UNIVERSITY OF TARTU
DEPARTMENT OF ENGLISH STUDIES

A CORPUS-BASED STUDY OF *LIKE* IN ESTONIAN EFL LEARNERS' SPEECH

BA thesis

JOHANNA KONSO
SUPERVISOR: ***Assoc. Prof.*** **JANE KLAVAN**

TARTU
2021

ABSTRACT

Pragmatic markers (PMs) are an essential part of conversation. However, with PMs such as *like* appearing everywhere, they are stigmatised as being only used for hesitation and therefore being meaningless. While definitely an important function of *like*, the word has a multitude of other uses that are prominent in speech. This paper hopes to disprove the common belief that *like* is only used for hesitation, as well as determine how different is the use of *like* in speech between Estonian learners of English as a Foreign Language (EFL) and native speakers. In order to answer that question, a corpus-based study was carried out, using the Estonian subcorpus of the Louvain International Database of Spoken English Interlanguage (LINDSEI-EST) and Louvain Corpus of Native English Conversation (LOCNEC) respectively. While there have been theses written utilising the LINDSEI-EST corpus, PMs in Estonian EFL learners' speech have received almost no attention so far.

The following section provides an introduction to the topic, revealing the motivation for writing this paper, as well as giving a quick overview of the thesis. The first part of the paper discusses previous research on PMs in learner language. The empirical part of the paper begins with an overview of the methodology, explaining how 400 random samples were analysed. The following subsections introduce the corpora and define the functions of *like* that were observed in the sample. After that, the data is analysed, and the results are presented and discussed. The final section summarises the key points and draws conclusions based on the results of the thesis.

TABLE OF CONTENTS

ABSTRACT 2

LIST OF ABBREVIATIONS 4

INTRODUCTION 5

1. PRAGMATIC MARKERS IN LEARNER LANGUAGE 7
    1.1 The pragmatic marker *like* 9
    1.2 Learner language and corpora 11

2. EMPIRICAL ANALYSIS OF *LIKE* 13
    2.1 Methodology 14
        2.1.1 The LINDSEI and LOCNEC corpora 15
        2.1.2 Functions of *like* 16
            2.1.2.1 Grammatical functions 17
            2.1.2.2 Pragmatic functions 19
    2.2 Data analysis and results 23
    2.3 Discussion 27

CONCLUSION 31

REFERENCES 33

APPENDIXES 35
    Appendix 1: Total number of *like* uses by each speaker in LINDSEI-EST 35
    Appendix 2: Total number of *like* uses by each speaker in LOCNEC 36

RESÜMEE 38

LIST OF ABBREVIATIONS

EFL—English as a Foreign Language
KWIC—Keyword in context
LINDSEI-EST—The Estonian subcorpus of the Louvain International Database of Spoken English Interlanguage
LOCNEC—Louvain Corpus of Native English Conversation
PM—Pragmatic marker
TCELE—Tartu Corpus of Estonian Learner English

INTRODUCTION

Pragmatic markers (PMs) are very common in speech, especially *like* which has a certain stigma to it. As PMs are not taught in class, using them can come across as incorrect (Diskin 2017: 146) when in fact they can make the speaker seem more natural and even polite (Andersen 2001: 229; Svartvik 1980: 171). As argued by Fraser (1996: 188) the meaning of each sentence consists of "propositional content" and "a set of pragmatic markers", which makes PMs essential to a conversation. However, according to popular belief, the use of *like* is attributed to people who are not as educated or intelligent (D'Arcy 2007: 388). Andersen (2001: 216) and D'Arcy (2007: 395) want to disprove the claim that *like* is a mere hesitation marker. Similarly to these studies, the current thesis also explores how much *like* is used for hesitation to help dispel the myth of it being meaningless.

Despite the high frequency of *like* as a PM in speech, it has not yet been researched among Estonian learners of English as a Foreign Language (EFL). There have been theses written using the Estonian subcorpus of the Louvain International Database of Spoken English Interlanguage (LINDSEI-EST), for instance Rahusaar (2019), Toom (2020) and Piiri (2020), but none on the word *like* due to how new the corpus is. The research most relevant to the thesis at hand is the one by Rahusaar (2019) who conducted a short empirical study of the PM *well*.

Learner language in general is an important aspect to study. Learner corpora make it possible to see where learners commonly make mistakes, which can be used to great effect when compiling textbooks. However, learner language is not easy to collect or analyse. As mentioned by Granger (2008: 260), learners do not use the language in natural conditions, which means that the authenticity of the data is always questionable. In order to

receive the most natural data, open-ended tasks should be used as they allow the learner to express themselves as freely as possible.

The aim of the present thesis is to determine how different is the use of *like* in speech between Estonian EFL learners and native speakers. The thesis will focus mostly on the pragmatic functions of *like* in an effort to see how often *like* is used for hesitation, both by learners and native speakers. In order to answer the research question, an empirical corpus-based study of *like* was carried out.

The first part of this paper provides an overview of previous research on the topic. It is explained why the term pragmatic marker (PM) is hard for researchers to define due to its multifunctional nature and for the purposes of this thesis, a definition is provided by the author. This discussion is followed by a look at the existing research of the PM *like* and an overview of learner language and learner corpora is given.

The second part of the paper describes the empirical analysis of *like*. The section begins with an overview of the methodology, followed by an introduction of the two corpora used for the thesis at hand. In the next subsection, each function that was observed in the corpora is explained briefly and examples from the corpora are provided to illustrate the point more clearly. After the methodology subsection, the data is analysed and the results are presented. The section concludes with a discussion of the results.

1. PRAGMATIC MARKERS IN LEARNER LANGUAGE

Despite being an interest of study for a while, knowledge on PMs is still limited (Aijmer & Simon-Vandenbergen 2011: 223). One of the first challenges that researchers face is choosing a term for the oftentimes meaningless words and phrases in speech (such as *like*, *well*, *sort of* and *I mean*) that have been referred to as PMs in the current thesis. The reason for the difficulty of picking a term is due to the literature on the topic using different terms to refer to it to the point that those various terms have almost as if been used interchangeably. Aijmer and Simon-Vandenbergen (2011: 226) and Brinton (1996: 29) both point out terms such as *pragmatic marker, discourse marker, discourse particle, pragmatic particl*e and *filler* among numerous others that were not mentioned. The authors of both articles (Aijmer & Simon-Vandenbergen 2011: 227; Brinton 1996: 29–30) prefer the term *marker* to *word* or *particle* as it does not exclude phrases unlike *word*, nor does it confuse the reader, unlike *particle* which in grammar is a fixed term for a part of speech. Aijmer and Simon-Vandenbergen (2011: 227) mention that the term *pragmatic marker* is often used as an umbrella term, although it has a slightly different function when compared to a *discourse marker*. Diskin (2017), for instance, avoids the problem of distinguishing between the functions of the two terms by referring to it as a *discourse-pragmatic marker* instead.

