Tampereen yliopisto

Jenna Honkanen

A CORPUS-BASED STUDY ON THE USE OF THE PROGRESSIVE FORM IN ADVANCED LEARNER ENGLISHES

Faculty of Information Technology and Communication Sciences
Optional Studies Thesis
April 2022

ABSTRACT

Jenna Honkanen: A corpus-based study on the use of the progressive form in advanced learner Englishes
Optional Studies Thesis
Tampere University
Advanced Studies in English Language and Literature as Elective Studies
April 2022

This thesis examines the progressive form in written, advanced learner Englishes. The progressive, i. e. a form of the auxiliary *be* and the present participle of the main verb, has been considered a challenging feature for learners of English mostly because of the various meanings and functions attached to it. Studies have found different patterns of usage of the form by non-native speakers compared to native speakers, such as overuse or use in non-standard contexts. This study attempts to examine the use of the form within the learner English variety by comparing different learner populations.

The data for the study comes from the International Corpus of Learner English (ICLE), which contains written essays by higher intermediate to advanced learners of English as a foreign language from different mother tongue (L1) backgrounds. The varieties examined in the present study are those from L1 backgrounds of Finnish, German, and Spanish. The analysis covers comparisons of the normalised frequencies of the progressives and of the frequencies of different main verbs used in the progressive across the corpora. In addition, a closer examination of the verb types and their contexts is conducted by means of semantic categorisation.

The results of the analysis show moderate differences in the frequencies of the progressive and the type/token ratios between the corpora. The frequencies were lowest and the type/token ratios the highest among the Finnish students whereas this was in reverse among the Spanish students. These differences may possibly be explained through differences in genre, style and topical elements of the writings or factors relating to the proficiency level of the learners. Similarly, the differences in the distribution of different main verbs, especially, between the German L1 speakers and the other learner populations, may be explained by topic-related, or stylistic factors, such as the narrative style or formality level of the texts. The semantic categorisation of the progressive main verbs revealed, however, certain interesting usage-patterns among the Spanish learner population, such as, non-standard usage in stative contexts with mental verbs, verbs of existence, and aspectual verbs. These results may possibly be explained through L1 transfer, as the Spanish L1 differs from Finnish and German by being an aspect language.

The results corroborate findings from previous research, in that multiple factors, i. e. genre and topic of the text, proficiency level, and L1 transfer, may be seen as influencing the use of the progressive. Some provisional discussion and explanations were provided on which of these variables may be influencing the different aspects considered here, i. e. the frequency of the form, the proportions of different verbs, and the different types of verbs used in combination with the form. Controlling certain variables, such as the thematic elements would be required to understand the influence of other variables, such as proficiency and L1 transfer, more confidently. Regarding the learner's choice between the progressive and the simple aspect, more studies are needed to determine which factors guide this, possibly including more qualitative study on the individual native languages, as well.

Keywords: corpus linguistics, progressive form, learner Englishes, lexical variation, L1 transfer

TIIVISTELMÄ

Jenna Honkanen: A corpus-based study on the use of the progressive form in advanced learner Englishes
Valinnaisten opintojen tutkielma
Tampereen yliopisto
Englannin kielen ja kirjallisuuden syventävät opinnot valinnaisina opintoina
Huhtikuu 2022

Tämä tutkielma tarkastelee englannin kielen progressiivia eli kestomuotoa edistyneiden englannin opiskelijoiden kirjoitetussa kielessä. Progressiivi, eli *be*-apuverbi yhdistettynä pääverbin partisiipin preesensiin (*ing*-muotoon) on usein nähty hankalana muotona englannin kielen oppijoille sen monien merkityksien ja käyttötarkoitusten takia. Tutkimuksissa on havaittu, että muut kuin äidinkieliset puhujat käyttävät muotoa eri tavoin, esim. enemmän tai eri yhteyksissä, kuin äidinkieliset puhujat. Tässä tutkimuksessa pyritään tarkastelemaan muodon käyttöä englannin kielen opiskelijoiden keskuudessa vertailemalla eri äidinkielisiä oppijaryhmiä.

Tutkimuksen aineiston muodostaa International Corpus of Learner English (ICLE) -korpuskokoelman suomen, saksan ja espanjan oppijaryhmien osiot. Korpukset sisältävät englantia ylemmällä keskitasolla ja edistyneellä tasolla puhuvien opiskelijoiden kirjoittamia esseitä. Analyysissä tarkastellaan progressiivin yleisyyttä sekä eri pääverbien käyttöä progressiivimuodossa kussakin osakorpuksessa. Progressiivisissa käytettyjen verbien tyyppejä ja käyttöyhteyksiä tarkastellaan lisäksi tarkemmin semanttisen luokittelun avulla.

Analyysin tulokset osoittavat, että korpusten välillä on kohtalaisia eroja progressiivin yleisyydessä sekä eri pääverbien käyttöasteissa. Suomalaiset opiskelija käyttivät progressiivia vähemmän kuin espanjan äidinkieliset, mutta eri progressiivissa käytettyjen verbien määrä oli suurempi. Nämä erot voidaan mahdollisesti selittää kirjoitusten genre-, tyyli- ja aihepiirieroilla tai opiskelijoiden taitotasoon liittyvillä tekijöillä. Vastaavasti eroja eri pääverbien käytössä, erityisesti saksan äidinkielisten ja muiden oppijaryhmien välillä, voi selittää mm. tyyliseikat kuten tekstien kerronnallinen tyyli tai muodollisuuden taso. Progressiivisten pääverbien semanttinen luokittelu paljasti kuitenkin myös joitakin mielenkiintoisia eroavaisuuksia espanjan kielen oppijaryhmän ja muiden oppijaryhmien välillä. Eräissä verbiluokissa espanjan äidinkieliset puhujat käyttivät progressiivia statiivikonteksteissa, mikä on yleisesti nähty epätyypillisenä käyttönä äidinkielisten keskuudessa. Tätä tulosta voidaan mahdollisesti selittää siirtovaikutuksella (*transfer*) oppijan äidinkielestä, sillä espanja, toisin kuin suomi ja saksa, on aspektikieli, jossa kestomuotoa voidaan helpommin käyttää statiivikonteksteissa.

Tulokset vahvistavat aiempien tutkimusten tuloksia siitä, että useiden tekijöiden, kuten tekstin genren ja aiheen, opiskelijan taitotason sekä äidinkielen siirtovaikutuksen, voidaan nähdä vaikuttavan progressiivin käyttöön. Tutkimuksessa esitettiin pohdintoja näiden tekijöiden mahdollisesta vaikutuksesta tuloksissa esitettyihin eroavaisuuksiin progressiivin yleisyydessä, eri verbien käyttöasteissa sekä eri verbityypeissä. Muuttujien tarkempi kontrollointi, esim. tekstien aiheen rajaus, voisi auttaa tulevaisuudessa muiden tekijöiden, esim. siirtovaikutuksen vaikutuksen luotettavammassa havainnoinnissa. Lisätutkimusta tarvitaan myös, jotta voidaan selvittää, mitkä tekijät, esim. äidinkielen piirteistä mahdollisesti ohjaavat oppijan valintaa progressiivisen ja yksinkertaisen aspektin välillä.

Avainsanat: korpuslingvistiikka, progressiivimuoto, englannin opiskelu, leksikaalinen vaihtelu, siirtovaikutus

TABLE OF CONTENTS

1 INTRODUCTION

The progressive form (i. e. a form of the auxiliary *be* and the present participle, *–ing* form, of the main verb) has long been a topic of interest in linguistics. This is mostly due to the general increase in the use of the form, as well as its newly developed functions in the English language in the past centuries (Leech et al. 2009: 120–2; Meriläinen 2017: 760). The progressive has been studied in different varieties of English, especially in varieties of English as a native language (ENL or Inner Circle), but also increasingly in varieties of English as a second language (ESL or Outer Circle), and as a foreign language (EFL or Expanding Circle). This MA thesis focuses on the latter one, the so-called learner Englishes, and examines the use of the progressive by speakers of English who have learnt the language primarily in formal educational settings.

In this study, an attempt is being made to investigate the use of the progressive in learner Englishes as such, that is, without comparing them to any native varieties of English. This approach draws on the idea of English as a world language in the sense that it does not only belong to those who speak it as a native language, and that the norms of language use should not always be determined by ENL contexts but could be observed as such (Kachru 1985, quoted in Schneider 2007: 14). The goal is to see whether the learner English varieties chosen show clear differences in the frequencies of the progressive and in the lexical patterns with the progressive. The study takes a corpus-driven approach in that no hypotheses are made beforehand, and the closer examination of any lexical or morphosyntactic features is conducted only if seen relevant or interesting based on the data. Furthermore, the study aims to lay ground for future, more comprehensive research on the progressive form in learner Englishes. My research questions are the following:

1. Are there significant differences in the frequencies of the progressive in the different learner English / EFL varieties?

2. Are there differences in the distribution of different progressive main verbs in the learner English varieties?

3. Are there differences in lexical patterns, i. e. the types of main verbs used in the progressive in the learner English varieties?

4. Which variables are possibly influencing the use of the progressive form? The variables considered in this study are the learners' L1 (aspect vs. non-aspect), the task- or genre-specific elements of the essays, and the learners' proficiency level (advanced vs. higher intermediate learners)

2 THEORETICAL FRAMEWORK

2.1 The progressive

This study examines the use of the so-called progressive form. By this I mean the verbal construction of a form of the auxiliary verb *be* combined with the present participle, *-ing* form, of the main verb. Although almost any verb or tense in the English language can theoretically be used in the progressive (Römer 2005: 3), the present study focuses on the following types of progressives:

- present progressives, e. g. *are doing*
- past progressives, e. g. *were doing*
- present perfect progressives, e. g. *has been doing*
- past perfect progressives, e. g. *had been doing*
- modal progressives (i. e. progressives preceded by a modal auxiliary), e. g. *will be doing*
- progressive infinitives, e. g. *to be doing*
- passive progressives, e. g. *is being done*

The progressive form has interested linguists for multiple reasons. The core perspective that a progressive verbal phrase marks, is aspectual, which means that the progressive is used to indicate activity in progress that has a limited duration (Quirk et al. 1985: 197–8, quoted in Meriläinen 2017: 763). This aspectual meaning of the progressive, also called *imperfectivity* (see Chapter 2.1.1), is not uncommon in languages in general, it is, however, unusual among Germanic languages (Kirk 2015: 88). Moreover, studies have shown that the use of the progressive is increasing, both in certain ENL varieties, such as British English and Irish English, and in ESL and EFL varieties, where it has also developed new, extended uses that have attracted the attention of scholars. According to Collins (2009: 118) this development of new forms and uses of the progressive has contributed to its overall growth. Another reason for the increased use may have to do with stylistic factors: as the written norms of the language continuously move closer to those of the spoken, the increased popularity of the form in spoken English may influence the use of the form in the written register, as well (Collins 2009: 118).

