains would not be valid since, the two varieties are distinctly different, as demonstrated above. Thus, the general relationship between the domains and genres is examined. This means that the relative difference or similarity of the domains and genres is investigated. For example, how the BNC genre 'meeting' is situated, according to its typological composition, analyticity, syntheticity, and grammaticity, in contrast to the rest of the BNC genres and whether its VOICE domain counterpart is similarly placed and composed among the domains. In other words, while the most synthetic domain PR includes spoken texts from research and science surroundings, is this result replicated in the BNC genres? Research and science is an area of language use generally deemed relatively formal and thus high syntheticity is an expected result, and BNC appears to follow this assumption. The three most synthetic genres in this restricted selection of BNC genres are 'tutorial', 'interview', and 'lectures', of which 'interview' and 'lectures' (university

lectures i.e. fall under research and science. See Table 4 for sub-types) can at least be considered

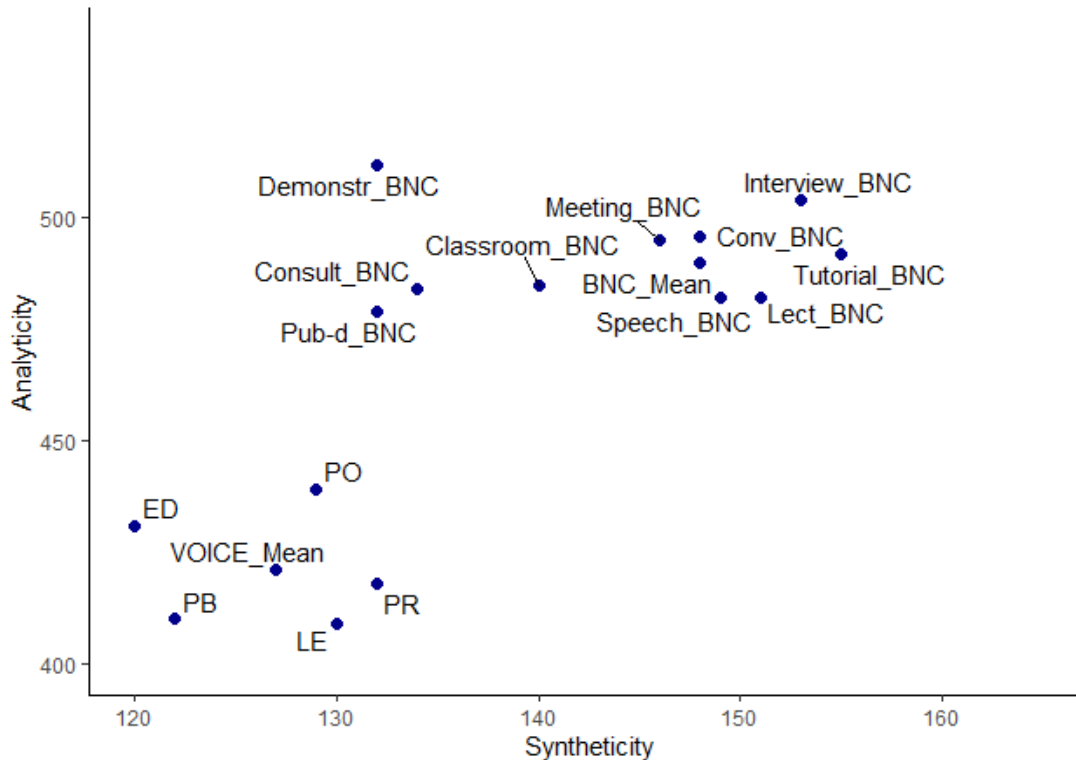formal and corresponding to PR in some respect.

Figure 7. *Scatterplot of the VOICE domains and selected BNC genres.*

The least synthetic VOICE domain ED is also the second highest in analyticity and this sits

well with previous research. Explicitness and understandability are important in educational

settings and the typological composition of ED compared to the other domains follows this.

This perception also applies to the BNC genres if examined in this fashion. The 'classroom'

genre is not the least synthetic but still clearly among the less synthetic genres of the BNC, and

approximately same in analyticity as the BNC mean. The 'demonstration' genre is significantly

less synthetic and more analytic than the BNC mean. A demonstration could be perceived as a

type of educational situation, in which the comprehension of the listener is of importance and

the presenter should aim for explicitness. After all, lower syntheticity, higher analyticity, and

higher grammaticity are associated with transparency, enhanced comprehension, and decreased complexity (Szmrecsanyi 2009, 323), and these qualities are more or less fulfilled and perhaps aimed at in educational speech environments.