Trying to define the term is a challenge on its own due to the conflicting interpretations of it in literature. Brinton (1996: 31, 34–35) mentions that PMs are not mandatory but rather seem to be grammatically optional, as well as their function in keeping the discourse continuous and filling moments of silence while the speaker thinks of what to say. Aijmer and Simon-Vandenbergen (2011: 225–226) highlight that PMs are frequent in conversation and due to that are regarded as informal, but most importantly of all carry little to no propositional meaning. Fraser (1996: 188) argues that “the sentence

(read ‘semantic’) meaning is comprised of two parts: A propositional content; and a set of pragmatic markers”. Andersen (2001: 39) similarly describes PMs as “generally hav[ing] little lexical import but serv[ing] significant pragmatic functions in conversation”. What this means is that PMs are an essential part of one’s speech. As Andersen (2001: 40) points out, their meaning “lie[s] beyond [their] propositional meaning”, but they are most certainly meaningful (Andersen 2001: 59).

Another important aspect of PMs is their connection with sounding native-like. Svartvik (1980: 171) highlights the importance of PMs in the communication between EFL learners and native speakers—omitting or incorrectly using a PM is not an error that a native speaker is likely to correct, unlike making a grammatical error with irregular plural forms (e.g. *two gooses*) or tenses (e.g. *she cutted*). However, as Svartvik (1980: 171) further notes native speakers are presumably going to regard the EFL learners as “dogmatic, impolite, boring” or “awkward to talk to”. Hasselgren (2002, as cited in Santos 2019: 4) establishes a connection between the use of PMs and the speakers’ proficiency through her study and considers PMs as a marker of native-like fluency. Santos (2019: 4) similarly claims that PMs are “fundamental to successful communication”, but also points out how the misuse and lack of them can negatively affect the learner’s ability to contribute to a conversation. This further confirms the previously mentioned argument that Fraser (1996: 188) made regarding PMs being one of the two parts to give a sentence its meaning.

Based on the previous discussion, the definition of the term could be as follows: A PM is a word or a phrase in speech that usually does not change the propositional meaning of a sentence. It helps the speaker organise their thoughts and maintain the flow of the conversation, not to mention that using PMs can help EFL learners sound more native-like.

1.1 The pragmatic marker *like*

As it is something learners are not formally taught, PMs often have a negative connotation and are encouraged to be avoided (Diskin 2017: 146). This is a view supported by D'Arcy (2007: 388) who has also pointed out that the users of *like* are frequently seen as unintelligent and in general people have a negative attitude towards them. Diskin (2017: 146) describes the language that learners encounter in class as an "idealised standard language" that gives them the wrong perception of the actual spoken language that native speakers use.

To further elaborate on the connotation behind the PM *like*, D'Arcy (2007) investigates the myths surrounding the use and users of *like*. The data D'Arcy (2007: 389) uses is from the Toronto English Archive, a corpus of spoken contemporary English, collected in Toronto in 2002–2004. D'Arcy (2007: 412) concludes by stating that *like* is not only something that adolescents, women and Americans use, but it is more so a part of everyone's speech. Although it is more frequent in teenager talk, contrary to popular belief, people do not outgrow the use of *like* (D'Arcy 2007: 403). The reason why D'Arcy's (2007: 390) research is so valuable is because she is one of the few to have analysed the uses of *like* among all ages of a community, rather than focusing on one specific age group as was done, for instance, by Andersen (1997).

Another myth that D'Arcy (2007: 390) mentions is "talk[ing] of *like* as a single, monolithic entity" and therefore being meaningless in speech. The reason being that to an ordinary person all pragmatic uses of *like* sound the same (D'Arcy 2007: 411). This idea explains why users of the PM *like* are seen in a negative light, as the PM is often seen as a filler and a marker that uneducated people use. Andersen (2001: 227) rejects the idea of *like* as a mere filler, while still acknowledging that the PM can be used with planning

difficulties, false starts and self-repairs. Andersen (2001: 228-229) supports his claim with three points:

1. *Like* occurs frequently in conversation; however, not all cases suggest planning difficulties as *like* can be uttered during a continuous and quick flow of speech, and pronounced in the same manner as the rest of the words, the so-called real parts of the sentence.
2. *Like* has other meanings that cannot be associated with filled pauses (such as *um* and *er*). As such, it is important in the interpretation of an utterance, signalling loose talk among other things. Omitting the PM could even affect the truth condition of what was said and "lead to pragmatic anomaly" (Andersen 2001: 228).
3. *Like* shows traces of its original lexical meaning 'similar to', which indicates that the PM cannot be placed anywhere in a sentence as opposed to fillers.

These points clearly indicate that *like* is not simply a hesitation marker that must be avoided, but something that is even necessary in conversation.

When talking about the studies on *like* the focus is usually on different varieties of native English, such as Irish English (Diskin 2017; Schweinberger 2015), Canadian English (D'Arcy 2007; Tagliamonte 2005) and varieties of British English (Andersen 1997; Schweinberger 2015). There have also been studies (Diskin 2017; Magliacane & Howard 2019; Santos 2019) on second language acquisition. Aijmer and Simon-Vandenbergen (2011: 230) also mention that PMs are more commonly studied in native versus non-native speaker communication, made possible with the existence of comparable learner and native speaker corpora, such as LINDSEI. To expand on a few studies, Diskin (2017) explored the use of *like* among Polish and Chinese migrants who came to Ireland 1–11 years prior, and found that its frequency is largely dependent on the migrants' length of residence and that they had trouble adopting the clause-final *like*, a

feature of Irish English. Magliacane & Howard (2019) analysed the use of *like* by Italian EFL learners studying abroad in Ireland for six months and determined that the learners' use of the PM increased but did still not compare to the native speakers' use of *like*. Finally, Santos (2019) explored the use of *like* by six Brazilian university students who had been living in Ireland for at least six months and found that the learners overused *like* and noted it could be simply to make themselves better understood.

1.2 Learner language and corpora

Another term besides PMs that must be explained in detail is learner language, commonly studied with the help of learner corpora. As defined by Granger (2008: 259) learner corpora can in simple terms be said to be "electronic collections of texts produced by language learners". However, with English as such a widely-spoken language all over the world, it is not that clear who exactly are language learners (Granger 2008: 259). It is generally agreed upon that they are foreign language learners, a notion that excludes native speakers and those who live in a country where the language is institutionalised (Granger 2008: 259). This definition supports Estonian EFL learners since English is hardly ever used in their day-to-day life and has no official status in Estonia.