2.1.1 The progressive aspect

According to Leech et al. (2009) the term *aspect* as a cognitive category and as a grammatical construction refers to "the manner in which the internal temporal constituency of a situation is represented" (2009: 118). This definition goes back to Comrie (1976: 3). The two main aspects, the *imperfective* and *perfective* aspect, have different characteristics in various languages, including meaning, use and markedness in language. In general, the imperfective aspect views a given situation as ongoing, incomplete whereas the perfective views the situation as complete or as a whole. (Binnick 1991: 136, 153.) Kranich (2010: 26–7) gives the following examples to show how the temporal aspect is different between the imperfective (example 2) and the perfective (example 1):

(1) Paul ran when he noticed me.
(2) Paul was running when he noticed me. (Kranich 2010: 26 –7.)

Furthermore, in prototypical aspect languages, such as Russian, the imperfective aspect (so-called general imperfective) can be seen as rather the default form and the perfective as its marked counterpart (Binnick 1991: 136; Kranich 2010: 30). The progressive is generally understood as a subtype of the imperfective aspect (Kranich 2010: 30). The English language does not have perfective aspect, which means that the binary choice is instead made between the progressive and the non-progressive, the simple aspect, which covers anything that the progressive cannot (Kirk 2015: 91). The progressive thus contrasts the simple aspect, which according to Kirk typically expresses perspectives such as "durativity, punctuality, and stativeness" (ibid.: 87). Kirk (ibid.) gives as examples "*it rains vs it is raining*" and "*John sings vs John is singing*", where the simple and the progressive portray the event differently: the simple forms express something general or stative about the event whereas the progressives indicate that the events are in progress. The simple aspect can, however, also make claims about the completeness of the situation, much like the perfective aspect.

2.1.2 Basic and special functions of the progressive

Leech et al. (2009: 119) consider the progressive form of the English language a subcategory of aspect. According to them, a majority of instances of verbs used in the progressive contain the aspectual notion of incompleteness or being in progress (ibid.). However, while the primary function of the progressive is aspectual, the progressive is nowadays considered acceptable in certain situations with stative verbs and habitual activities, as well, e. g. *Mary's living in a flat in London* (ibid.: 129). Furthermore, the progressive has certain non-aspectual functions, such as the futurate progressive, e. g. *He's coming back tomorrow* (ibid.: 133) and the subjective progressive, e. g. *I imagine you're always battling in school* (ibid.: 134). In addition to these generally accepted, standard uses of the progressive, studies have reported certain extended, non-standard uses of stative and habitual progressives especially in ESL varieties, e. g. *You have you must be having waterfalls?* (Rautionaho 2014: 200).

2.1.3 Lexical aspect: categorising progressive main verbs

Studies have shown that the progressive should not be treated purely as a grammatical phenomenon, independent of lexical patterns (e. g. Römer 2005). Römer's (2005) data on the spoken English by native British English speakers suggest that individual verbs prefer certain contexts or functions, one of those being the progressive vs. non-progressive aspect (ibid.: 169). This has raised the question whether the speakers of English as a foreign language also use the progressive together with verbs that in ENL varieties have been seen as preferring the form and if EFL learners are thus sensitive to these types of lexical patterns.

Traditionally verbs have been divided into verbs that can be used in the progressive and verbs that cannot, where this division seems to correspond with the classes of stative and non-stative or dynamic verbs (e. g. Comrie 1976: 35, quoted in Römer 2005: 116). Stative verbs refer to states (e. g. *be, know*) whereas non-stative or dynamic verbs refer to actions or events

(e. g. *do, come*) (Römer (2005: 116). Since there is generally no element of change or progression with stative verbs, the progressive is rarely used in these contexts (Rautionaho 2014: 29). However, Römer (2005: 116) notes with reference to Comrie (1976: 35–6) that there are contexts in which these lexically stative verbs can appear in the progressive. An example of this are verbs of perception (e. g. *see, hear*), which can on occasion express temporariness or certain emotionality, as was seen in Römer's (2005) data on spoken British English:

(3) You '*re* not *seeing* the real me! I am not just a teacher, I 'm so much more than that! (BNC_spoken) (ibid.: 116)

The simplistic division between stative and non-stative verbs is therefore not enough to understand which verbs are generally used in the progressive. More often considering the entirety of the situation expressed in a sentence is necessary, to determine if the progressive is compatible. As the first step into analysing broader situation types[1] of sentences, studies have often used classification originally from Biber et al. (1999: 360–5, quoted in Collins 2008 and Rautionaho 2014), which categorises verbs into semantic domains depending on their semantic characteristics. This approach was chosen for the present study as it specifically considers the lexical meaning of the verb in a sentence (Rautionaho 2014: 162).

Table 1. Semantic domains of verbs (Collins 2008: 234–6; Rautionaho 2014: 162–4)

Semantic domain	Characteristics	Example verbs
Activity verbs	subject has the role of an (volitional) agent	e. g. *work, come, try*
Communication verbs	activities involving spoken/written communication	e. g. *say, talk, tell*
Mental verbs	cognition, perception, emotion	e. g. *think, see, love*
Verbs of occurrence	events occurring without volitional activity	e. g. *become, get*
Verbs of existence	a. state of existence b. relationship; entity & description	e. g. *live, stay* e. g. *contain, be* (copular)
Causative verbs	person or an inanimate entity brings about a new situation	e. g. *cause, help, permit*
Aspectual verbs	stage of progress of an event or activity	e. g. *start, continue, keep*

[1] A typical classification is Vendler's (1957) model that labels so-called situation types as *activities, accomplishments, achievements,* and *states* (quoted in Shirai 2013: 271). This model is referred to in numerous studies (e. g. Kranich 2010; Shirai 2013; Zeng, Shirai & Chen 2021).

As seen in Table 1, the classification of verbs into semantic domains includes seven classes. The progressive form combines with these classes to varying degrees: for example, activity and communication verbs generally combine easily with the progressive, whereas mental verbs and verbs of existence have more variation within the class regarding which verbs are compatible with the form. The classes are described in more detail together with the analysis in Chapter 4.2.2.

2.2 The progressive in learner Englishes

2.2.1 Defining categories ENL, ESL and EFL

Concerning classifications of different varieties of English, scholars have often referred to the Three Circle model of Kachru (1986), in which the status of English in a particular country is equated with the terms ENL (English as Native Language), ESL (English as Second Language) and EFL (English as Foreign Language) as follows:

- ENL or Inner Circle countries, where English is spoken as the dominant native language (e. g. UK, US)
- ESL or Outer Circle countries, where English has spread as a result of colonialism and serves as an additional official language (e. g. India, Singapore)
- EFL or Expanding Circle countries, where English is learnt as an important foreign language in formal instructional settings (e. g., France, Japan). (Rautionaho, Deshors & Meriläinen 2018: 42.)

With reference to these categories, the use of the progressive has been studied extensively and most often by comparing either one or more ESL or EFL varieties with a native variety (e. g. Virtanen 1997; Axelsson & Hahn 2001; Rautionaho 2014; Meriläinen 2017, see Chapter 2.2.3). In recent years, there have been studies that have brought ENL, ESL and EFL varieties together, e. g. Hundt and Vogel (2011), who examined the frequencies and non-standard uses of the progressive in multiple ESL and EFL varieties comparing them to multiple ENL varieties. These more recent studies have adopted a view in which the varieties (especially ESL and EFL) are not regarded as strict, clear-cut categories but more as a continuum (e. g. Gilquin & Granger 2011; Hundt & Vogel 2011; Meriläinen, Paulasto & Rautionaho 2017; Rautionaho, Deshors &

Meriläinen 2018). As Meriläinen (2017: 780) notes, the variation between different learner populations shows that the so-called learner Englishes do not form a homogenous group within the EFL variety. For example, Gilquin & Granger (2011), who compared learner Englishes of L1 Dutch, French, Spanish and Tswana, found clear differences in these learner populations that they attributed to the degree of exposure to English in the population (i. e. L1 Dutch and Tswana showed similarities, and both had high exposure to English). Therefore, before drawing conclusions about the EFL or learner English variety as a whole, it is important to consider the variation within the group and take different variables, such as the language background and proficiency of the speakers, into account (Hundt & Vogel 2011: 146; Meriläinen 2017: 780).

The learner English varieties chosen for this study consist of English by native speakers of Finnish, German and Spanish. These varieties can all be described as EFL or Expanding Circle varieties since English is spoken among these learner populations mostly as a foreign language taught in educational settings. The following chapter describes certain characteristics of these learner English varieties with special focus laid on the aspectual marking in the L1, native languages, that may be relevant to examination of the use of the progressive. Further background on variables such as the proficiency levels of the learner populations is given in Chapter 3.2.

2.2.2 Characteristics of learner Englishes: the role of transfer from L1?

The progressive form is generally seen as a challenging feature for learners of English, partly because of the "semantic and functional confusion" (Römer 2005: 1) related to it, as well as inadequate pedagogical description in English textbooks (ibid.: 275). In addition, as Römer (2005: 172–3) notes, native speakers of so-called non-aspect languages, such as Finnish and German, that miss the direct counterpart of the progressive often have problems with using the form appropriately, even on a comparatively high level of proficiency.

Both German and Finnish lack the progressive grammatical category, the marker for ongoingness of a situation. Aspect can, however, be expressed by using a temporal adverb, such as *gerade* in German and *juuri* in Finnish, which in both languages carry the meaning 'at the moment'. In addition, there exists exceptional grammatical constructions in both languages, which can be used to express aspect. There have been debates on whether some of these constructions or temporal adverbs (like *gerade* in German) could be seen as systematic aspectual markers. (Dose-Heidelmayer & Götz 2016: 231.) As their usage is, however, optional, it is generally agreed that languages like German and Finnish do not have a functional, morphological marking of aspect and are thus so-called non-aspect languages. Spanish, on the other hand, similarly to English, marks events as ongoing or being in progress by using the auxiliary *estar* ('to be') together with the main verb, which is used in the present participle (e. g. *estoy hablando* 'I am talking') (Zeng, Shirai & Chen 2021: 272). In addition, Spanish distinguishes between imperfective and perfective aspect (see Chapter 2.1.1) in the past tense. Since the progressive in Spanish is seen as a highly productive feature, Spanish is normally considered an aspect language. (ibid.)

There is a general understanding that the learner's L1 may influence the tense-aspect marking of English, there is, however, no clear conclusion as to the extent of the influence (Zeng, Shirai & Chen 2021). Studies have shown conflicting results on comparisons between aspect and non-aspect languages. Some have concluded transfer as a possible explanation for extended uses of the progressive or differing frequencies compared to a native variety, e. g. Meriläinen (2017) who concluded that L1 influence is a likely explanation for extended uses of the progressive in Japanese and Chinese learner Englishes. On the other hand, Zeng, Shirai & Chen (2021: 273) report that some studies (e. g. Kleinmann 1977) have found no significant differences between aspect and non-aspect L1 populations, or that other variables, such as task types and proficiency levels need to be considered, as well (e. g. Sugaya and Shirai 2007).