Of the BNC genres under scrutiny, lowest in syntheticity are the genres 'pub debate', 'demonstration', 'consultation', and 'classroom'. In all of these four BNC genres, speech economy could be said to be a secondary attribute as long as the opposing side understands what is attempted to convey, which is why analyticity frequencies are high and syntheticity low. As already exemplified, ED, 'classroom', and 'demonstration' could all be assessed as belonging to a broadly similar category. However, reflecting on the other two BNC genres, 'pub-debate' and 'consultation', it could be stated that they both also require the aforementioned transparency, easy comprehension, and low complexity in order for a successful communication. Of course, compared to BNC mean, "consultation" and "pub-debate" are not higher in analyticity but on a similar level. However, it needs to be remembered that overall, the spoken BNC genres vary more in syntheticity than in analyticity (Figure 7).

Interestingly, the least grammatic domain PB is low in both analyticity and syntheticity. The VOICE domain definition (2013) states that "the professional business domain includes all social situations connected with activities of making, buying, selling or supplying goods or services for money". This description indicates that most likely speech events in PB are not long discussions but more like fast transactions. Thus, it could be assumed, and the grammaticity levels support the reading, that the recorded interactions comprise largely of lexical word utterances. As the example (3) from PB below exhibits, the utterance and social situation is not casual conversation, either is it specific professional jargon but something from between these two.

(3)   hello I want to apply for the visa of my family - - I went to the district office

and they told me that I need a letter - - That I'm living here and paying my rent

and

(VOICE, PBsve430)

None of the BNC genres resemble PB by genre label although, the largest speech event type in the domain is 'meeting'. While PB is the least grammatic of the VOICE domains, 'pub-debate' is the lowest in grammaticity in BNC. These two genres hold little similarities. Therefore, the BNC 'meeting' is the genre of comparison for PB here. The 'meeting' genre is higher in analyticity and slightly lower in syntheticity in relation to the mean BNC values. Whereas, PB is lower in syntheticity than the mean VOICE values but also low in analyticity so in this respect ELF and British standard appear to diverge. As already mentioned, some reservations must be held with these comparisons since the corpora are gathered form different environments, at a different time, and not categorised to correspond.

The domains PO (AI 439, SI 129) and LE (AI 409, SI 130) are interesting as they are similar in syntheticity but significantly differ in analyticity. Furthermore, the largest speech event type in terms of number of tokens in the PO domain is 'working group discussion' while in LE it is 'conversation' meaning that the texts are from similar speech events (VOICE 2013). However, the distinction is that PO is a 'Professional' domain which could explain the analyticity. In professional settings there is a need to be understood in communication so that work efficiency is guaranteed. Leisurely conversation on the other hand probably has no other goal than that of conversation, so there is room for possible misunderstandings. An interesting observation is that although LE among the VOICE domains is high in syntheticity, the syntheticity level of LE resembles the syntheticity of the BNC genres 'demonstration', 'consultation', and 'pub debate', previously deemed informal genres. In particular, 'pub-debate' represents highly

informal and casual interaction that could perhaps be compared with the LE domain. Thus, in this instance, ELF appears to conform to native Standard British syntheticity instead of opting for explicitness. The examples below demonstrate the language used in the 'pub-debate' genre and LE domain.

(4)   I want to say the thing that 's happening at the moment in my difficult life that I 'd hold onto like a drowning man with a log. One thing that works at this time forget today and yesterday and all the week it 's the number of children that come into this playhouse now we have us we have a thing called work experience where they come from the school 's and the poor little bugger 's have **got ta** work with me

(BNC, D91PS003)

(5)   yeah yeah yeah so I mean I have s- do some thinking first what we 've **got ta** buy

(VOICE, LEcon565)

In both extracts the modal *gotta* is present which demonstrates the colloquial nature of the language used. Furthermore, in VOICE data overall *gotta* is not frequent, so it is interesting that it appears specifically in the leisure domain. As non-native Englishes are generally lower in analyticity, the low analytic frequencies among the VOICE domains is no surprise either and this is reflected in the examples above. The BNC extract utterance is longer and laden with redundancies, words and idioms that do not necessarily provide anything towards the idea attempted to convey. Respectively, in the ELF example these redundancies have been eliminated as previous research by Björkman (2018) has exhibited, and what is left is merely the information in a colloquial form. This is further supported by Szmrecsanyi (2009, 323), who states that redundancies and increased grammaticity are connected, which is perfectly demonstrated in these examples. What could also contribute to the relatively high syntheticity in LE is that in informal, low pressure situations like the speech events in LE, ELF speakers

might be more inclined to try native like constructions and customs. This then appears in this data as high syntheticity in relation to other domains, and in adoption of constructions like *gotta*. Furthermore, casual conversation can be fast paced, and speech turns change quickly so language economy is highlighted, which might further promote syntheticity compared to the other domains. However, all this is hypothetical speculation and needs further confirmation through research. Nevertheless, in light of current data and resources available, these conclusions appear reasonable.