In general, learner corpora can, for example, be categorised by use (commercial or academic), size, language and form (written or spoken) (Granger 2008). Granger (2008: 261) notes that written corpora are more common, but it is expected to change in the near future as information and communication technologies start to be used more in teaching. When looking at the different corpora compiled in Estonia, the two most recent notable ones are the Tartu Corpus of Estonian Learner English (TCELE) and the Estonian subcorpus of the Louvain International Database of Spoken English Interlanguage

(LINDSEI-EST), which contain written and spoken data of Estonian EFL learners' English usage respectively. Using these corpora as a basis, a number of theses have been written by students of the University of Tartu. For instance, Toom (2020) used both corpora to analyse the use of phrasal verbs by Estonian EFL learners, while Piiri (2020) studied formulaic language. On the other hand, Rahusaar (2019) described the compilation of the spoken subcorpus of TCELE, which would later be used for LINDSEI-EST, and conducted a short empirical analysis of the PM *well*. The fact that there is already interest in the LINDSEI-EST corpus while it is still being compiled indicates that spoken language is starting to be studied more.

Many more studies have been created with the help of the LINDSEI corpus to analyse aspects of learner English, most notably lexis, syntax, phraseology, discourse and pragmatics (UCLouvain n. d. (a): para. 4). For instance, Aijmer has done a few studies on PMs in learner English such as *I don't know* and *dunno* (2009) and *well* (2011). Aijmer's (2011) comparative study shows that Swedish learners overuse *well*, mostly using it for speech management to compensate for their level of English. At the same time, they also underuse it for attitudinal purposes. Such studies give a good overview of the English that learners actually use, helping teachers make adjustments to their methods, as well as improve textbooks.

Learner data, however, is difficult to work with for a number of reasons. Granger (2008: 260) mentions the authenticity issue, as foreign language learners do not often use the language under non-experimental conditions, unlike native speakers. The tasks that provide the highest degree of naturalness are informal interviews, with fill in the blanks and reading aloud exercises being the most limiting (Granger 2008: 260). It must be mentioned that in that regard, the LINDSEI-EST corpus being informal and using

open-ended tasks, allows the speakers to express themselves rather freely, making the learner data quite authentic.

Another issue concerns data processing. Granger (2008: 263) points out that more often than not, learner corpora contain no annotations, which means researchers are usually working with raw data. Annotating tools are commonly used; however, they work best on native speaker data and as such "there is no guarantee that they will perform as accurately when confronted with learner data" (Granger 2008: 263). Undo (2018) calculated the error percentage of an automated part-of-speech tagger using TCELE, the results of which confirmed that taggers such as the one used in Undo (2018) have a fairly high success rate with advanced learner data, especially with learner data in written form. Another type of annotation is error annotation, which helps researchers locate words that have, for instance, been misspelled (Granger 2008: 264). Further work with the LINDSEI-EST corpus could possibly invest time and effort into automatic part-of-speech tagging of the corpus; this would benefit future research looking at the different lexical and grammatical aspects characteristic of Estonian EFL learners' speech.

2. EMPIRICAL ANALYSIS OF *LIKE*

This part of the paper begins with an overview of the methodology, explaining the whole process in detail. The following subsections introduce the corpora and define the functions of *like* that were observed in the sample of 400 concordance lines with examples from the corpora. Finally, the data is analysed, and the findings are presented and discussed.

2.1 Methodology

The aim of the thesis is to see how the use of *like* differs between Estonian EFL learners and native speakers of English. In order to answer this question, 200 random samples were analysed from both the LINDSEI-EST and LOCNEC corpora, 400 samples in total (see the following subsection for details about the two corpora). Before the analysis itself, the corpus data had to be cleaned in order to analyse the interviewees' use of *like* only. This means that all A turns, in other words the interviewer's turns, were removed. While this made analysing the data harder as roughly half of the context had been erased, it assured that the analysed data were produced by learners.

Both corpora were then individually opened in a concordance software called AntConc, which is a free toolkit used for corpus analysis (Anthony 2020). The software was mainly used to identify all instances of *like* in the Concordance tab (see Figure 1 for a screenshot of the software). The File View tab was utilised when assigning functions later on, to give more context, since it is easy to click on the keyword in context (KWIC) that is marked in blue and be taken to the exact place the word was mentioned in the file.

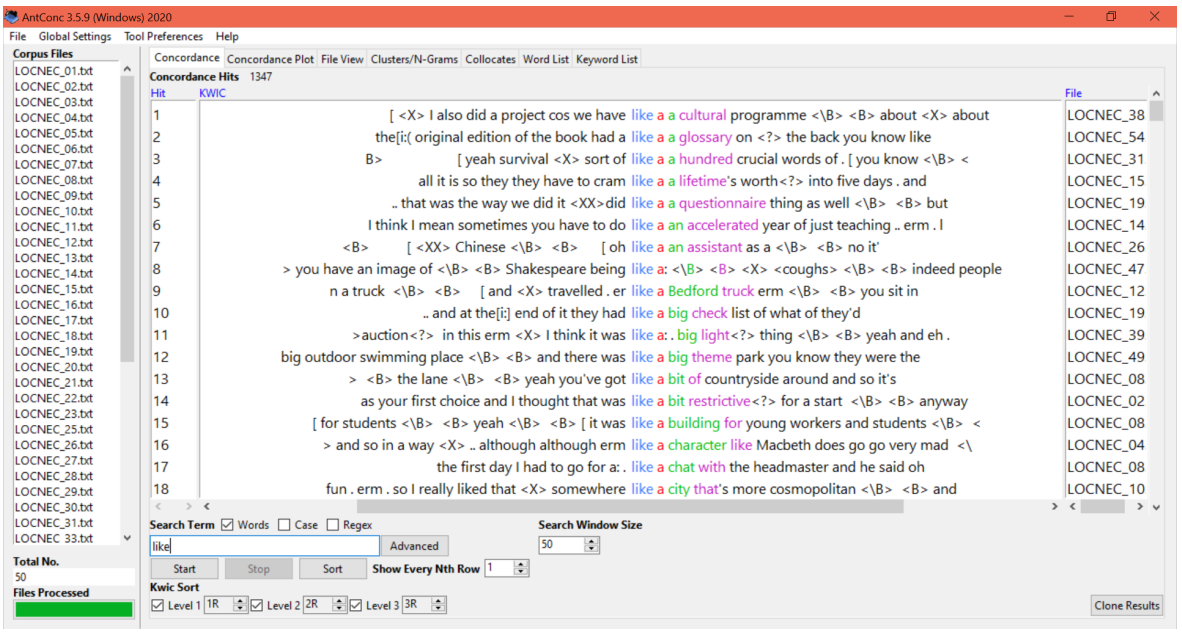

Figure 1. AntConc Concordance view of *like* in LOCNEC

After loading the corpus data in AntConc, the concordance lines were copied into Google Sheets in a similar format. As the sizes of the corpora are not comparable, a representative sample needed to be selected from both. One column was added to the spreadsheet with the random value of each concordance line, utilising the RAND() formula in Google Sheets. The first 200 concordance lines sorted by random value from both sets of data were copied into another sheet where each use of *like* could be assigned its function and be analysed in more detail.