2.2.3 Previous studies on the progressive form in learner Englishes

As mentioned in the introduction, the progressive form has been studied comprehensively especially in ENL varieties. Among the most extensive studies are those of Römer (2005), who examined the progressive in a book-length corpus-based study on spoken British English, and of Kranich (2010), who examined the diachronic changes of the progressive from various perspectives. Furthermore, the use of the progressive in ESL varieties has continued to attract the attention of scholars, presumably because of new, extended uses of the form that have been discovered in these varieties (e. g. Collins 2008; Rautionaho 2014). Rautionaho (2014) examined the use of the progressive comprehensively in spoken World Englishes, a group that consisted of eight native and second language English varieties. She included various aspects in her study, such as the frequency of the progressive, morphosyntactic features that relate to the progressive, lexical variation in progressive main verbs, as well as the functions of the progressive. As Rautionaho's (2014) study focused on spoken English in ENL and ESL varieties, her results cannot be directly compared with the results of this study. However, because of the extent of her research, it offers a useful model of different perspectives and methodological approaches for studying the progressive. This study partly follows her approach in the examination of progressive main verbs and their categorisation into semantic domains (see Chapter 4.2). Furthermore, Collins (2008) included both written and spoken registers in his study on the progressive in ESL varieties, where he also made comparisons with ENL varieties as reported in Biber et al (1999) and in Smith (2002). Because of the comparisons between registers, his study (ibid.) will be referenced in the analysis when seen appropriate.

Research focusing on the progressive form in learner Englishes has lately gained importance, as well. Learner Englishes are increasingly being equated with ESL varieties both in terms of equal importance in the field of study on linguistic variety and change as well as because of the parallel findings in both varieties concerning certain features, such as the

extended uses of the progressive. Research on the progressive in learner Englishes vary by varieties or L1 populations included, by mode of communication considered, and aspects (e. g. lexical, morphosyntactic features) studied. Some of these various studies are presented in the following.

Early research comparing the use of the progressive in learner Englishes was conducted by Virtanen (1997) and Axelsson & Hahn (2001), for example. Both studies compared two written learner English varieties, Virtanen (1997) L1 Finnish and Swedish, and Axelsson & Hahn (2001) L1 Swedish and German, with native varieties of British and American English. Both Virtanen (1997) and Axelsson & Hahn (2001), who referenced Virtanen's (1997) results, examined the frequencies of the progressives and the distribution of progressive main verbs. The studies reported a slight overuse of the progressive form by foreign language learners in comparison to native speakers of British English, with the highest overuse found among the German L1 speakers. However, Virtanen (1997) also noticed that American English had the highest frequency of the progressive out of all the varieties, which was attributed to informality and choice of register in the writings.

Wulff & Römer (2009) examined the progressive in academic texts by comparing two upper and lower lever corpora by L1 German speakers with two respective native speaker corpora. Their results show that the preferences concerning the choice of aspect preferences relate more to "the writers' mastery of text type-specific conventions" (2009: 1, 130–1) than to general language proficiency. The study by Wulff & Römer is relevant for the present study as it examines the use of progressive in the same mode of communication, written English, and in a similar genre, academic texts, even if it does not consider variation between multiple different learner populations.

Various studies have examined the progressive in spoken English, instead of written, mostly because the progressive is generally seen as typical to the spoken register. For example,

Dose-Heidelmayer & Götz (2016) examined spoken English by German L1 speakers and detected underuse compared to an ENL variety. Zeng, Shirai & Chen (2021), on the other hand, included multiple spoken sub-corpora from the LINDSEI corpus[2] as they examined the progressive in spoken learner English by L1 German, Spanish and Chinese speakers. Special attention was given to the role of transfer in these two studies. Furthermore, comparison between the spoken and written modes of communication was conducted by Meriläinen (2017), who compared Finnish, Swedish, Japanese, and Chinese L1 speakers' written and spoken corpora (ICLE and LINDSEI) with native British and American English varieties. Meriläinen (2017) investigated multiple variables in her study, such as L1-influence, mode of communication (written or spoken) as well as differences in proficiency and exposure to the language.

As established in this chapter, the learner varieties chosen for the present study (L1 Finnish, German and Spanish) have been included in various previous studies. Especially L1 German has been treated as an interesting research subject, possibly because of the lack of grammatical aspect in the language (see Chapter 2.2.2) and the perceived "errors" made by the learners in the use of the form (e. g. Römer 2005; Dose-Heidelmayer & Götz 2016). The studies presented in this chapter will be, when seen relevant, compared with the corpus findings of this study.

[2] Louvain International Database of Spoken English Interlanguage (Gilquin, Gaetanelle, Sylvie De Cock & Sylviane Granger (eds.). 2010. *The Louvain international database of spoken English interlanguage: Handbook and CD-ROM.* Louvain-la-Neuve: Presses universitaires de Louvain.)

3 DATA AND METHODS

3.1 International Corpus of Learner English (ICLE)

The data used in this study comes from the International Corpus of Learner English (ICLE). The ICLE corpus contains written essays by higher intermediate to advanced learners of English as a foreign language from different mother tongue (L1) backgrounds. The essays are mostly argumentative[3], written of varying topics, and on average 605 words in length. The writers of the essays are university undergraduates, most of them in their twenties, who study English in a non-English-speaking country. (Granger et.al 2020: 14.)

There have been two earlier versions of the ICLE: the first version appeared in 2002 and contained essays from 11 mother tongue backgrounds, and the second in 2009, representing 16 language backgrounds. This study makes use of the third and most recent (2020) version of the ICLE, which contain c. 5 million words from 25 different L1 backgrounds. (Granger et.al 2020.) However, it should be noted that the varieties used in this study already appeared in the first version of the ICLE, in year 2002. The time of data collection needs to be considered when discussing certain factors, such as the social, educational, and cultural background or learning contexts of the learners in different countries, as some of these factors may be influenced by changes occurred during the last 20 years (such as the increased use of the Internet).

[3] In the Spanish sub-corpus, the proportion of literary essays (instead of argumentative) is somewhat higher than in the other two: 20.72%. In the Finnish corpus the respective proportion was 8.46% and in the German 3.43%.

3.2 Varieties included in the present study

The varieties chosen for this study are learner Englishes from the following L1 backgrounds: Finnish, German, and Spanish. The main reason for choosing these varieties was to include different kinds of languages, i. e. languages from different language families. I selected Finnish, a Uralic language, and two Indo-European languages, one of which is a Germanic language and one a Romance language. This decision was motivated by the interest in possible influences by L1 transfer (see Chapter 2.2.2). In addition to this, some practical and subjective reasons influenced the selection of the varieties, such as the corpus used (i. e. selection was made from the varieties that are represented in the ICLEv3) and the researcher's knowledge of the L1s and/or countries in question, which was seen as relevant for considering L1 transfer and any further variables possibly influencing the results.

Although the ICLE corpus is ideally meant to contain data from advanced learners, the external criteria used to define "advanced", i. e. "university undergraduates in English" (Granger et al. 2020: 11) do not automatically ensure a standard proficiency level due to factors, such as the learning context in each country[4]. Therefore, the proficiency levels of the essays vary both across and within the sub-corpora (ibid.). To gain at least a tentative idea of the differences in proficiency across the corpora, Granger et al. (2020), i. e. the compilers of the ICLE, submitted a random sample of 20 essays from each corpus to be rated on the basis of the Common European Framework of Reference for Languages (CEFR). The CEFR results for the samples of the relevant corpora are presented in Table 2 below.

[4] For example, lesser exposure to English due to dubbing of movies and TV programmes is generally seen as one variable affecting the proficiency level of an individual learner population, especially when keeping in mind that the ICLE data consist of essays written pre-2002, i. e. during time when the increased use of Internet had not evened the possible differences between the learner Englishes. Concerning the varieties included in the present study, it should be noted that in Finland, TV programmes are generally not dubbed but subtitled, whereas in Germany and Spain this has traditionally been in reverse.

Table 2. CEFR results – 20 essays per sub-corpus (Granger et al. 2020: 12)

MOTHER TONGUE	B2 (AND LOWER)	C1	C2	TOTAL
Finnish	3	8	9	20
German	1	12	7	20
Spanish	12	8	0	20
ICLEv3	194	217	89	500

As reported by Granger et al. (2020: 11), the proficiency level of the essays ranges from higher intermediate (B2) to advanced (C1 or C2). In Table 2, we can see that the proficiency level of learners from L1 Spanish seems to be centred on the lower end of the range (B2), whereas L1 Finnish group is on the higher end. These differences in proficiency levels need to be considered in the analysis.

The new interface of the ICLEv3 enables the controlling of certain variables in selecting the essays to be included in the study. Such variables are the country, institution, and language background of the students. As the individual sub-corpora are rather small, most of the variables were not controlled so as not to limit the data too much. However, to be able to draw comparisons between the corpora, the following restrictions were applied. The sub-corpora are from now on referred to as FICLE (Finnish component of ICLE), GICLE (German component) and SICLE (Spanish component).

Table 3. Learner variables in each sub-corpus

Learner variables / Corpus	Native language	Country	1st language at home	2nd language at home	University (cities)	Texts	Words
FICLE	Finnish	Finland	Finnish	none	Helsinki, Joensuu, Jyväskylä, Oulu and Turku	218	165,111
GICLE	German	Germany	German	none	Dresden and Augsburg	270	136,395
SICLE	Spanish	Spain	Spanish	none	Madrid, Alcala	236	186,594

Controlled variables included the native language, home country, 1st language spoken at home, 2nd language spoken at home, and the educational institution. Of the German part of the ICLE, only the essays of writers from Germany and in universities in Germany (and not in Austria, for example) were included. Similarly, of the Finnish sub-corpus, the university Åbo Akademi was excluded as the students in this university mostly study in the Swedish language. As one

of the variables considered in this study relates to the different L1s of the learners, the language background of the learners was controlled, in that only students within each learner population were included whose 1st language at home is the L1 in question and who have reported no 2nd language used at home, i. e. bilinguals were excluded from the study.

As mentioned above, the essays in the ICLE corpora are mostly argumentative, however, among the Spanish students' essays there are some more variation. as around 20% of the essays were written on literary topics. It was not possible to control the essay topics included in the corpora, as it would have limited the size of the data too much. However, the possible variation resulting from the genre differences will be considered in the analysis of the results.