Another interesting cross-variety comparison is the PO domain from VOICE and the 'demonstration' genre from BNC. Both of these are the highest in analyticity within their variety and considerably similar in syntheticity overall. The professional organisational domain has an analyticity index of 439 and syntheticity index of 129 markers per thousand words. For 'demonstration' the same values are 512 analytic markers and 132 synthetic ones. Thus, the analyticity differs by 73 index points between these two spoken genres and syntheticity only 3 index points. The examples from both corpora show the kind of language used:

6) Sixty to eighty you can feel it at pulse point , ca n't you , here , here , we 'll show you so more in a minute , okay ? So you know how to check whether the circulatory system is working because if it 's working you 'll be able to find a pulse pounding away , yes , that 's how you check it , you know the respiratory system 's working because if you want to check it what do you want to look for ?

    (BNC, PS1P9)

7) arrange er such a structure of of this manual manual so things to be decided before you set up the consortium things you have to find before you set up the consortium and then the things that we have to deal wi- with having set up the consortium and those will be academic and administrative things or academic and organizational things

    (VOICE POmtg314)

Both extracts display instructive language with some repetition, but again the BNC example appears to utilise more constructions that could be assessed redundant. For instance, the tag question "can't you" and the added descriptive end in "pulse *pounding away*". This kind of language use is absent from the ELF example, which could provide towards the more pronounced analyticity in the 'demonstration' genre. The PO domain is composed of primarily recordings from different types of discussions or meetings (VOICE 2013). In other words, the texts come from relatively casual speech events despite being from professional environments. From these genre labels and descriptions, it is therefore surprising that the professional organisational domain and 'demonstration' genre share resemblances in some respect.

For example, the pair that could be presumed to resemble each other, and does resemble to a certain limit, is the educational (ED) domain and 'classroom' genre. ED is lowest in syntheticity of the domains with a SI of 120, and while the 'classroom' situates as the fourth lowest BNC genre it is still considerably more synthetic than the least synthetic genre (Figure 7 exhibits that 'classroom' clearly is not in the same grouping as the three least synthetic BNC genres). Furthermore, ED has an AI of 431, the second highest in VOICE, while 'classroom' is close to the BNC mean (AI 490) with an AI of 485. So, although they share titular similarity, and are both less synthetic than the variety mean, it cannot be stated that the distribution of educational domain and genre inside the varieties is the same.

The observations between individual domains and genres have demonstrated how ELF and Standard British English compare. Interestingly, even though similarities could be recorded in the larger general scale those results do not replicate necessarily onto the genre level. It was above detected that, both in VOICE and in BNC more academic genres are more synthetic while

informal genres are generally less synthetic and more analytic. However, when individual domains and genres are examined the two varieties appear to have differing internal genre distributions. For instance, arguably the most informal VOICE domain LE (leisure) is second highest in syntheticity, a result which based on previous research and theory should be opposite. Indeed, in BNC data the least synthetic genre 'pub-debate' is also the most informal. Reflection on the reasons behind this result was discussed more above, but it could be that in this instance ELF users mimic native speech which results in this unexpected typological composition in comparison to the other domains.

Furthermore, the domain and genre pair that do resemble each other in some respect, PO and 'demonstration', are by genre description dissimilar. These two genres are both highest in analyticity among their respective varieties as well as having similar synthetic frequency. Despite the unifying factors it must be admitted that when examined inside the varieties (Figure 7) these two genres are distributed differently in relation to variety means. PO is more synthetic than the VOICE mean, when 'demonstration' is significantly lower in syntheticity than the BNC mean. Furthermore, although both genres are highest in analyticity in their varieties, there is still the 73 index point difference between PO and 'demonstration'. Therefore, these similarities cannot be considered more than incidental similarities.

Due to the dissimilarity of the data classifications in the two corpora (domains and genres), any declarations on typological difference or resemblance between the VOICE domains and BNC genres cannot be given. This means that based on this study's data, it is not reliable to suggest that ELF and Standard British English genres are organised in the same way inside the varieties. It appears that for example, in both varieties the more academic domains and genres are more synthetic than the variety mean, whereas in educational settings syntheticity is significantly

lower. Which implies that ELF users accommodate their language use to some respect according to situation, like native English speakers. However, the VOICE domains and BNC genres are gathered and arranged so differently that a one-to-one comparison is not applicable.

Nevertheless, what can be relatively confidently confirmed is that there is variation among ELF domains. The five ELF domains in VOICE appear to vary significantly in their grammatical marker composition. Also, from corpora examples, a pattern could be observed that ELF users reduce redundancies actively, like suggested in previous research (Björkman 2018). While, ELF and Standard British English can be partially and superficially contrasted based on these results, the truth is that the VOICE domains and Szmrecsanyi's BNC genres differ. The domains appear to differ in genre distribution but also in corpus composition, and genre division, which is why the domain-genre comparisons are not reliable. Therefore, a more detailed and more appropriately suited comparison would require a native corpus with divisions of speech categories more similar to VOICE, or either a new spoken ELF corpus. In the next section, further discussion on ELF variation and these results.