2.1.1 The LINDSEI and LOCNEC corpora

The Louvain International Database of Spoken English Interlanguage, better known as the LINDSEI project, is a corpus of spoken learner English. More specifically, it consists of numerous subcorpora by advanced learners of English who come from different language backgrounds (UCLouvain n. d. (a): para. 1). One of these corpora is the Estonian subcorpus LINDSEI-EST that is also used for the thesis at hand. At the present moment, the Estonian subcorpus as well as four others (Saudi Arabian Arabic, Lebanese Arabic, Croatian and Iranian) are being compiled, with twenty subcorpora already completed (UCLouvain n. d. (b)).

Each completed subcorpus consists of approximately fifty interviews which all have the same structure and tasks—set topic, free discussion and picture description (UCLouvain n. d. (a): para. 2). The conversation starts with one of the three available topics, either an important lesson they learnt from an experience, an impressive country they visited, or a film/play that they found memorable (Rahusaar 2020: 22). After the short three- to five-minute monologue, the interviewer asks some further questions (Rahusaar 2020: 22). The interview concludes with a picture description task where the interviewee has to create a story based on four pictures around a painter and a woman (Rahusaar 2020:

22). In order for the talk to remain spontaneous, the interviewees are not allowed to take any notes during any of the tasks (Rahusaar 2020: 22). The interviews are transcribed in accordance with the guidelines set by LINDSEI and linked to a profile that provides information about the interview, as well as information about the learner and the interviewer (UCLouvain n. d. (a): para. 2). Thanks to the profile, the influence of various aspects on learner language can be researched in more detail (UCLouvain n. d. (a): para. 2).

In addition to learner English, a native speaker counterpart (LOCNEC) is available, which makes comparing the two possible (UCLouvain n. d. (a): para. 3). An example of that can be seen in Toom's (2020) thesis where the LINDSEI-EST and LOCNEC corpora were used to analyse the use of phrasal verbs between Estonians and native speakers of English. Having each subcorpus following the same guidelines makes it fairly easy to also compare studies, as was done by Rahusaar (2019: 35) who conducted a comparative analysis in her thesis with Aijmer's (2011) prior research on the topic.

At its current stage, the LINDSEI-EST corpus consists of 25 interviews (33,842 words of transcribed text) recorded in 2018. All of the interviewees (18 female, 7 male; average age 23 years) were native speakers of Estonian. They were third or fourth year students of English language and literature at the University of Tartu. The LOCNEC corpus consists of 50 interviews (118,159 words of transcribed text) with native speakers of English, all of them undergraduate and graduate students at Lancaster University in the UK.

2.1.2 Functions of *like*

Like has a multitude of grammatical functions that combined with its pragmatic functions in speech makes it seem as though *like* can be used anywhere (D'Arcy 2017: 3).

The following subsections will explain the contexts in which the widely-used *like* can be used, with examples from the two corpora, LINDSEI-EST and LOCNEC. For clarity the usage of the *like* analysed in the thesis will be provided in *italics* in all examples. In sentences where multiple uses of *like* are present only the one corresponding to the analysed concordance line will be italicised.

2.1.2.1 Grammatical functions

D'Arcy (2017: 3) has proposed seven main grammatical functions of *like*—verb, adjective, noun, preposition, conjunction, complementiser and suffix—that she refers to as "unremarkable" due to their long-standing use in English. It must be mentioned that the reason these functions are considered grammatical is because in such constructions *like* is not optional (D'Arcy 2017: 3). For the purposes of the current thesis, these categories were reduced to just four: verb, preposition, complementiser and suffix. There were no instances of adjectives, nouns nor conjunctions in the analysed samples. Adjectives and nouns especially are fairly infrequent as pointed out by D'Arcy (2017: 5). A common construction of *like* as a noun is *the likes of*. As only the form *like* is analysed in the present thesis then it is very probable that the exclusion of the plural form as the search term is the explanation for the absence of noun usage in the samples. The following sections will go into more detail regarding the grammatical functions of *like* that were analysed in the present thesis.

Like as a **verb** is one of the best recognised functions of the word. This is due to the fact that it has been a feature of English since Old English (D'Arcy 2017: 4). The Cambridge Online Dictionary (n. d.) mentions that *like* can mean *enjoy*, as can be seen in (1a) and with *to*-infinitive in (1b). It is also commonly used to make offers and requests

with the *would like* construction (Cambridge Online Dictionary n. d.). An example of the construction can be seen in (1c) where the speaker politely expresses wanting to teach.

(1a) you know it's not as if I'm living there and I'm working there but .[1] I really did *like* it a lot (LOCNEC_49[2])

(1b) (mm) I *like* to bake .. which I can't do at my dorm . because we don't have a stove . or an oven (LINDSEI-EST012)

(1c) I'd *like* to teach (em) literature and culture (LINDSEI-EST002)

As a **preposition**, *like* can be replaced with *similar to* (Cambridge Online Dictionary n. d.). It commonly appears with verbs such as *look*, *sound*, *feel*, *taste* and *seem* (Cambridge Online Dictionary n. d.). Due to the nature of the picture task in the interviews used for the corpora, *look* and *like* often occur together, as is the case in (2a). The same meaning of similarity can be seen in yet another common construction: *something like that* (2b).

(2a) erm she she doesn't really look very much *like* the woman in the final portrait (LOCNEC_35)

(2b) (mm) .. I don't I don't think I have like this de= definite . fav= favourable genres or or something *like* that (LINDSEI-EST013)

According to D'Arcy (2017: 7), a **comparative complementiser** is a newer construction of *like* where it can be substituted with *as if*, *as though*, *that* and a null complementiser at the beginning of a finite subordinate clause (3a). Conveying similarity or comparison, *like* usually follows verbs such as *seem*, *appear*, *look*, *found* and *feel*, which are more commonly known as experiencer or perception verbs (D'Arcy 2017: 7). Another construction where *like* is used as a comparative complementiser can be seen in (3b),

[1] The period signals a pause, in this case a short pause. Up to three periods can be used to mark empty pauses.
[2] The number represents the number of the interview in the corpus, in this case LOCNEC interview 49.

where the infinitive *seemed to be* has been replaced with the complementiser *like* and its finite subordinate clause (D'Arcy 2017: 8).

(3a) looking back on it I really feel *like* I achieved a lot (LOCNEC36)

(3b) in general it seemed *like* a nice place and not . too different from Estonia (LINDSEI-EST025)

As described by D'Arcy (2017: 8) *like* as a **suffix** gives a word qualities of resemblance. One of the two instances of suffixes in the analysed samples can be seen in (4).