3.3 Obtaining data

The ICLEv3 is a part-of-speech (POS) tagged corpus, which makes it fairly easy to retrieve different types of progressives from the data. The POS-tag Vbe (any form of the verb *be*) combined with Ving, i. e. the tags VVG / VHG / VD / VBG[5], allowed the retrieval of following types of progressives:

- present progressives, e. g. *are doing*
- past progressives, e. g. *were doing*
- present perfect progressives, e. g. *has been doing*
- past perfect progressives, e. g. *had been doing*
- modal progressives (i. e. progressives preceded by a modal auxiliary), e. g. *will be doing*
- progressive infinitives, e. g. *to be doing*
- passive progressives, e. g. *is being done*

[5] VVG (*-ing* participle of a lexical verb), VHG (*-ing* participle of *have*), VD (*-ing* participle of *do*), VBG (*-ing* participle of *be*) (Granger et al. 2020: 71).

In addition, the "any word" function of the ICLEv3 was used to retrieve instances where one or two words appear between the form of *be* and the present participle[6]. Table 4 shows the full search string that was used to retrieve the progressives.

Table 4. Search strings used in ICLEv3 to retrieve progressives and example instances

	Search string				
context	Vbe (= any form of the verb *be*)	any word	any word	Ving VVG / VHG / VD / VBG (= Vlex *-ing* / Vhave *-ing* / Vdo *-ing* / Vbe *-ing*)	type of progressive retrieved
	am, 'm, are, 're, is, 's			Ving	present progressive
	was, were			Ving	past progressive
e. g. has, had	been			Ving	perfect progressive
e. g. can, could, may, might, will	be			Ving	modal aux. + progressive
to	be			Ving	progressive infinitives
	Vbe			Vbe *-ing*	passive progressives

Corpus examples

GICLE	I've	been			making	music now for 20 years.
SICLE	He	is	always		thinking	in a superficial way.
FICLE	[…] it seems to me that it	is	not	only	becoming	but already is a world language.

It should be noted that fragmentary forms of progressives, i. e. progressives, where the auxiliary *be* is missing[7] (see Rautionaho 2014) were excluded in the present study. Fragmentary progressives could not be retrieved using the search string above since the tag Vbe was used as part of the search string. Including these forms would have required laborious manual checking

[6] During the retrieval process, the search string was tested with three or more words between Vbe and Ving. However, as this produced hardly any instances of actual progressives but would have resulted in extensive manual checking of irrelevant instances, the search string was limited to two words between Vbe and Ving.

[7] an example for a fragmentary progressive as presented in Rautionaho (2014: 134): "(6.28) You just *trying* to wind me up (ICE-IRL, S1A-025)".

of irrelevant hits since the search would have been expanded to include all present participle forms of verbs and not only those in the vicinity of a form of the verb *be*. The fragmentary forms were excluded from this study as it was seen unlikely that these forms would appear in the written English corpora under examination here.

3.4 Methods of analysis

After retracting the data from the ICLE corpus, the data was copied to an Excel worksheet, where the occurrences of the progressive were manually checked. Certain restrictions were applied to the instances included in the analysis following Rautionaho (2014: 74–5). The following types of instances were excluded from the data:

- structure *be going to* with future reference
  (4) But this "attention" will be what is going to provoque the Leonte's jealousy (SICLE, SPAL1009)
- gerunds
  (5) A part of that fight is demanding free contraception and abortion (FICLE, FIJO3001)
- appositively used participles
  (6) Thas's what life is worth living for (GICLE, GEAU3058)

The POS-tagging of the ICLEv3 allowed the exclusion of some of these unwanted instances, for example, most instances of the structure *be going to*[8]. Some further irrelevant instances were removed during the manual checking of the data, such as some incorrectly tagged adjectives and nouns ending in *-ing*, e. g. *The results are really discouraging* (SICLE, SPM02007). Unlike in Rautionaho (2014), progressive infinitives were included in the present study as they were seen relevant based on other studies, e. g. Collins (2008).

As a starting point for the analysis, frequencies of the progressive forms were calculated for each sub-corpus (Chapter 4.1). In addition to calculating the normalised frequency, i. e. comparing the number of progressives with the total number of words in each corpus, as done

[8] The structure *be going to* was excluded from the search by excluding the tag VVGK (*-Ing* participle catenative (*going* in 'be going to')) from the search string (see Granger et al. 2020: 71).

most often in previous research, I calculated the frequencies by comparing the number of progressives to the non-progressive forms in the corpora. As the progressive form can only occur with verbs and not with adverbs or nouns, for example, it can be argued that the comparison should be made specifically with the number of verb phrases in the corpus[9] (e. g. Rautionaho (2014: 76). However, since most previous studies utilise the normalised frequency where the number of progressives is related to the number of words[10], both frequencies were calculated to allow comparisons with previous studies.

The first part of the analysis of progressive main verbs (Chapter 4.2.1) concentrates on the frequencies of different main verbs in the corpora. First, a so-called type/token ratio was calculated to see how many different main verbs are used in each corpus compared to the number of progressive occurrences. After this the focus was shifted to the most common main verbs used in the progressive: the distributions of these verbs were compared across the three corpora concentrating on the verbs that appear frequently in all three corpora as well as verbs that show variations in their distributions.

The second part of the analysis (Chapter 4.2.2) concerns the semantic domains of the most common main verbs. The classification by Biber et al. (1999) as presented in Collins (2008) and Rautionaho (2014) (see Chapter 2.1.3) was applied to the most common main verbs (more than 1% of the progressives) found in the three corpora. Along with the categorisations of the verbs into semantic domains, the differences in the distribution of these verbs were discussed in more detail and with examples drawn from the data. It should be noted that this part of the analysis involved a degree of subjective interpretation as the progressive main verbs were manually categorised into the semantic domains, albeit the categorisation followed the model introduced by Biber et al. (1999, quoted in Collins 2008 and Rautionaho 2014).

[9] so-called V-coefficient, where V = NPROG / NVERB x 10,000 (Rautionaho 2014: 77)

[10] so-called M-coefficient, where M = NPROG / NWORD x 100,000 (Rautionaho 2014: 77)

4 RESULTS

4.1 Frequencies of the progressive

The frequencies of the progressive form in each sub-corpus were examined as a general starting point for the closer examination of the different main verbs that appear in the progressive.

Table 5. Frequencies of the progressive form in the corpora

	Number of words in corpus (Nwords)	Number of verb phrases[11] (Nvp)	Number of progressives (Nprog)	M (Nprog / Nwords x1,000)	V (Nprog / Nvp x100)
FICLE	165,111	32,546	413	2.50	1.27
GICLE	136,395	25,993	426	3.12	1.64
SICLE	186,594	35,726	575	3.08	1.61
Total	488,100	94,265	1,414	2.90	1.50

The left side columns of Table 5 show the number of words and the number of verb phrases found in each corpus. The frequencies of the progressive were first calculated by dividing the total number of progressives (middle column) by the total number of words in the corpus and multiplying it by a thousand: M = Nprog / Nwords x1,000. This calculation method is most often applied in corpus linguistics, and was used here, as well, so as to be able to compare the results with those reported in previous studies. As the progressive can only occur as part of a verb phrase, another frequency for the progressive was calculated by comparing the number of progressives with the number of verb phrases in each corpus: V = Nprog / Nvp x100. As seen on the right side columns of Table 5, both calculation methods, M and V, produced similar results: the frequency of the progressive was lowest in FICLE, whereas the frequencies in GICLE and SICLE are rather identical.

The frequencies of the progressive in the present data correspond to those reported by Axelsson & Hahn (2001: 11) in that, according to their research, the frequency of the

[11] any Vlex, Vbe, Vhave or Vdo

progressives per 1,000 words was higher (3.6) in the German learners' essays compared to the frequencies in the British (2.0) and the Finnish students' essays (2.3) that Virtanen (1997) had reported. Axelsson & Hahn (2001) thus concluded that there is a slight overuse of the progressive by the German native speakers compared to the British English native speakers. As seen in Table 5, the Spanish students' essays compared closely with the German students' ones regarding the frequency of the progressive in the present data. The Finnish students, on the other hand, have generally been comparable to native speakers concerning the frequency of the form (e. g. Virtanen 1997; Axelsson & Hahn 2001). Meriläinen (2017), however, found that the progressive was used even less frequently in the Finnish students' written corpora than in the BrE reference corpus (LOCNESS), which, using same terms, could be seen as "underuse" of the form. The differences between the varieties are, however, smaller in the present data than in the previous studies.

Furthermore, Virtanen (1997) found that the normalized frequency of the progressive was considerably higher in the AmE variety included in her study than in the learner corpora and in the BrE variety. The higher frequency of the form was attributed to informality and choice of register in the writings since the progressive is generally seen as more common in informal or spoken registers. Therefore, examining the level of informality or the genre of the texts seems necessary before making any conclusions about proficiency or L1 transfer as reasons for the differences in frequencies.

4.2 Progressive main verbs

This section examines the distribution of different verbs that are used in the progressive form in the corpora. This part of the analysis only concerns progressive main verbs: passive progressives, where the verb *be* is used as an auxiliary in the progressive, are excluded from the analysis.

4.2.1 Frequencies of different progressive main verbs

Altogether 374 different main verbs were used in the progressive across the three corpora. Table 6 presents the number of different progressive main verbs (Types) found in each corpus, which are compared to the total numbers of the progressives[12] (Tokens). This produces a type/token ratio for each corpus: a higher ratio means that more different verbs were used in the progressive form.

Table 6. Type/token ratios of FICLE, GICLE and SICLE

	Types	Tokens	Type/Token ratio
FICLE	167	381	43.8
GICLE	179	421	42.5
SICLE	190	539	35.3
total	374	1,341	27.9

As seen in Table 6, the type/token ratio is highest in FICLE (43.8), that is, the Finnish students had the most variety in the main verbs used in the progressive form. The type/token ratios are, however, nearly identical between FICLE and GICLE (42.5). The Spanish students, on the other hand, had less variety in the use of different main verbs in the progressive as the type/token ratio in the SICLE corpus is somewhat lower (35.3). Axelsson & Hahn (2001: 14) note that a higher type/token ratio could indicate a higher level of language proficiency. They found in their study that the German and Swedish learners of English had lower type/token ratios than

[12] The number of tokens presented in Table 6 for each sub-corpus differ from the number of the progressives given in Table 5 because the passive progressives were subtracted from the total number.

the Finnish and Finland-Swedish learners and native British and American English speakers studied by Virtanen (1997). The present study presents a contrasting view as the ratios were almost identical between FICLE and GICLE. The lower ratio in SICLE could, however, be attributed to the lower proficiency level of the Spanish learners as reported by Granger et al. (2020: 12; see Chapter 3.2, Table 2). The repetition of a limited number of verbs could according to Axelsson & Hahn (2001: 14) also indicate greater informality of the text, since the essays by AmE speakers in Virtanen's (1997) study had clearly lower type/token ratio compared to the other native speaker group (BrE). The differences between the type/token ratios in the present study are, however, much smaller than those that Axelsson & Hahn (ibid.) reported, and it is therefore difficult to draw any definite conclusions.