7. DISCUSSION

The objective of this study was to investigate whether various spoken ELF domains, such as professional and educational, differ in their grammatical marker composition, and whether the possible differences between the domains are similar in relation when compared with a native variety, in this case Standard British English. These results give further insights to ELF structure, adding to previous knowledge of the typological status of ELF (mainly Laitinen 2018). Szmrecsanyi (2009, 333) has provided evidence that British English shows typological differences between different kinds of spoken and written interactions. Furthermore, Laitinen (2018) discovered similar genre distinctions in written ELF. Therefore, an assumption that spoken ELF could also portray similar behaviour was deemed probable. The five ELF domains of the VOICE corpus under examination in this study represent different speech environments thus enabling comparisons between the domains. The results that emerged imply that ELF does exhibit variation in spoken domains, and hence it also implies that ELF users adapt their language to suit the situational conventions.

Although ELF users are not as fluent speakers as natives, this study found that, based on the domain variation, ELF users appear to accommodate and change their speech so that the domains can be typologically specified as distinct. However, although ELF appeared to vary in a similar fashion as the native counterpart in some respect (research, science, and other academic genres are, on average, in both varieties more synthetic than other genres, for instance), a dissimilarity in some genre distributions was visible when contrasted to Standard British English (previous research by Szmrecsanyi (2009) was employed to represent the native English counterpart). It was found that inside the two varieties, similar genres are not consistently and systematically similar in relation to their respective variety means. Of course, there are limitations to these results which are further expressed below. The results obtained

give new insights onto ELF, its typology, and its essence. ELF is currently considered somewhat controversial since it cannot be accepted as a full-scale distinct variety while, on the other hand, it is becoming increasingly common, gathering new speakers and users every day in this global era of internet and freedom of movement. The results, what they imply, study limitations, and suggestions for future research are discussed in this section more in depth.

Firstly, the aim was to figure out spoken ELF typology by determining analyticity, syntheticity, and grammaticity in different spoken domains. The VOICE corpus was analysed to retrieve the figures. The frequencies acquired exhibit that the five VOICE domains hold differing typological compositions. The professional organisational (PO) domain was found to portray highest analyticity index (AI) of 439 and syntheticity index (SI) of 129. The educational (ED) domain has an AI of 431 and a SI of 120. Professional business (PB) has an AI of 410 and a SI of 122, and the last professional domain research and science (PR) portrays AI of 418 and SI of 132. Finally, leisure (LE) shows lowest analyticity with AI of 409 markers and a SI of 130. Thus, the grammaticities of the domains are: PO 568; ED 551; PB 532; PR 550; and LE 539. Based on these values the second research question can be answered.

Statistical testing of the analyticities and syntheticities reveals whether the frequency differences between the five domains are significant. The loglikelihood statistical testing revealed that nearly all comparisons among the domains are statistically significant (table 9). This indicates domain variation in ELF and provides quantitative confirmation for the qualitative hypothesising why certain domains might differ. In analyticity, only LE and PB are not statistically significant, otherwise the domains are all distinct in relation to each other. In syntheticity there is more dispersion but still, only two calculations produce not significant results. The syntheticity between LE and PR, and between PO and LE are not statistically

significant, meaning that the syntheticity levels between these domains do not vary enough to produce statistically detectable differences.

Those insignificant results however prove interesting. Figure 3 portrays the domains in a scatterplot according to analyticity and syntheticity. What can be observed from the scatterplot is that syntheticity appears to divide the domains into two groups with a definite gap in between. ED and PB are clearly less synthetic than PO, PR, and LE, which form the more synthetic grouping. There is no direct cause available to explain this division. The reason cannot be found in the speech event types, for each domain includes several speech event types. For example, 'meeting' type is one of the most common ones in both PB and PO, and 'conversation' speech event type can be found in all domains. Although, this analysis on speech event types does not solve the syntheticity gap question, it does support the finding that ELF varies according to domain. These findings show that the speech event type does not influence the domain typology. It appears thus, that the speech event 'conversation', for instance, is accommodated by ELF users to suit the domain, but not vice versa so that domain language use would change because of the speech event. Without obvious cause, it could be then simply a coincidence which produced this syntheticity division. Afterall, the synthetic marker frequency difference between the groups is only 7 index points (PB 122; PO 129). Of course, there might be underlying factors contributing to this phenomenon, like the first languages of the ELF speakers. Whatever the reason, there is no room in this study to delve into this further. Whether there is a domain gap in VOICE or whether it is due to chance, the question remains unanswered for future research.