(4) so in the third picture we see the[i:] beautiful beautiful young women on the paint= woman on the painting that should re= represent the[i:] actual model who is. less less model *like* (LINDSEI-EST016)

D'Arcy (2017: 8) mentions that the suffix *like* is uncommon, which is exemplified by the results of the current thesis. While it could be the case of its infrequency, it is more likely that many instances of *like* as a suffix were skipped as only the word *like* on its own was analysed, forms such as *childlike* or *ball-like* were not taken into consideration. As such, no definite conclusions about the suffix *like* can be drawn from this research.

2.1.2.2 Pragmatic functions

Santos (2019: 5) and Magliacane and Howard (2019: 75–76) have both identified the following pragmatic functions: hedge, approximator, exemplifier, hesitation marker, focuser and quotative. Out of the mentioned functions, Diskin (2017: 148) analysed all of the functions except for *like* as a focuser and quotative. While *pragmatic* suggests that *like* cannot be omitted in these instances then there are two exceptions (approximator and quotative) that will be discussed in more detail in the following section.

The first pragmatic function of *like* to be discussed is the PM used as a **hedge**. Using a hedge makes the utterance less direct (Magliacane & Howard 2019: 76) and therefore more polite. Beltrama and Hanink (2019: 4) similarly highlight that *like* is used for weakening of commitment. Diskin (2017: 154) points out that it is commonly used with "short statements of opinion that could be perceived as face-threatening or opposing the interlocutor's views". Schweinberger (2015: 122) explains the importance of expressing vagueness, where *like* can "function as a face-saving device". In (5a), an example of hedging one's statement with the help of a clause-final *like* can be seen. Whereas in (5b) the speaker uses *like* to signal the vagueness of *very north*. In both examples *like* is preceded by other hedging devices such as *just* (5a) and *sort of* (5b).

(5a) yeah and then we went over to New Jersey and stayed there we were just so eager to get back into New York . cos like New Jersey there's not a lot happening there really it was just *like* (LOCNEC_53)

(5b) erm .. that was in the north erm it's a village called <name of the village> and I can't think where it was near now but it it was sort of *like* very north (LOCNEC_18)

The approximate function of *like* has been referred to as an **approximator** (Andersen 2001: 50; Magliacane & Howard 2019), approximate adverb (D'Arcy 2017; Diskin 2017; Maddeaux & Dinkin 2017), and approximate device (Santos 2019). However, in the current thesis, the term approximator will be used as it is the most neutral of the three. *Like* is used as an approximator before numerals (D'Arcy 2017: 9; Magliacane & Howard 2019: 4; Santos 2019: 10) where it can be replaced with words denoting approximation such as *about* or *roughly* (D'Arcy 2017: 10), an example of which can be seen in (6a). Both Santos (2019: 10–11) and Magliacane and Howard (2019: 4) point out that the approximator *like* does not necessarily have to be used before numerical

expressions, for instance, it can indicate that the term the speaker has used may not be the most appropriate (6b). This is supported by the fact that after saying *the older generation* the speaker has specified that what they meant were *the middle-aged people*.

(6a) you only have *like* one or two little freezer boxes in the fridge (LOCNEC_28)

(6b) a lot of them were bad at English anyway . but (eh) *like* the older generation the middle-aged people spoke no English (LINDSEI-EST020)

As omitting *like* in such a context can alter the meaning of the sentence, D'Arcy (2017: 9–10) argues against it being a PM. Andersen (2001: 260), however, suggests that it "signal[s] that the utterance contains a loose interpretation of the speaker's thought", but still considers it a PM like any other.

A common function for *like* as a PM is exemplifying, as observed by Diskin (2017: 153) and Magliacane and Howard (2019: 79), where *like* can be used instead of *for example* (Magliacane & Howard 2019: 75; Santos 2019: 11). For instance, the speaker's second usage of *like* in (7) takes on the role of an **exemplifier** as they provide a list of phrases needed for communication.

(7) so I could do like basic conversation *like* hello thank you . have a nice day the weather is nice it's sunny outside (LINDSEI-EST020)

The PM *like* is notorious for its function of expressing hesitation. In many cases, *like* conveys that what was said should not be taken word for word (Andersen 2001: 229). However, using *like* as a **hesitation marker** can make the speaker seem more polite and not as assertive (Andersen 2001: 229). Andersen (2001: 229) also mentions that *like* can be used when the speaker has trouble with planning; therefore, it helps connect prepositional elements with little to no connection to one another. It allows the speaker time to organise their thoughts and to either clarify or change the structure of their utterance (Santos 2019:

11). Andersen (2001: 249) similarly points out that *like* can "occur[] in connection with false starts and self repairs [sic]". An example of a structural change can be seen in (8a) where the speaker's hesitation can already be seen by the use of the filler *eh* and the slight pause after *like*. This is something Santos (2019: 12) mentions in regard to recognising the speaker's unfamiliarity with a topic or their limited vocabulary. In such cases, *like* usually appears with other hesitation markers, in particular, filled pauses (such as *er*, *em* and *mm*), and pauses in speech (Santos 2019: 12). In (8b) the speaker's hesitation is expressed in the repetition of *like a* and admitting they do not know the correct word.

(8a) it was (eh) very *like* . makes you feel so small (LINDSEI-EST020)

(8b) takes (er) the Mother Earth's sort of heart this sort of crystal . and puts it in a . I don't know how do you say *like* a . like a little statue or something (LINDSEI-EST003)

Underhill (1988, as cited in Schweinberger 2015: 120) states that *like* as a **focuser** mostly marks new information. Santos (2019: 11) and Magliacane and Howard (2019: 75) also mention that the focuser *like* is used to emphasise or highlight a word or phrase. In (9), for instance, the speaker adds stress with the focuser to signal that there were very many dogs.

(9) in the capital . (er) there were . *like* . so many dogs like people had . a lot of pets (LINDSEI-EST008)

The final pragmatic function observed in the corpus is that of the **quotative** *like*, the *be like* construction, which is the most recent addition of *like* (D'Arcy 2017: 16). While D'Arcy (2017: 16) does not consider the quotative *like* entirely pragmatic due to it carrying some propositional meaning (Andersen 2001: 50), there seems to be no consensus on whether the quotative is pragmatic or not, for instance both Diskin (2017: 149) and Schweinberger (2015: 118) excluded that function in their research. In order to follow the

pragmatic categories that were suggested by Santos (2019: 5) and Magliacane and Howard (2019: 75–76), the present thesis will also consider the quotative a pragmatic function of *like*. The *be like* construction introduces reported speech, sound or the speaker's feelings and inner thoughts (D'Arcy 2017: 16; Magliacane & Howard 2019: 75; Santos 2019: 12). Beeching (2016, as cited in Magliacane & Howard 2015: 75) supports the idea that *like* cannot be removed from the sentence without affecting its meaning, but argues that what was so to speak reported is a loose account of it. The previously mentioned nuances of the use of *like* can also be seen in (10) where the speaker lightly mocks the attitude of their fellow Estonians. While (11) does not follow the *be like* construction, it functions in the same way (Andersen 2001: 269). The quotative *like* also collocated with phrases such as *say*, *scare*, *think* and *point towards* in the analysed sample. Although the *be like* construction is still prominent, the mentioned phrases show how diverse the quotative *like* can be.