Out of the 374 progressive main verbs (Types) found in the corpora, 217 were used only once and 114 from two to five times. Because of the small size of the corpora, this study concentrates on the more frequently used verbs as this establishes better grounds for conclusions. The most frequently used verbs (more than 1% of all occurrences) across all three corpora are presented below. These 22 verbs comprise nearly half of the instances of the progressive form (46.46%) in the corpora.

Table 7. The most frequent progressive main verbs in the present learner English varieties

Verb	Occ.	%
do	60	4.47
try	52	3.88
go	46	3.43
talk	45	3.36
get	41	3.06
look	36	2.68
live	31	2.31
become	29	2.16
work	29	2.16
change	26	1.94
watch	24	1.79
study	23	1.72
happen	20	1.49
say	20	1.49
sit	20	1.49
fight	19	1.42
think	19	1.42
have	18	1.34
take	18	1.34
make	17	1.27
wait	16	1.19
refer	14	1.04
total	623	46.46

The most frequently occurring verbs in the learner English varieties examined in this study are *do, try, go, talk,* and *get*. These results are not surprising as they are consistent with those of Virtanen (1997) and Axelsson & Hahn (2001). Similarly, Rautionaho (2014) found that *go* and *do* are the most common verbs occurring in the progressive in spoken World Englishes, while *talk*, *get,* and *try* appear among the eight most frequently used main verbs, as well. Furthermore, Römer (2005: 114) lists *go, talk, try,* and *do* among the top 9 and *get* as the 18th most frequent main verb in spoken British English. The results indicate consistency in the use of these main verbs in the progressive across ENL, ESL, and EFL and across spoken and written registers.

While it is interesting to examine learner Englishes as a group, this study mainly focuses on the variation between the learner populations of L1 Finnish, German and Spanish. The most frequent progressive main verbs presented in Table 7 were distributed differently in the three sub-corpora. The following presents an analysis on the use of these main verbs in each individual corpus. Table 8 lists the most frequent progressive main verbs (more than 1% of all

occurrences) in descending order for each corpus (FICLE, GICLE and SICLE), and indicates the number of occurrences (Occ.) and the proportion out of all the progressives in the respective corpus (%) for each individual verb. At the bottom of the table, the total proportions of these verbs out of all the progressives in each corpus are given. The proportion is highest with SICLE, which aligns well with the slightly smaller type/token ratio of the corpus reported above.

Table 8. The most common main verbs in the progressive in FICLE, GICLE and SICLE

FICLE			GICLE			SICLE		
Verb	Occ.	%	Verb	Occ.	%	Verb	Occ.	%
do	18	4.72	sit	20	4.75	do	30	5.57
get	17	4.46	talk	19	4.51	try	26	4.82
go	16	4.20	go	18	4.28	talk	22	4.08
become	14	3.67	get	16	3.80	change	16	2.97
try	14	3.67	look	14	3.33	look	14	2.60
work	12	3.15	do	12	2.85	watch	14	2.60
happen	10	2.62	try	12	2.85	live	13	2.41
live	10	2.62	live	8	1.90	refer	13	2.41
change	9	2.36	watch	8	1.90	go	12	2.23
make	9	2.36	work	8	1.90	study	12	2.23
look	8	2.10	come	6	1.43	become	11	2.04
say	8	2.10	stand	6	1.43	have	10	1.86
fight	7	1.84	think	6	1.43	think	9	1.67
study	7	1.84	wait	6	1.43	work	9	1.67
take	7	1.84	become	5	1.19	be	8	1.48
face	4	1.05	deal	5	1.19	die	8	1.48
increase	4	1.05	fight	5	1.19	get	8	1.48
miss	4	1.05	make	5	1.19	wait	8	1.48
raise	4	1.05	ride	5	1.19	fight	8	1.48
talk	4	1.05	say	5	1.19	happen	7	1.30
think	4	1.05	wear	5	1.19	say	7	1.30
use	4	1.05				take	7	1.30
						begin	7	1.30
						give	6	1.11
						play	6	1.11
						see	6	1.11
total	194	50.92		194	46.08		297	55.10

Table 8 shows that there are differences between the sub-corpora in the order of the most frequently used verbs. To better illustrate these differences, any verbs that marked more than 2% of the progressives in one or more of the three corpora (marked with blue in Table 8) were included in Table 9 and compared with the proportions in the other two corpora.

Table 9. Comparison of the frequencies of the most common progressive main verbs between FICLE, GICLE, and SICLE

	FICLE		GICLE		SICLE	
	Occ.	%	Occ.	%	Occ.	%
become	14	3.7	5	1.2	10	1.8
change	9	2.4	1	0.2	16	3.0
do	18	4.7	12	2.9	30	5.5
fight	7	1.8	5	1.2	7	1.3
get	17	4.5	16	3.8	8	1.5
go	16	4.2	18	4.3	12	2.2
happen	10	2.6	3	0.7	7	1.3
have	3	0.8	4	1.0	10	1.8
live	10	2.6	8	1.9	13	2.4
look	8	2.1	14	3.3	14	2.6
make	9	2.4	5	1.2	3	0.6
refer	1	0.3	0	0.0	13	2.4
say	8	2.1	5	1.2	7	1.3
sit	0	0.0	20	4.8	0	0.0
study	7	1.8	4	1.0	12	2.2
take	7	1.8	3	0.7	7	1.3
talk	4	1.0	19	4.5	22	4.1
think	4	1.0	6	1.4	9	1.7
try	14	3.7	12	2.9	26	4.8
watch	2	0.5	8	1.9	14	2.6
work	12	3.1	8	1.9	9	1.7

Main verbs that comprise more than 2% of the progressive main verbs in all three corpora are *do*, *go*, *look*, and *try* (marked blue in Table 9). Aside from the verbs that are common in all three varieties, there are verbs that are used significantly more frequently (i. e. with more than 2% difference in the normalised frequencies) in one or two of the varieties compared to the others (yellow in Table 9). Notable differences are found in the frequencies of *change*, which is common in FICLE (2.4%) and SICLE (3.0%) but not in GICLE (0.2%); in *sit*, which was the most common verb in GICLE (4.8%) but did not occur in FICLE and SICLE; and in *talk*, which was used frequently in GICLE (4.5%) and SICLE (4.1%) but less so in FICLE (1.0%).

The following chapter examines the most common progressive main verbs presented in Table 8 (see above) more closely by means of analysis of the semantic domains of these main verbs. The differences in the frequencies of some of these verbs, as illustrated in Table 9, are brought into the analysis, and discussed in more detail.

4.2.2 Semantic domains of the most common progressive main verbs

In this chapter I will categorise the most common progressive main verbs, i. e. main verbs that comprised more than 1% of the occurrences in each sub-corpus (see Table 8 above), according to their semantic characteristics. The analysis was limited to the most common verbs in the data because, due to the small size of the corpora, it did not seem worthwhile to investigate the verbs that appeared in the progressive less than 1%, i. e. less than 4–6 times depending on the corpus. The categorisation follows the seven-class taxonomy by Biber (1999), as presented in Rautionaho (2014) and Collins (2008), which classifies verbs into semantic domains (see Chapter 2.1.3). Each of the occurences of the most common main verbs were individually categorised into these domains. As noted in previous studies (Collins 2008 and Rautionaho 2014) certain verbs present a challenge for the categorisation, as they have multiple meanings. A good example is the verb *change* which can be used as an activity verb, as in *Luckily they are changing this system with Clinton 's new health-care program* (FICLE, FIJY1025) and as a verb of occurrence, as in *During the history, this situation has been changing though slowly* (SICLE, SPM05001). Therefore, all instances of the verb could not automatically be categorised but required manual checking.

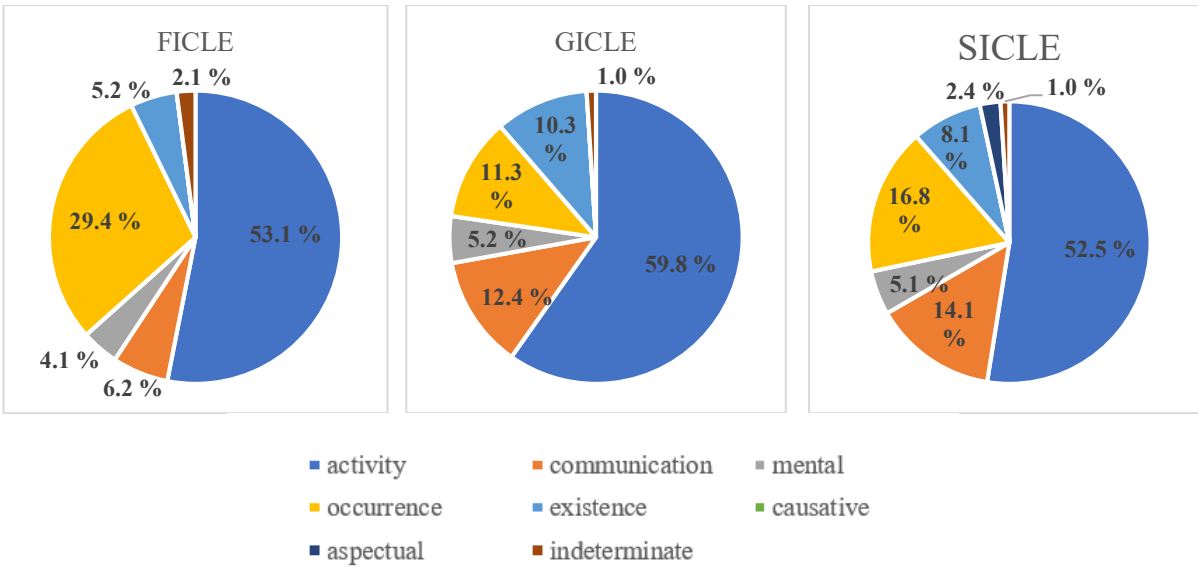

Figure 1. Semantic domains of the most common progressive main verbs in FICLE, GICLE and SICLE[13]

i. Activity verbs

As seen in Figure 1, more than 50% of the most common progressive main verbs in each corpus are classified as activity verbs. Activity verbs are controlled by a subject that has the role of an agent (Rautionaho 2014: 162). Typical activity verbs are verbs like *come, go, try*.

(7) Their wives *will be coming* home soon, worn out and edgy after an exhausting day at the banks […] (GICLE, GEAU3093)

The high ratio of activity verbs is not surprising as the progressive form generally prefers durative and dynamic verbs, as stated in Chapter 2.1.3. Especially verbs like *do, go, and try*, which were among the most common in all sub-corpora (see Table 9), contributed to this high frequency.