Let us now return back to the main findings, where a more detailed discussion on the results is due. Firstly, the initial marker frequency levels of this study support the findings by Laitinen

(2018) that spoken ELF exhibits low grammaticity overall. This is revealed when the VOICE mean grammaticity (548) is compared to spoken BNC mean grammaticity (638). As per Szmrecsanyi's (2009, 323) typological definition the low grammaticity signifies increased speaker output economy and decreased redundancy. However, from the non-native English perspective, the expected ELF result would have been decreased complexity which in theory would in turn suggest high grammaticity and especially high analyticity. On the other hand, Björkman (2018, 260) has proposed that for non-native English speakers the decreased redundancy in fact promotes comprehension while producing low grammaticity as a by-product. This view would in some respect explain the ELF results. It could be proposed that ELF users reject markers that are not essential for the communicated utterance to be understandable and thus concentrate on the lexical words. Furthermore, the domain differences imply that accommodations in the composition are made according to the situation. Despite the overall low grammaticity, the VOICE data shows that when enhanced explicitness and clarity is required also the marker frequency is increased in the domains. For example, the educational domain (ED) shows higher frequencies of analyticity than the leisure domain (LE). This appears reasonable when the nature of both domains is considered. It could be assumed that educational settings try to produce easily understandable speech, and one way to achieve this explicitness is increased analyticity. As Szmrecsanyi notes "increased analyticity increases explicitness and transparency and decreases hearer/reader comprehension complexity (2009, 323).

Certainly, this increased analyticity is exemplified in the domains ED and PO, the two being the most grammatic and analytic of the domains. It can be argued that from the five domains, these two place most emphasis on the ease of comprehension. Educational settings are rather self-explanatory in their need for minimal complexity. PO, the professional organisational domain, on the other hand, is the least formal of the professional domains, covering "activities

of international organizations or networks which are not doing research or business" (VOICE 2013). In other words, most likely work-related discussion, including speakers giving mutual advice and explaining or planning projects, which entails easily understandable language, as has been exemplified in the corpus extracts in the results section above (Examples (1) and (7)).

Conversely, PR the most synthetic domain is arguably also the most formal one. Research and science as a language genre is formal and generally it is acknowledged that formal genres tend to exhibit higher syntheticity with decreased analyticity. In this way, ELF appears to act like its native variety of comparison. In the big picture, the typological compositions of the domains happen to more or less follow the native conventions even though the overall grammaticity is significantly lower. More relevantly, these results suggest reliably that there is internal variation in ELF, and that that variation is not entirely arbitrary.

The third research question placed ELF and Standard British English parallel to each other. The intent was to compare the VOICE domains and BNC genres to determine whether they are distributed in a similar fashion in the two varieties. Before comparing individual genres, the variance in the two varieties was examined more closely. Since there are only five domains in the VOICE data, ELF variation is not as pronounced as in the native BNC comparison data which has sixteen spoken genres. Possibly due to this data size difference, it was discovered that the ELF variation does not seem to be systematic in the way Standard British English is. In Szmrecsanyi's (2009) BNC data, the spoken genres showed correlation of analyticity and syntheticity levels, a result that was not replicated on the ELF data. Statistical testing recorded no correlation whatsoever among the VOICE domains, but Szmrecsanyi (2009, 337) reports a correlation in the BNC that increased analyticity denotes lower levels of syntheticity in the spoken genres. This means that the variation detected in ELF, once again, does not follow the

native convention, a phenomenon that was observed in Standard British English. To interpret this further, it partly answers the third research question placed; overall variation in spoken ELF and Standard British English is not identical. Again, however, it must be reminded that with only five domains to compare in VOICE, it might not be enough for a correlation pattern to emerge. Nevertheless, with the VOICE corpus data, the result depicting correlation of analyticity and syntheticity could not be detected in ELF.

In order to answer the question of genre distribution, individual VOICE domains and BNC genres were compared. This study has already exhibited that ELF domains vary, and Szmrecsanyi (2009) likewise demonstrated how native British speakers accommodate their speech according to genre. The scatterplot in Figure 7 illustrates visually this variety comparison. It is apparent then that both varieties, ELF and Standard British English, are varied according to domain and genre, but what needed further examination was whether the genre distributions are similar in these two varieties. A surface level investigation into the domains and genres showed that ELF does conform to Standard British English conventions. This is exemplified in PR, the professional research and science, domain, for instance. As PR is the most synthetic domain ELF obeys the native practice of concentrating increased syntheticity in the more academic domain. It is generally acknowledged that formal and academic English is more synthetic. Among the selected BNC genres, the third most synthetic is the 'lecture' genre, entailing university level lectures, which is the closest equivalent to PR. Furthermore, the quick overlook showed that in both varieties speech from an educational situation is low in syntheticity in relation to other genres or domains. In this respect ELF and Standard British English were recorded to share resemblances. However, when individual genres and domains were studied more closely the similarities were not as prominent. Although, there are not as many domains in VOICE as there are genres in BNC, and the domains are not divided to

correspond the genres, a general comparison was nevertheless attempted based on similar qualities in the domains and genres.