(10) even though . we were all tired and they were tired as well they still kept on this cheery very like happy mood . and it was amazing to see because (eh) Estonians they were all *like* <sighs> I need a rest I need to lay down (LINDSEI-EST007)

(11) and another time it just goes *like* <whistles> a flash of inspiration and like you're just scribbling away (LOCNEC_10)

2.2 Data analysis and results

Searching *like* in AntConc generated altogether 672 concordance lines in LINDSEI-EST and 1,347 in LOCNEC. However, it must be mentioned that the LOCNEC corpus is currently twice as large, so the numbers must be normalised before they can be

compared. In order to calculate the relative frequency, the following formula was used: $x = \frac{absolute\ frequency\ (N) \times 100{,}000}{corpus\ total\ word\ count}$. Using this formula, it appears that in LINDSEI-EST the frequency of *like* per 100,000 words is 1,985, and 1,139 in LOCNEC. The difference between the frequencies indicates that Estonian EFL learners overuse *like*. Whether the same could be said about other EFL learners is outside of the scope of the study, but it is very much possible to study this with the help of other available LINDSEI subcorpora.

When taking a closer look at the use of *like* produced by individual speakers, an interesting picture emerges (see Appendixes 1 and 2). In LINDSEI-EST the mean value of *like* per interview is 26.88 (median 23), in LOCNEC the results are quite similar with a mean of 26.94 (median 21.5). While the mean is in fact higher in LOCNEC than in LINDSEI-EST, the median values show that LOCNEC has more speakers in the lower half of the dataset. Estonian EFL learners' higher usage of *like* can also be seen in the highest and lowest absolute frequencies (N) per speaker. In LINDSEI-EST the most a speaker has used the word is 104 times, while in LOCNEC this number is 83. The same can be seen with the lowest absolute frequencies, which are 8 and 3 respectively. All of this is illustrated in Table 1.

Table 1. Use of *like* by speaker

	Mean	**Median**	**Highest N**	**Lowest N**
LINDSEI-EST	26.88	23	104	8
LOCNEC	26.94	21.5	83	3

In Table 2 the 15 most frequently used words in both corpora can be seen. In order to compile this ranking of words, the Word List tab in AntConc was used. However, it counted contracted forms such as *'s* and *'t* as words. Especially with *'s* it is often not clear what it represents, as such these forms were excluded from the list. Being based on spoken

data, a number of filled pauses (such as *erm*, *eh*, *em*) also appeared. Similarly to the contracted forms, these fillers were also removed. While 14 out of 15 most commonly used words match between the two, with *is* and *yeah* being the exceptions, the words themselves are not in the same order. The same can be seen in the case of *like*, which ranks eight places higher in LINDSEI-EST. This data gives further support to the claim that *like* is more common in Estonian EFL learners' speech compared to native speakers.

Table 2. The 15 most frequently used words by the interviewees

Rank	LINDSEI-EST	LOCNEC
1	I	I
2	the	and
3	and	it
4	to	the
5	it	a
6	a	to
7	**like**	yeah
8	of	you
9	that	of
10	in	was
11	so	that
12	was	so
13	but	in
14	you	but
15	is	**like**

Looking further into the reasons why *like* is very frequent in both corpora, its functions must be examined. To achieve that, all 400 concordance lines were manually checked and assigned their function according to the definitions outlined in section 2.1.2. The results of the distribution of the word's functions can be seen in Table 3. What is

noticeable is that the overall balance is off between grammatical and pragmatic functions. In LINDSEI-EST 28.5% of the sample were grammatical uses of *like*, while 71.5% were pragmatic. In the case of LOCNEC, these numbers are divided fairly evenly, 49% and 48.5% respectively. It must be mentioned that there were five instances in the native speaker corpus that could not be assigned a function based on the little context that was available (marked as N/A in Table 3). Many words were missing in these sentences due to the poor quality of the audio file or the speaker not pronouncing the words clearly enough.

Table 3. The distribution of functions of *like*.

Function	Total number of uses	LINDSEI-EST	LOCNEC
Grammatical total	**155 (38.75%)**	**57 (28.5%)**	**98 (49%)**
Verb	76 (19%)	32 (16%)	44 (22%)
Preposition	69 (17.25%)	20 (10%)	49 (24.5%)
Comparative complementiser	8 (2%)	4 (2%)	4 (2%)
Suffix	2 (0.5%)	1 (0.5%)	1 (0.5%)
Pragmatic total	**240 (60%)**	**143 (71.5%)**	**97 (48.5%)**
Hedge	17 (4.25%)	5 (2.5%)	12 (6%)
Approximator	18 (4.5%)	12 (6%)	6 (3%)
Exemplifier	27 (6.75%)	15 (7.5%)	12 (6%)
Hesitation marker	46 (11.5%)	31 (15.5%)	15 (7.5%)
Focuser	109 (27.25%)	66 (33%)	43 (21.5%)
Quotative	23 (5.75%)	14 (7%)	9 (4.5%)
N/A	5 (1.25%)	0 (0%)	5 (2.5%)
Total	**400 (100%)**	**200 (100%)**	**200 (100%)**

Based on absolute frequency, the three most dominating functions in LINDSEI-EST are focuser (66), verb (32) and hesitation marker (31), while in LOCNEC these functions are preposition (49), verb (44) and focuser (43). Overall, the most used functions are focuser (27.25%), verb (19%) and preposition (17.25%). *Like* might be

notorious for its use as a hesitation marker, but it is clear that a large part of its usage in speech is still grammatical. The reason the preposition *like* is as prominent with native speakers is because they used constructions such as *something like that* and *things like that* considerably more. As mentioned in section 2.1.2.1 the high usage of the preposition *like* was certainly influenced by the picture task itself. While the frequencies were not analysed per task, the collocation *looks like* appeared often while the interviewees were telling the story during the picture description task.

The results show that speakers in both corpora use *like* in the same way as a comparative complementiser, suffix and exemplifier. A slight deviation can be seen in the use of *like* as a hedge, approximator and quotative, though they appear to have been used in a similar way. However, where Estonian EFL learners and native speakers differ the most are with the more frequently used functions: verb, preposition, hesitation marker and focuser. It is clear that in these cases Estonian EFL learners overuse the pragmatic *like* (hesitation marker, focuser) and underuse the grammatical one (verb, preposition). Estonians were more than twice as likely to use the hesitation marker *like* in speech (compare 15.5% to 7.5%). Despite being quite prominent, these results show that *like* is not a mere hesitation marker, it has a multitude of other uses, most of which could be observed in the sample.