Some interesting differences were found between the corpora in the use of certain activity verbs, especially that of *sit*. As seen in Table 8, *sit* is the most common verb used in the progressive in GICLE, as it appears 20 times (4.8% of the progressives) in the corpus. In FICLE

[13] Causative verbs did not occur in any of the sub-corpora and are therefore not visible in the pie chart.

and SICLE, however, there are no instances of *sit* in the progressive. An example of the use of *sit* in GICLE is presented in (8) below.

(8) I can still very well remember the last package holiday I went on. We *were* all *sitting* in a large coach rolling at a constant speed on noisy and busy motorways towards our destination - Venice. (GICLE, GEAU3054)

As demonstrated in (8), the verb *sit* often appears in the past progressive in GICLE and is used to describe the situation that the speaker finds (or found) themselves in. Römer (2005: 104) names these contexts as *framing*, where a new interrupting event happens while something else is in progress. These events are mostly set in the past and the progressive is used to frame the new interrupting event (ibid.). Römer (ibid.) found these types of progressives to be rare in spoken BrE, however, as the present data concerns written essays, the differing result is not that surprising.

While all three corpora consist of mostly argumentative essays, in GICLE, the topics of the essays seem to include more narrative elements than the ones in FICLE and SICLE. Some essays seem to concern the writer's personal experiences or events in the past or even descriptions of a personal idol, for example. The narrative elements of the essay topics could therefore in part explain the frequent use of the verb *sit* in the progressive and especially in these framing contexts in the past tense. The somewhat higher proportion of activity verbs in general in the GICLE corpus (59.8%, see Figure 1) can possibly be explained by this, as well: personal narrative style may induce the use of concrete, physical verbs. However, as the semantic analysis only considers the most common verbs used in the corpora (approximately half of the occurrences) such small differences should not justify any definite conclusions.

ii. Verbs of occurrence

The second highest frequencies in FICLE (29.4%) and SICLE (16.8%) were found with verbs of occurrence, that is, verbs that report an event without volitional activity (Collins 2008: 236;

Rautionaho 2014: 163). Collins (2008: 236) states that these verbs, e. g. *become, change*, are easily compatible with the progressive, because of their durative nature.

(9) Nowadays, finding a job is not that casy and lt *is becoming* more and more difficult even if you have recelved an universlty education. (SICLE, SPM10006)

Notable is especially the high proportion of these verbs in the Finnish speakers' essays (29.4 %), compared to the German speakers' relatively low proportion (11.3 %). The higher proportions in FICLE and SICLE correspond to the results by Collins (2008: 239), in that, in his data, verbs of occurrence were the group with the second highest proportion in the written register in ESL varieties, whereas in the spoken register, both communication and mental verbs had higher proportions. In his data, the proportion of verbs of occurrences was, however, much lower (16.2%) than in the present data with the Finnish students.

Looking at the distribution of different main verbs, some lexical variation can be seen between the three corpora. As seen in Table 9 above, typical verbs of occurrence, such as *become* and *change*, were common in FICLE (with frequencies of 3.7 and 2.4) and SICLE (2.0 and 3.0), but less so in GICLE (1.2 and 0.2). Certain lexical patterns emerge in combination with *become*, especially in FICLE and SICLE, one of which is the use of the adverb construction, *more and more,* as seen in example 9 above. It should be noted, that even though the frequency of *become* in GICLE was lower than in the other two corpora, there are still five instances found and in those instances the verb is used rather similarly as in FICLE and SICLE (in combination with the adverb *more and more,* for example), which shows that native speakers of Finnish, German and Spanish are combining similar structures and elements with the progressive.

On the other hand, the verb *get* is more frequently used in FICLE (4.5) and GICLE (3.8) whereas the frequency is lower in SICLE (1.5).[14] In FICLE *get* is used most often as a verb of occurrence and thus rather similarly as the verb *become*:

(10) Even in the factories everything *is getting* **more and more** computerized […] (FICLE, FIJO3027)

In GICLE, *get* is frequently used as a verb of occurrence, as well, but the contexts are somewhat different:

(11) I don't want to know how my teeth *are getting* as white as snow or how my hair *is getting* silky (GICLE, GEAU1085)

In GICLE *get* is mostly used to refer to physical changes and, for example, never in combination with *more and more.* The difference is notable since in SICLE, where *get* is used less frequently, the contexts are still similar to the ones found with *get* in FICLE (as in example 10) and the ones with *become* in both corpora (as in example 9). One explanation for the choice of the verb *become* in FICLE and SICLE may be its possibly higher formality, which fits the genre of an academic essay. The verb *get* is generally seen as more informal. The frequent use of *get* and the different contexts, relating to more physical events, in GICLE may relate to the essays possibly including somewhat more informal elements or elements typical to spoken language than those in FICLE and SICLE. The smaller frequency of *become* and higher frequency of *get* could therefore rather concern genre-related issues than a lower proficiency among the L1 German speakers. This will be further discussed in the following when considering communication verbs like *talk* and *say*.

[14] Instances where *get* was clearly used in the so-called *get*-passive structure were categorised as indeterminate, e. g. *people stay in this situation without being able to change it and they are getting destroyed by it* (GICLE, GEDR1020).

iii. Communication verbs

Verbs of communication are a sub-category of activity verbs that involve an element of communication (Rautionaho 2014: 162). Included are verbs like *talk, say* and *write*. Communication verbs were frequent among the most common progressive verbs in GICLE (with proportion of 12.4%) and SICLE (14.1%). The relatively low frequency in FICLE (6.2%) was somewhat surprising compared to the other two corpora but also compared to previous studies on different varieties of English. For example, Collins (2008: 237) found, comparing the results by Biber et al. (1999) and Smith (2002), that communication verbs had a proportion between 13 and 17% in progressive main verbs in both ENL and ESL varieties across written and spoken register. The proportions of communication verbs in GICLE and SICLE are thus more consistent with the ENL and ESL varieties reported in Collins (2008) than that in FICLE. Collins's (ibid.: 239) data show, however, differences between the ratios in written and spoken registers, in that, communication verbs were used proportionally more frequently in speech whereas the second highest class in writing were verbs of occurrence. Similar pattern was seen in Smith's (2002) data on written British English as reported by Collins (2008: 237). This, in contrast, is consistent with the frequencies in the FICLE corpus and suggests similarities between FICLE and the ESL and ENL varieties examined in Collins's (ibid.) study. Therefore, if higher proportions of communication verbs are seen as a feature of spoken register whereas higher proportions of occurrence verbs that of written register, this suggests that the proportions seen in FICLE (see Figure 1) correspond better to written register whereas those in GICLE to spoken register.

Looking at the most frequent communicative verbs in the corpora (*talk* and *say* in FICLE and GICLE; *talk*, *say* and *refer* in SICLE), clear difference can be seen in their frequencies. A notable difference was found in the use of the verb *talk*, which occurred more frequently in GICLE (4.5%) and SICLE (4.1%) than in FICLE (1.0%). At a closer look at the instances, it is

evident that the verb is often used when the speaker is referring to themselves either with the pronoun *I* or *we*, as in examples (12) and (13) below.

(12) We *have been talking* about how television has influed in children, but we do n't have to forget adults (SICLE, SPM05006)

(13) We *are* not *talking* about hitting and slapping the children in their faces. We *are* not *talking* about causing severe injuries with almost torture-like instruments such as canes or hard wooden rulers. We *are* not *talking* about teachers who indulge their choleric temper by literally beating up a pupil who was only exchanging a few words with his neighbour. What we *are talking* about is making use of mild forms of corporal punishment in cases in which verbal chastisement has ceased to be effective. (GICLE, GEAU3007)

Examples (12) and (13) seem to be parts of clearly argumentative essays, where the verb *talk* is used presumably as a stylistic or emphasizing method. In the case of example (13), the writer uses the expression *we are (not) talking* even four times within a couple of sentences. The fact that *talk* appears multiple times in the same writer's text, could slightly diminish the role of the verb within the most common verbs used in GICLE. However, if only different writers' essays were considered, *talk* appears in 16 essays, which still leaves the verb among the most common progressive verbs in the corpus.

One possible explanation for the differences in the frequencies of *talk* could relate to the formality or informality of the texts and to the fact that in academic writing, the use of personal pronouns, especially that of *I*, is often discouraged. Even as the essays in the corpora were not exactly academic in nature and should not necessarily have required references to literature, for example, it is possible that the conventions of academic writing or of the written essay genre in general may have had a limiting effect on the use of *talk* in FICLE. Rautionaho's (2014) results could be seen as supporting this view, as the verbs *say* and *talk* were the third and fourth most common main verb used in the progressive in spoken World Englishes, which shows that these verbs are not uncommon in the progressive in general and that the frequencies of *talk* in GICLE and SICLE correspond to the use of the progressive in spoken World Englishes. It may suggest that the essays in GICLE and SICLE may be more informal and in some ways closer to spoken language compared to those in FICLE. On the other hand, when looking at the verb *say,* we can

see that it appears more frequently (2.1%) in FICLE than in GICLE (1.2%) and in SICLE (1.3%). evident that also the Finnish students use *say* in a similar manner as the verb *talk* is used in GICLE and FICLE, as shown in example (14):

(14) Of course, I am not saying that there are n't also those who are sincerely concerned of nature itself (FICLE, FIHE1002)

This suggests that also Finnish students use this type of self-reference or meta-discoursal elements in their essays, although possibly somewhat less than the German and Spanish students, and with different lexical preferences.

iv. Verbs of existence

Verbs of existence are divided into two sub-classes: verbs that express a state of existence or stance, e. g. *live*, *stay*, and verbs that concern relationships between two entities, e. g. *contain*, or between an entity and its description, e. g. *be, seem* (Collins 2008: 236). According to Collins (ibid.) the first group is generally compatible with the progressive, which gives the event the meaning of temporariness.

(15) I came home one evening to the Paris hall of residence where I *was living* during the year I spent studying in France (GICLE, GEAU3034)

Contrarily, verbs of existence that are used to express relationships between entities and descriptions, so-called copular verbs (Rautionaho 2014: 163), are rare with the progressive. These verbs are mostly used in stative situations (see Chapter 2.1.3) and do not readily combine with the progressive.