The most prominent difference between the two varieties could be seen in a comparison of the more casual domain and genres. For example, leisure domain and 'pub-debate' genre are actually very similar in syntheticity levels (LE 130; 'pub-debate' 132) but instead of contributing towards similarity of varieties it differentiates them. When considered in relation to each variety mean they are situated completely differently. LE is the second most synthetic domain in the VOICE while 'pub-debate' shares the title of least synthetic genre in the BNC. So, in effect this speech category is on the opposing side in ELF compared to Standard British English. It was reasoned that this could be due to ELF speakers attempting more native like language since the domain mainly includes texts from informal and casual conversations. However, this observation does not support the notion of similarity of ELF domain and Standard British English genre distribution.

Another comparison found to further diminish the ELF and Standard British English genre distribution similarity was that of PB and 'meeting'. This pairing was chosen since the largest speech event type in PB is 'meeting' and none of the other BNC genres better correspond to the professional business domain. The PB is the least grammatic domain in VOICE with a GI of 532. This means 26 index points lower grammaticity than the corpus mean (548). Conversely, 'meeting' is close, and slightly higher than the BNC mean (638) with a GI of 641. In other words, while PB is overall lower in both analyticity and syntheticity in relation to VOICE mean, 'meeting' almost represents the BNC mean. Therefore, in this occasion again the ELF domains do not comply with BNC genre distribution.

These results and findings have demonstrated how spoken ELF and Standard British English variation and genres compare to each other. In some instances, and when examined in a more general fashion, it has become apparent that the genres hold resemblances. However, a closer examination of individual genres and their relation to corpora means revealed that there are differences in the distributions.

Lastly, a remark on the frequencies is necessary at this point. Compared with the results by Laitinen (2018), especially the analyticity index requires attention. Laitinen (2018, 122) reports an analytic marker frequency of 431 and a synthetic marker frequency of 122. The analytic frequency by Laitinen is higher than the respective frequency of 421 in this study. Since the corpus (VOICE) and the analytic markers in both studies are almost identical, a more similar result would have been expected. It is likely that a systematic error has occurred during the analysis of this or both studies, resulting in this difference. The synthetic marker frequencies, on the other hand, behave as could be expected. The added marker tags, although not high in frequency across the domains, elevate the value slightly in this study to 127 markers per 1000 words. Overall, the GI difference of five index points (Laitinen reports a GI of 553, the GI of this study is 548) is not statistically significant.

In conclusion, this study has demonstrated that spoken ELF does portray internal variation like the established varieties of English. The frequency differences between the ELF domains were statistically verified which suggests reliable results that the variation is factual. However, the results concerning spoken Standard British English comparison were more imprecise. There were similarities recorded between spoken ELF and Standard British English domains and genres, but all the while the similarities did not extend through all of the data. Therefore, further research is required in order to establish how spoken ELF genres are distributed. However, to

do that, a wider spectrum of genres is needed to ascertain the spoken ELF genre typology in relation to a native variety. The five domains in VOICE do not cover sufficiently all speech environments. With more domains or genres, the analyticity-syntheticity correlation could be studied anew. Furthermore, the texts and speakers in VOICE are all from relatively academic backgrounds and largely European, so the author would like to see a similar study conducted on data which includes more wide variety of speakers as well. Perhaps entirely new spoken ELF corpus would be due, considering the recordings in VOICE were gathered between 2001 and 2007. In almost twenty years ELF speech might have already evolved. Therefore, a new spoken ELF corpus would also enable a diachronic study on ELF. Other suggestions for future research on spoken ELF would be to examine which particular markers contribute to domain analyticity, syntheticity, and grammaticity, and if the markers differ from domain to domain. By examining the frequencies of markers, it would reveal more of the relationship between spoken ELF and spoken native varieties. That way structural and typological differences could be assessed more accurately. All in all, this study has provided new intel on spoken ELF and that it internally varies according to domains. What is more, it was exhibited that this ELF internal variation does follow some native conventions but that a similar genre distribution should not be presumed without some reservations.

8.   REFERENCES

Baker, W. 2015. *Culture and Identity Through English As a Lingua Franca: Rethinking Concepts and Goals in Intercultural Communication*. Berlin/Boston: De Gruyter Mouton.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. 1999. *Grammar of Spoken and Written English*. London: Longman.