2.3 Discussion

The results of the current thesis clearly indicate that Estonian EFL learners overuse *like*. This might be due to the fact that the LINDSEI-EST corpus is currently not completed, having only the data of 25 interviewees, making it twice as small as LOCNEC. Even though a random sample was generated from both corpora and the numbers were

made comparable during all steps of the analysis, it is possible that a few of the *like* users had more effect on the data than others. For instance, with the focuser *like* 20 out of the 66 uses in the LINDSEI-EST sample were by Speaker 20, who used the focuser numerous times while talking about their experience abroad.

Based on the results, some conclusions can still be drawn. Estonians overusing *like* and using it considerably more for pragmatic functions could be attributed to the Estonian equivalent of *like—nagu*. While this prediction would need to be researched on its own, it could be a viable reason for the PM's overuse by Estonian EFL speakers. It would be interesting to see whether these results are reflected in other EFL learners' usage of *like*. Overall, the numbers did not show any drastic differences in any function. Most functions had been used in a similar way to the native speakers, with some being under- or overused. The reason for such behaviour will be discussed shortly.

A function that is more commonly attributed to learners is the hesitation marker. As discussed in section 2.1.2.2 it is used to delay time and oftentimes to change the structure of the sentence. It frequently appears in connection with filled pauses and actual pauses in speech. In the case of LOCNEC, there were even instances of the speakers using longer vowels with articles, which appear as a: and the[i:] in the transcriptions. The repetition of the hesitation marker *like* could also be seen, although only with Estonian EFL learners. Aijmer (2004: 185) similarly found that Swedish learners use clustering and collocations more, most likely due to communicative stress. According to Aijmer (2004: 185), "[r]epetition indicates non-fluency", so in this aspect Estonian EFL learners are not very native-like.

What Aijmer (2004: 185) also mentions is that when PMs cluster they assume the same function, which helps determine the functions of such uses, and not only in the case of hesitation markers. While there are functions (such as verb, comparative

complementiser, suffix and quotative) that can be assigned without question, then the rest are almost all highly subjective. The ones found most difficult in the current thesis was distinguishing between hedges and approximators. Both uses were used to communicate a sense of vagueness.

Another point to mention is the fact that some uses of *like* were used for two functions at the same time, which made it difficult to pinpoint the so-called correct one. An example of this can be seen in (12) where at first glance it looks to be an approximator, but based on the repetition of *forty-five degrees* it seems that *like* was used to stress how hot it was, rather than give an estimation, which makes it a focuser. In cases where *like* appeared with a phrase such as *something like that* which further indicates approximation, it was easier to make the distinction between the focuser and approximator *like*.

(12) inside the greenhouse it was *like* forty-five degrees forty forty-five degrees (LOCNEC_41)

A similar issue appeared with *like* as a focuser and exemplifier. When bringing an example the speaker is simultaneously focusing attention on it, as illustrated in (13). In this particular case the repetition of *you can* suggests that the speaker is trying to think of an example rather than focus attention on the fact that travelling is possible.

(13) you can you can leave *like* to travel but at <X>[3] I mean . you work like you have one day off a week (LOCNEC_34)

The fact that it was so difficult to assign the focuser function in two different situations indicates that *like* as a focuser might not be very accurately represented in the data analysis of this thesis.

Another shortcoming of this thesis is the fact that audio files were not available for both corpora, as such it was decided that the audio should not be used at all to analyse the functions of *like*. Having the possibility to check the audio of the transcription would have

[3] The <X> signals an unclear syllable or word in the transcription.

made it easier to assign functions based on clues from the voice tone and how stressed the word is. Hearing the actual conversation would have helped determine the use of *like* in sentences where some words could not be transcribed. The more context that is available during the analysis the easier it is to make the right decision in regard to its function.

CONCLUSION

PMs are an essential part of spoken communication, helping the speaker by giving them more time to think about what they wanted to say, as well as expressing nuances that would otherwise not be possible. While usually dismissed as meaningless in speech, PMs are an important aspect to study, especially in learner language, as their correct use can help the speaker sound native-like. While the authenticity of learner language is questionable, as the language that they use is not natural to them, informal interviews and open-ended tasks help the speakers express themselves better and help the researchers to collect authentic learner language data.

Despite being a frequent occurrence in speech, PMs have not been researched much in the context of Estonian EFL learners. The thesis at hand aimed to fill that gap by determining how different is the use of *like* in speech between Estonian EFL learners and native speakers. Another aim was to focus on the pragmatic functions of *like* in an effort to see how often *like* is used for hesitation.

In order to achieve this, a comparative corpus-based analysis was conducted, using the LINDSEI-EST and LOCNEC corpora. The search term *like* was entered into a concordance software, which revealed 672 and 1,347 instances of *like* in LINDSEI-EST and LOCNEC respectively. To compare the numbers, the relative frequencies were calculated. Estonian EFL learners had a relative frequency of 1,985 per 100,000 words, and native speakers 1,139 per 100,000 words. This data, as well as looking at the mean and median values revealed that Estonian EFL learners have a tendency to overuse *like*, which can likely be due to how much the Estonian equivalent *nagu* is used by Estonians.

After that the functions of *like* were defined and each use of *like* in the random sample of 400 concordance lines was assigned one of the ten functions discussed in the thesis. This process was not as easy and straightforward as it might seem, some functions

such as focuser, approximator and hedge proved more difficult than others, while some could be determined without issue. A study like this is subjective, as such further research could produce different results. Having said that, the analysis still reveals some trends in Estonian EFL learners' usage of *like*. For instance, the use of *like* was fairly similar in both corpora. The largest differences occurred in the case of the four overall most used functions—verb, preposition, hesitation marker and focuser—where learners tended to underuse the grammatical functions (verb and preposition) and overuse the pragmatic ones (hesitation marker and focuser).

Showing that *like* is not a mere hesitation marker also proved successful as nine other functions of *like* were observed in the sample. It is true that Estonian EFL learners show a tendency of overusing it as a hesitation marker, which indicates that they have trouble expressing themselves on a native level, frequently using clusters of PMs for instance.

The thesis is one of the first studies on the topic of PMs in Estonian EFL learners' speech. As such, further research can be carried out on this topic, analysing, for instance, the position of *like* or how *like* is used in every task of the interview in the LINDSEI corpora. With the methodology and *like* functions described in detail, it should be fairly easy to replicate the study. Another possibility would be to analyse other PMs such as *you know* and *so*. With research in this field of study in Estonian EFL learners' speech being in its early stages, a lot can still be researched.