(16) This system *is being* the subject of a lot of argument in Spain at the moment, since it remains […] (SICLE, SPM04026)

In the present data, verbs of existence were used somewhat more frequently in GICLE (10.3%) and SICLE (8.1%) than in FICLE (5.2%). The verb *live* appeared in each corpus and as the verb is generally easily compatible with the progressive, these instances were not examined more closely. *Live* was also the only existence verb found in FICLE. In GICLE, instances of *wear*

and *look* were found in addition to those of *live,* which increased the overall frequency of existence verbs. Due to the narrative elements in the German essays, occurrences of these verbs in the progressive were not surprising as they were often used in decsriptions of events and scenes. Finally, in the SICLE corpus, the verbs of existence found included the verbs *be, have* and *live*. Among these verbs in the SICLE corpus stative, extended uses were found, such as with the verb *be* in example 16 above and with *have* below:

(17) Do you really think speaking hastly this kind of people *should be* rehabilitated and *having* a second chance? (SICLE, SPM07025)

The differences between the use of the verbs *be* and *have* are seen in the distribution of these verbs between the three corpora, as well. As seen in Table 9 above, both *be* and *have* appeared among the most common progressive main verbs in SICLE (proportions of 1.5 and 1.9) but not in FICLE (0.5 and 0.8) or GICLE (0.5 and 1.0). Even with the small size of the corpora in the present study, this result is interesting, as it leads to the conclusion that L1 Spanish speakers extend the progressive form even to stative situations, such as with verbs *be* or *have*, more than L1 Finnish and German speakers. As one of the most distinctive differences between these learner populations is the aspect-marking of the L1[15], this suggests that L1 transfer may explain this difference at least to a certain degree. This result is also in line with those of Meriläinen (2017), who discovered significantly higher proportions of the extended stative progressive in written learner English by L1 speakers of Chinese and Japanese compared with L1 Finnish and Swedish speakers. Chinese and Japanese are both aspect-marking languages, which lead Meriläinen (ibid.) to the conclusion that the extended, non-standard use of the progressive is influenced by the learners' L1 and that speakers of aspect-marking languages, such as Chinese and Japanese, are "more likely to stretch the boundaries of the progressive into stative and habitual contexts" (ibid.: 773), than native speakers of non-aspect languages, who do not

[15] Spanish marks aspect whereas Finnish and German do not, see Chapter 2.1.3.

experience such (negative) transfer from their L1. Further support for these findings was reported by Zeng, Shirai & Chen (2021: 280) who found L1 transfer effect to be a factor in the non-standard use of the progressive with stative verbs among Spanish native speakers.

It should be noted that even though the verb *have* was fairly common in the progressive in SICLE, not all instances of the verb were considered verbs of existence. Looking at the instances, only three were categorised as existence verbs whereas the remaining eight instances were categorised as activity verbs. Notable here is that in seven instances, *have* was used as part of the phrase *have an affair with*, as in:

(18) Leontes thinks that his wife, Hermione, is having an affair with Polixenes (SICLE, SPAL1001)

This diminishes the significance of the verb *have* when it comes to extended contexts, although the impact of *be* naturally remains the same in this respect. Furthermore, the frequent use of the phrase *have an affair with,* which influenced the proportion of the verb *have* of the progressive main verbs in the entire SICLE data, gives rise to the notion that genre- and topic-related elements are significant in considering lexical aspect, or especially the frequency of a specific verb in the progressive. If an essay topic focuses on a specific literary work, it can be expected that certain common phrases appear more often in the data and may thus skew the results.

v. Mental verbs

Mental verbs concern experiences or activities by a (human) subject and include meanings such as cognition (e. g. *think*), perception (e. g. *see*), or emotion (e. g. *love*) (Collins 2008: 234–5; Rautionaho 2014: 163). Collins (2008: 234–5) notes that the compatibility with the progressive with these verbs is especially various or meaning-dependent. For example, certain volitionally performed actions of perception or cognition can more easily be used in the progressive.

(19) they forgive each other and leave for Sweden again, although he *has been thinking* about staying and leaving his wife. (FICLE, FITU2003)

Contrarily, mental verbs that "have an inert, experiential, stative sense" (ibid.: 235), such as, *know, hear,* and *feel* are less compatible with the progressive. The combination is therefore often seen as extended, non-standard.

The proportions of mental verbs were fairly similar in FICLE (4.2%), GICLE (5.2%) and SICLE (5.7%). These results differ significantly from the results by Smith (2002) as reported by Collins (2008: 237), where mental verbs were found the second most prominent group (26.5%) in written BrE. Also Collins's (ibid.) data on ESL varieties showed higher frequencies for mental verbs, 14.2% in written and 12.3% in spoken register.

Mental verbs posed a particular challenge for the present analysis as analysing the use of these verbs would have often required extensive examination of the context[16]. Some differences were, however, found between SICLE and the other two corpora. In both FICLE and GICLE, the only mental verbs among the most common progressive verbs were *think* and *look* in the construction *look forward to*, which are both commonly used in the progressive especially in less formal contexts. In SICLE, on the other hand, the verb *see* appeared among the most common progressive main verbs, in addition to the verb *think.*

(20) […] you are aware of the misunderstanding and you *are seeing* how Marlow is behaving in a bad way towards Mr. Hardcastle […] (SICLE, SPM08014)

As Rautionaho (2014: 30) notes, verbs that express such involuntary perception are not usually compatible with the progressive. It is therefore interesting to see that the L1 speakers of Spanish use similar possibly extended, non-standard progressives with mental verbs as seen above with verbs of existence.

[16] For example, three instances of possible mental verbs (with *look*) in the SICLE corpus were labelled as indeterminate because of the ambiguity in meaning related to presumably erroneous use of prepositions in phrasal verb structures (*look at / look for / look back*).

vi. Causative and aspectual verbs

The final two semantic domains are those of causative and aspectual verbs. Causative verbs express a sort of facilitation, where an entity or person "brings about a new state of affairs" (Collins 2008: 236), e. g. *cause, help* and *allow*. These verbs are reported to be rare in the progressive, by both Collins (2008) and Rautionaho (2014), which was also the case in the present data: no instances of causative verbs were found among the most common progressive main verbs. Aspectual verbs indicate a state of progress of an activity, e. g. *begin, continue* (Collins 2008: 236) and are reported to be equally rare as causative verbs by Collins (ibid.) and Rautionaho (2014). In the present data, instances of an aspectual verb, *begin*, were found in the SICLE corpus. Its frecuency in the corpus was 1.3% and it formed 2.4% of the most common progressive verbs.

(21) In XV century protestantism *is beginning* (SICLE, SPM08003)
(22) […] situation, where women has more responsabilities whithin society: she *is beginning* to be taken into account, but only by the radicals of this religion. (SICLE, SPM05010)

In contrast, *begin* occurred only once in both FICLE (0.3%) and GICLE (0.2%). Although the frequencies in the data are small, this result, when combined with the extended uses found among verbs of existence and mental verbs, leads to the tentative conclusion that learner English by Spanish L1 speakers shows signs of extended or non-standard uses of the progressive in combination with certain semantic domains.

5 DISCUSSION AND CONCLUSIONS

This chapter summarises and discusses the results of the study, thereby answering the research questions set in Chapter 1. Research questions 1–3 concerned the frequencies of the progressive, the frequencies of the different progressive main verbs and the types (i. e. semantic domains) of main verbs used in the progressive in each learner English variety. In particular, this chapter attempts to provide answers to the fourth research question, that is, concerning the variables possibly influencing the results: the learners' L1 (aspect vs. non-aspect), the task- or genre-specific, thematic elements of the essays, and the learners' proficiency level (advanced vs. higher intermediate learners). Furthermore, this chapter discusses the reliability and significance of the results and presents suggestions for further research.

5.1 Discussion and summary of results

This study examined the use of the progressive form in academic, argumentative writing in three learner English or EFL varieties from native language backgrounds of Finnish, German, and Spanish. The primary goal was to examine lexical variation in the use of different types of main verbs in the progressive. The idea was to study the three varieties independently, i. e. not to compare them directly to any native English variety, but to see which patterns or differences emerge from the data through the examination of the frequencies and types of different progressive main verbs across the corpora.

The calculation of normalised frequencies of the progressive showed that the progressive occurred somewhat more frequently in the essays by the L1 German and Spanish students than in the Finnish speakers' ones. These results were consistent with those reported in previous studies (e. g. Virtanen 1997; Axelsson & Hahn 2001; Meriläinen 2017). Some studies have regarded higher frequencies of the progressive compared to native varieties as "overuse" and connected it with a lower proficiency level (e. g. Axelsson & Hahn 2001). As reported by

Granger et al. (2020: 12; see Table 2 in Chapter 3.2), the proficiency levels were presumably higher among the Finnish students which could possibly explain the differences between the varieties. However, studies have also suggested that thematic or task-dependant elements, which may influence the formality level of the texts, need to be considered carefully before making judgements based on frequencies of the form alone (e. g. Virtanen 1997; Dose-Heidelmayer & Götz 2016). Dose-Heidelmayer & Götz (ibid.: 239), for example, concluded that certain narrative tasks. e. g. story-telling, seem to invite use of the progressive. As discussed in connection with the different progressive main verbs, the essays in the German corpus contained more narrative elements than the other two sub-corpora. This could explain the higher frequency of the form in the German corpus compared to the Finnish one. Genre variation may further explain the differences between the Spanish and Finnish sub-corpora, as well. As reported by Axelsson & Hahn (2001: 7), a study on Swedish learners by Blomberg (2001) suggests that the progressive form was more than twice as frequent in literary essays than argumentative essays. As the Spanish students' essays contained literary references and quotations and were thus clearly written on more literary topics than most of the essays in the other two corpora, this could explain the differences in the frequencies of the progressive.

The analysis of the lexical variation in the progressive between the three corpora focused on the progressive main verbs. The comparison of the type/token ratios between the corpora showed that the Spanish students used fewer different main verbs in the progressive than the Finnish and German students. Based on previous studies (Virtanen 1997; Axelsson & Hahn 2001), this could possibly indicate a lower proficiency level among the Spanish students, or a lower degree of formality of the texts in SICLE. As discussed above, SICLE included more essays on literary topics. The style and level of formality in literary essays should generally be comparable to academic genre, and the formality levels of the Spanish students' essays should thus correspond to the ones in FICLE. Informality should therefore not explain the lower

type/token ratio. The combination of the results, i. e. of the higher frequencies of the progressive together with the lower type/token ratio, may possibly be considered an indication of the lower level of proficiency of the Spanish students influencing the use of the progressive. The differences between the frequencies and the type/token ratios were, however, too small to draw any definite conclusions.

Because of the small size of the data, the examination of the progressive main verbs concentrated only on the most frequently used main verbs. In general, the analysis showed that there is consistency between learner Englishes, and ESL and ENL varieties, as the most common verbs used in the progressive across the three learner English corpora, *do*, *go*, *try* and *talk*, were also found among the most common verbs in ESL (Rautionaho 2014) and ENL (Römer 2005), albeit the ESL and ENL varieties concerned spoken English.