Björkman, B. 2013. *English as an Academic Lingua Franca: An Investigation of Form and Communicative Effectiveness*. Berlin: De Gruyter.

Björkman, B. 2018. Morphosyntactic Variation in Spoken English as a Lingua Franca Interactions. In Jenkins, J., Baker, W. & Dewey, M. (eds) *The Routledge Handbook of English as a Lingua Franca*, pp.255–266. Abingdon: Routledge.

Busch, B. 2012. The Linguistic Repertoire Revisited. *Applied Linguistics*. 33(5):503–523.

Caffarel, A., Martin J. R., & Matthiessen C. 2004. *Language Typology: A Functional Perspective*. Philadelphia.: John Benjamins Publishing Company.

Cienki, A. J. 2017. *Ten Lectures on Spoken Language and Gesture from the Perspective of Cognitive Linguistics: Issues of Dynamicity and Multimodality*. Leiden: Brill.

Cogo, A. 2018. ELF and Multilingualism. In Jenkins, J., Baker, W. & Dewey, M. (eds) *The Routledge Handbook of English as a Lingua Franca*, pp.357–368. Abingdon: Routledge.

Daniel, M. 2011. Linguistic Typology and the Study of Language. In Song, J.J. (ed.) *The Oxford Handbook of Linguistic Typology*, pp.44–68. New York: Oxford University Press.

Dewey, M. & Jenkins, J. 2010. English as a Lingua Franca in the Global Context: Interconnectedness, Variation and Change. In Saxena, M. & Omoniyi, T. (eds.) *Contending with Globalization in World Englishes. Multilingual Matters,* pp.72–92. Bristol: Multilingual Matters.

"E, n.1." OED Online, Oxford University Press, 2020. www.oed.com/view/Entry/58919. Accessed 29 July 2020.

Eberhard, D.M., Simons, G.F. & Fennig, D. C. (eds.). 2020. *Ethnologue: Languages of the World*. Twenty-third edition. Dallas: SIL International. Online version: http://www.ethnologue.com.

Ellis, N. C. 2008. Usage-Based and Form-Focused Language Acquisition: The Associative Learning of Constructions, Learned Attention, and the Limited L2 Endstate. In Robinson, P. & Ellis, N. C. (eds.) *Handbook of Cognitive Linguistics and Second Language Acquisition*, pp.372–405. New York: Routledge.

Giles, H. & Smith, P. 1979. Accommodation Theory: Optimal Levels of Convergence. In Giles, H. and St. Clair, R.N. (eds) *Language and Social Psychology*, pp.45–56. Oxford: Blackwell

Greenberg, J.H. 1960. A Quantitative Approach to the Morphological Typology of Language. *International Journal of American Linguistics*. 26(3):178–194.

Haselow, A., 2011. *Typological Changes in the Lexicon : Analytic Tendencies in English Noun Formation.* Berlin: De Gruyter.

Jenkins J. 2006. Current Perspectives on Teaching World Englishes and English as a Lingua Franca. *TESOL Quarterly* 40(1):157–181.

Jessner, U., Allgaüer-Hackl, E. & Hofer, B. 2016. Emerging Multilingual Awareness in Educational Contexts: From Theory to Practice. *The Canadian Modern Language Review / La Revue Canadienne des Langues Vivantes*. 72(2):157–182.

Kachru B. 1985. Standards, Codification and Sociolinguistic Realism: the English Language in the Outer Circle. In Quirk, R. & Widdowson, H.G. (eds) *English in the World: Teaching and Learning the Language and Literatures*, pp.11–30. Cambridge: Cambridge University Press.

Kettemann, B. & Marko, G. 2002. *Teaching and Learning by Doing Corpus Analysis: Proceedings from the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July 2000.* Amsterdam: Rodopi.

Kimura, D. & Canagarajah, S. 2018. Translingual Practice and ELF. In Jenkins, J., Baker, W. & Dewey, M. (eds) *The Routledge Handbook of English as a Lingua Franca*, pp.295–308. Abingdon: Routledge.

Kortmann, B. & Szmrecsanyi, B., 2011. Parameters of Morphosyntactic Variation in World Englishes: Prospects and Limitations of Searching for Universals. In Siemund, P. (ed) *Linguistic Universals and Language Variation*, pp.264–290. Berlin: De Gruyter.

Laitinen, M. 2018. Placing ELF Among the Varieties of English. In Deshors, S. C. (ed) *Modeling World Englishes. Assessing the Interplay of Emancipation and Globalization of ESL Varieties*, pp. 109-131. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Larsen-Freeman, D. 2018. Complexity and ELF. In Jenkins, J., Baker, W. & Dewey, M. (eds) *The Routledge Handbook of English as a Lingua Franca*, pp.51–60. Abingdon: Routledge.