The differences in the most frequent progressive main verbs between the learner Englishes were examined more closely through categorisation into semantic domains (Biber et al 1999, quoted in Collins 2008 and Rautionaho 2014). In all varieties, activity verbs were clearly the most prominent group (over 50% of the most common main verbs). However, the German students differed from the Finnish and Spanish students in their more frequent use of physical activity verbs, such as *sit*, which was the most frequent main verb in the GICLE corpus but had no occurrences in the other two corpora. Similar patterns were found with verbs of occurrence, which in the German students' essays were often used to refer to physical changes whereas in the Finnish and Spanish students' essays were used in a more abstract manner. Wulff & Römer consider such a shift in the written register from "more concrete, physical action verbs to abstract, metadiscoursal verbs" (2009: 127) a typical sign of higher level of proficiency. However, based on the report by Granger et al. (2020) on the proficiency levels of the ICLE students it does not seem probable that the proficiency level would explain the differences in the present data, especially since the Spanish students had, according to Granger et al (2020:

12) lower proficiency levels than the German students. Instead, the variation may be explained through thematic differences found between the essays: essays including narrative, personal style possibly use more techniques like setting the scene or "framing" (Römer 2005) with more physical verbs. The clear differences found in the proportions of verbs of occurrences and communication verbs, especially between the Finnish and German students, can be seen as supporting this view. Based on previous studies (e. g. Smith 2002, as reported in Collins 2008; Collins 2008; Rautionaho 2014), occurrence verbs have appeared more frequently in written registers whereas communication verbs in spoken registers. The differences in distributions of the verbs belonging to these domains could lead to the tentative conclusion that the German native speakers' essays were possibly more informal, speech-like in style. The observations on the choice of different verbs within these domains further support this conclusion: in the FICLE corpus more formal verbs were generally chosen, such as *become* over *get,* whereas the essays in GICLE contained more informal verbs like *get* and *talk*. The use of the verbs *become* and *get* in similar lexical-grammatical contexts, as for example, with the adverb *more and more*, in all varieties showed that general language proficiency is less likely the variable affecting these choices but more likely task-dependant elements or expertise. This view is supported by conclusions made in Wulff & Römer (2009: 130–1) and also aligns with the discussion above on the indications of the higher frequencies of the progressive in the German speaker corpus.

Looking at the differences between the Spanish sub-corpus and the other two corpora, however, it appears there may be more factors than thematic or genre-specific ones at play. The proportions of the semantic domains of the most common main verbs in the SICLE corpus did not show clear differences compared to the other two corpora. However, on closer examination of the different verbs belonging to each domain, some interesting observations were made. In SICLE, certain verbs occurred in the progressive more or were used in different contexts than in the other corpora. These verbs included the existence verbs, *be* and *have*, the mental verb *see,*

and the aspectual verb *begin*, none of which appeared among the most common progressive main verbs in the other two corpora. Non-standard or extended uses of the progressive, i. e. uses of the form in stative situations, were found on examination of the instances of these verbs. Spanish being an aspect language, these observations suggest that L1 transfer could be influencing the use of the progressive form in combination with these verbs. Support for this view was found in Zeng, Shirai & Chen (2021) and Meriläinen (2017) who concluded that L1 transfer is likely to have influenced the extended use of the progressive by the learners whose L1 is an aspect language.

5.2 Conclusions and outlook

Because of the limited scope of the study, certain restrictions had to be made on the data as well as on the factors and features analysed. Conclusions made based on the differences in frequencies and type/token ratios between the corpora can therefore only be tentative. Furthermore, concerning the semantic domains of the progressive main verbs, only the most frequent verbs, i. e. around half of the progressive instances were categorised. In addition, a more detailed examination of the instances was only conducted where seen relevant, i. e. based on the differences in the frequencies of the verbs or on lexical variation within the semantic domains. The categorisation into semantic domains and the individual examples drawn from the data on the uses of different main verbs involved a degree of subjectivity that needs to be considered in the interpretation of the results.

Nevertheless, this study produced some interesting results on the distribution of the progressive form and different progressive main verbs across the learner English varieties examined. Different factors were also considered that could explain the variation between the varieties. The analysis of the different main verbs showed how L1 native language is possibly influencing the use of the progressive, as seen among the Spanish learner population who

extended the progressive to stative contexts. A further important observation from the present study was, as also noted in various other studies (e. g. Dose-Heidelmayer & Götz 2016), that the thematic focus of the essay or task has an impact on the use of the progressive. Even though one of the goals of this study was to observe if any patterns in the use of the progressive could possibly be explained through L1 transfer, the conclusions drawn mostly concerned text type or thematic differences which presumably influenced the use of the progressive or certain progressive main verbs. In order to be able to examine the role of transfer more confidently between different learner English varieties, even more comparable corpora, i. e. corpora containing essays from the same topics, would be required. At any rate, controlling the variable of genre and task-dependant elements may still lead to similar results and conclusions as those presented in this study: as both Finnish and German are non-aspect languages, it is possible that the only remaining explanation for the differences between the corpora concerns the proficiency of the writers, at least within the scope of this study. The question then arises, which variables influence the acquisition of the form: matters relating to pedagogy and teaching or simply exposure to the target language. Understanding these aspects, as well as understanding L1 transfer more comprehensively, would require further, possibly more qualitative study on the individual native languages, as well.

Furthermore, as stated in Rautionaho, Deshors & Meriläinen (2018: 46), recent studies have shown a shift in focus towards the alternation of the progressive vs. the non-progressive. Comparisons between the use of individual verbs in the progressive versus in the non-progressive were not conducted in this study because of the small set of data which would (most likely) not have produced significant results. Examination of the progressive main verbs in non-progressive contexts could, however, reveal patterns that may help understand how the learners of English choose the progressive or non-progressive aspect in writing. Including other linguistic, contextual factors, such as adverbs used in combination with the progressive, which

could possibly impact the choice of the progressive aspect would be relevant, as well.

Finally, it should be noted that in light of the results of the present study, already, it seems likely that the tense-aspectual choices a learner of English as a foreign language makes, are influenced by multiple factors, such as the ones found here, L1 transfer, task-dependant, thematic elements, and language proficiency, simultaneously. Further variables include exposure to the target language as well as features of the model variety, i. e. BrE vs. AmE as the model variety in the given learner population. It may not be possible to control all variables to examine the impact of any one variable or even then, the results may indicate a combined effect, given the complicity of the construction in question. Although definitive conclusions could not be reached, the present study corroborated findings from previous studies, provided discussion and suggestions thus contributing to a more comprehensive study on the progressive in individual learner English varieties to be carried out in the future.

REFERENCES

Corpora

Granger, Sylviane, Maïté Dupont, Fanny Meunier, Hubert Naets & Magali Paquot. 2020. *International Corpus of Learner English. Version 3*. Louvainla-Neuve: Presses universitaires de Louvain, https://corpora.uclouvain.be/cecl/icle/ [accessed: 14 April 2022]

Sources cited

Axelsson, Margareta, W & Angela Hahn. 2001. The use of the progressive in Swedish and German advanced learner English - a corpus-based study. *International Computer Archive of Modern and Medieval English (ICAME) Journal* 25, 5–30.

Binnick, Robert I. 1991. *Time and the Verb: A Guide to Tense and Aspect*. Cary: Oxford University Press, Incorporated.

Collins, Peter. 2008. The progressive aspect in World Englishes: A corpus-based study. *Australian Journal of Linguistics* 28(2), 225–249.

Collins, Peter. 2009. The progressive. In Pam Peters, Peter Collins, and Adam Smith (eds.), *Comparative Studies in Australian and New Zealand English: Grammar and Beyond,* Amsterdam/Philadelphia: John Benjamins Publishing Company, 115–123.

Dose-Heidelmayer, Stefanie & Sandra Götz. 2016. The progressive in spoken learner language: A corpus-based analysis of use and misuse. *International Review of Applied Linguistics in Language Teaching* 54(3), 229–256.

Gilquin, Gaëtanelle & Sylviane Granger. 2011. From EFL to ESL. Evidence from the International Corpus of Learner English. In Marianne Hundt & Joybrato Mukherjee (eds.), *Exploring Second-language Varieties of English and Learner Englishes: Bridging a Paradigm Gap,* Amsterdam/Philadelphia: John Benjamins Publishing Company, 55–78.

Granger, Sylviane, Maïté Dupont, Fanny Meunier, Hubert Naets & Magali Paquot (eds). 2020. *International Corpus of Learner English. Version 3*. Louvainla-Neuve: Presses universitaires de Louvain.

Hundt, Marianne & Katrin Vogel. 2011. Overuse of the progressive in ESL and Learner Englishes: Fact of fiction? In Marianne Hundt & Joybrato Mukherjee (eds.), *Exploring Second-language Varieties of English and Learner Englishes: Bridging a Paradigm Gap,* 145–165. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Kirk, John M. 2015. The progressive in Irish English. Looking both ways? In Peter Collins (ed.), *Grammatical Change in English World-Wide,* 87–118. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Kranich, Svenja. 2010. *Progressive in Modern English: A Corpus-Based Study of Grammaticalization and Related Changes*. Amsterdam: Rodopi.

Meriläinen, Lea. 2017. The progressive form in learner Englishes: Examining variation across corpora. *World Englishes*, December 2017, Vol. 36(4), 760–783.

Meriläinen, Lea, Heli Paulasto & Paula Rautionaho. 2017. Extended uses of the progressive form in Inner, Outer and Expanding Circle Englishes. In Markku Filppula, Juhani Klemola, Anna Mauranen & Svetlana Vetchinnikova (eds.), *Changing English: Global and Local Perspectives*, Vol. 92, 191–216. Berlin/Boston: De Gruyter, Inc.

Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study.* Cambridge: Cambridge University Press.

Rautionaho, Paula. 2014. Variation in the progressive: A corpus-based study into world Englishes. Tampere: Tampere University Press.

Rautionaho, Paula, Sandra C. Deshors & Lea Meriläinen. 2018. Revisiting the ENL-ESL-EFL continuum: A multifactorial approach to grammatical aspect in spoken Englishes. *ICAME Journal* 42(1), 41–78.

Römer, Ute. 2005. *Progressives, patterns, pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics,* Amsterdam/Philadelphia: John Benjamins Publishing Company

Schneider, Edgar W. 2007. *Postcolonial English. Varieties around the world.* Cambridge: Cambridge University Press.

Shirai, Yasuhiro. 2013. Defining and coding data: Lexical aspect in L2 studies. In Rafael M. Salaberry & Lorenç Comajoan (eds.), *Research design and methodology in studies on L2 tense and aspect*, 271–308. Boston/Berlin: Walter de Gruyter.

Virtanen, Tuija. 1997. The progressive in non-native speaker and native speaker composition: Evidence form the International Corpus of Learner English. In Magnus Ljung (ed), *Corpus-based studies in English: Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17)*, 299–309. Leiden, The Netherlands: Brill.

Wulff, Stefanie & Ute Römer. 2009. Becoming a Proficient Academic Writer: Shifting Lexical Preferences in the Use of the Progressive. *Corpora* 4(2), 115–133.

Zeng, Xiaoyan, Yasuhiro Shirai & Xiaoxiang Chen. 2021. Universals and Transfer in the Acquisition of the Progressive Aspect: Evidence from L1 Chinese, German, and Spanish Learners' Use of the Progressive -Ing in Spoken English. *International review of applied linguistics in language teaching, IRAL* 59(2), 267–292.