Leppänen, S. et al.. 2011. *National Survey on the English Language in Finland: Uses, Meanings and Attitudes.* Helsinki: VARIENG. [Online] Available at:
http://www.helsinki.fi/varieng/series/volumes/05/index.html
[Accessed 12.10.2018].

Mauranen, A. 2018. Conceptualising ELF. In Jenkins, J., Baker, W. & Dewey, M. (eds) *The Routledge Handbook of English as a Lingua Franca*, pp.7–24. Abingdon: Routledge.

McEnery, T. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge ; New York: Cambridge University Press.

McKay, S. L. 2018. English as an International Language: What It Is and What It Means For Pedagogy. *RELC Journal*, 49(1):9–23.

Mollin, S. 2006. English as a Lingua Franca: A New Variety in the New Expanding Circle? *Nordic Journal of English Studies.* 5(2):41–57.

Mäkisalo, J. 2009. Kielentutkijan Kvantitatiivisia Menetelmiä. In Kujamäki, P. & Mäkisalo, J. (eds.) *Kieli- ja Käännöstieteen Opetusmonisteita, 2.* Joensuu: University of Eastern Finland.

Opetushallitus. 2014. *Perusopetuksen Opetussuunnitelman Perusteet 2014.* Helsinki: Next Print Oy. Available at: https://www.oph.fi/sites/default/files/documents/perusopetuksen_opetussuunnitelman_peruste et_2014.pdf. [Accessed 29.07.2020].

Prodromou, L. 2008. *English as a Lingua Franca : A Corpus-based Analysis.* London: Continuum.

Ranta, E. 2018. Grammar in ELF. In Jenkins, J., Baker, W. & Dewey, M. (eds) *The Routledge Handbook of English as a Lingua Franca*, pp.244–254. Abingdon: Routledge.

Rayson, P. 2020. *Log-likelihood and Effect Size Calculator*. Availble at: http://ucrel.lancs.ac.uk/llwizard.html. [Accessed 29.07.2020].

Seidlhofer, B. 2001. Closing a Conceptual Gap: the Case for a Description of English as a Lingua Franca. *International Journal of Applied Linguistics*. 11(2):133–158.

Seidlhofer, B. 2018. Standard English and the Dynamics of ELF Variation. In Jenkins, J., Baker, W. & Dewey, M. (eds) *The Routledge Handbook of English as a Lingua Franca*, pp.85–100. Abingdon: Routledge.

Siegel, J. 2012. Accounting for Analyticity in Creoles. In Kortmann, B. & Szmrecsanyi, B. (eds.) *Linguistic Complexity*, pp.35–61. Berlin: De Gruyter.

Szmrecsanyi, B. 2009. Typological Parameters of Intralingual Variability: Grammatical Analyticity Versus Syntheticity in Varieties of English. *Language Variation and Change,* 21(3):319–353.

Szmrecsanyi, B. 2012. Analyticity and Syntheticity in the History of English. In Nevalainen, T. & Traugott, E. C. (Eds.) *The Oxford Handbook of the History of English*. New York: Oxford University Press.

Szmrecsanyi, B. & Kortmann, B. 2011. Typological Profiling: Learner Englishes Versus Indigenized L2 Varieties of English. In Mukherjee. J. & Hundt, M. (eds.) *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*, pp.167–188. Amsterdam: John Benjamins.

Szmrecsanyi, B. & Kortmann, B. 2012. Introduction: Linguistic Complexity. In Szmrecsanyi, B. & Kortmann, B. (eds.) *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*, pp.6 – 34. Berlin: De Gruyter.

Velupillai, V. 2012. *An Introduction to Linguistic Typology*. Amsterdam: John Benjamins Publishing Company.

VOICE. 2013. *The Vienna-Oxford International Corpus of English* [version 2.0 online]. Director: Seidlhofer, B; Researchers: Angelika Breiteneder, Theresa Klimpfinger, Stefan Majewski, Ruth Osimk-Teasdale, Marie-Luise Pitzl, Michael Radeka. Available at: http://voice.univie.ac.at [Accessed 2.6.2020].

VOICE. 2014. *VOICE Part-of-Speech Tagging and Lemmatization Manual.* Available at: http://www.univie.ac.at/voice/documents/VOICE_tagging_manual.pdf. [Accessed 2.6.2020].

Wang, Y. 2018. Chinese English as a Lingua Franca: An Ideological Inquiry. In Jenkins, J., Baker, W. & Dewey, M. (eds) *The Routledge Handbook of English as a Lingua Franca*, pp.151–164. Abingdon: Routledge.

Widdowson, H.G. 2018. Historical Perspectives on ELF. In Jenkins, J., Baker, W. & Dewey, M. (eds) *The Routledge Handbook of English as a Lingua Franca*, pp.101–112. Abingdon: Routledge.