REFERENCES

Aijmer, Karin. 2004. Pragmatic Markers in Spoken Interlanguage. *Nordic Journal of English Studies*, 3: 1, 173–190.

Aijmer, Karin. 2009. "So er I just sort I dunno I think it's just because...": A corpus study of *I don't know* and *dunno* in learners' spoken English. In Andreas H. Jucker, Daniel Schreier & Marianne Hundt (eds). *Corpora: Pragmatics and Discourse. Papers from the 29th International Conference on English Language Research on Computerized Corpora (ICAME 29), Ascona, Switzerland, 14-18 May 2008*, 151–168. Amsterdam: Rodopi.

Aijmer, Karin. 2011. Well I'm not sure I think… The use of well by non-native speakers. *International Journal of Corpus Linguistics*, 16: 2, 231–254.

Aijmer, Karin and Simon-Vandenbergen, Anne-Marie. 2011. Pragmatic markers. In Jef Verschueren and Jan-Ola Ostman (eds). *Handbook of Pragmatics*, 223–247. United States: John Benjamins Publishing Company.

Andersen, Gisle. 1997. They like wanna see like how we talk and all that. The use of *like* as a discourse marker in London teenage speech. In Magnus Ljung (ed). *Corpus-based Studies in English: Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17) Stockholm, May 15-19, 1996*, 37–48. Amsterdam: Rodopi.

Andersen, Gisle. 2001. *Pragmatic Markers and Sociolinguistic Variation: A Relevance-theoretic Approach to the Language of Adolescents*. Amsterdam and Philadelphia: John Benjamins Publishing.

Anthony, Laurence. 2020. AntConc (3.5.9). [Software]. Tokyo, Japan: Waseda University. Available at https://www.laurenceanthony.net/software/antconc/, accessed 18 May 2021.

Beltrama, Andrea and Hanink, Emily A. 2019. Marking imprecision, conveying surprise: *Like* between hedging and mirativity. *Journal of Linguistics*, 55: 1, 1–34.

Brinton, Laurel J. 1996. *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Berlin and New York: De Gruyter Mouton.

Cambridge Online Dictionary. Available at https://dictionary.cambridge.org/grammar/british-grammar/like, accessed 23 March 2021.

D'Arcy, Alexandra. 2007. Like and language ideology: Disentangling fact from fiction. *American Speech*, 82: 4, 386–419.

D'Arcy, Alexandra. 2017. *Discourse-Pragmatic Variation in Context: Eight hundred years of LIKE*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Diskin, Chloé. 2017. The use of the discourse-pragmatic marker 'like' by native and non-native speakers of English in Ireland. *Journal of Pragmatics*, 120: October, 144–157.

Fraser, Bruce. 1996. Pragmatic Markers. *Pragmatics*, 6:2, 167–190.

Granger, Sylviane. (2008). Learner corpora. In Anke Lüdeling and Merja Kytö (eds.) *Corpus Linguistics. An International Handbook*. Volume 1, 259–275. Berlin and New York: Walter de Gruyter.

Maddeaux, Ruth and Dinkin, Aaron. 2017. Is *like* like *like*?: Evaluating the same variant across multiple variables. *Linguistics Vanguard*, 3: 1, 1–15.

Magliacane, Annarita and Howard, Martin. 2019. The role of learner status in the acquisition of pragmatic markers during study abroad: The use of 'like' in L2 English. *Journal of Pragmatics*, 146: June, 72–86.

Piiri, Andreas. 2020. *A corpus based study of formulaic language use by native and*

non-native speakers. BA thesis. Department of English, University of Tartu, Tartu, Estonia.

Rahusaar, Anne. 2019. *The compilation of the spoken sub-corpus for the Tartu Corpus of Estonian Learner English*. MA thesis. Department of English, University of Tartu, Tartu, Estonia.

Santos, Giovani. 2019. Second language pragmatics: a corpus-based study of the pragmatic marker like. *Letrônica*, 12: 4, 1–16.

Schweinberger, Martin. 2015. A comparative study of the pragmatic marker *like* in Irish English and in south-eastern varieties of British English. In Carolina P. Amador-Moreno, Kevin McCafferty, Elaine Vaughan (eds). *Pragmatic Markers in Irish English*, 115–135. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Svartvik, Jan, 1980. Well in conversation. In Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (eds). *Studies in English Linguistics for Randolph Quirk*, 167–177. London: Longman.

Tagliamonte, Sali. 2005. *So* who? *Like* how? *Just* what?: Discourse markers in the conversations of Young Canadians. *Journal of Pragmatics*, 37: 11, 1896–1915.

Toom, Sandra-Leele. 2020. *The use of phrasal verbs by Estonian EFL learners: a corpus-based study*. BA thesis. Department of English, University of Tartu, Tartu, Estonia.

UCLouvain, n. d. (a). LINDSEI, available at https://uclouvain.be/en/research-institutes/ilc/cecl/lindsei.html, accessed 20 April 2021.

UCLouvain, n. d. (b). LINDSEI, available at https://uclouvain.be/en/research-institutes/ilc/cecl/lindsei-partners.html, accessed 20 April 2021.

Undo, Aare. 2018. *Calculating the error percentage of an automated part-of-speech tagger when analyzing Estonian learner English: an empirical analysis*. MA thesis. Department of English, University of Tartu, Tartu, Estonia.

APPENDIXES

Appendix 1: Total number of *like* uses by each speaker in LINDSEI-EST

Speaker 1	42
Speaker 2	8
Speaker 3	55
Speaker 4	8
Speaker 5	8
Speaker 6	9
Speaker 7	49
Speaker 8	11
Speaker 9	12
Speaker 10	30
Speaker 11	19
Speaker 12	33
Speaker 13	38
Speaker 14	14
Speaker 15	25
Speaker 16	19
Speaker 17	23
Speaker 18	16
Speaker 19	36
Speaker 20	104
Speaker 21	14
Speaker 22	34
Speaker 23	24
Speaker 24	15
Speaker 25	26
Total	**672**

Appendix 2: Total number of *like* uses by each speaker in LOCNEC

Speaker 1	15
Speaker 2	61
Speaker 3	41
Speaker 4	21
Speaker 5	12
Speaker 6	22
Speaker 7	11
Speaker 8	23
Speaker 9	24
Speaker 10	83
Speaker 11	11
Speaker 12	22
Speaker 13	32
Speaker 14	17
Speaker 15	17
Speaker 16	17
Speaker 17	21
Speaker 18	54
Speaker 19	61
Speaker 20	3
Speaker 21	16
Speaker 22	59
Speaker 23	3
Speaker 25	24
Speaker 26	30
Speaker 27	14
Speaker 28	27
Speaker 29	13
Speaker 30	6
Speaker 31	32
Speaker 33	17
Speaker 34	19
Speaker 35	23
Speaker 36	18
Speaker 37	40

Speaker 38	13
Speaker 39	23
Speaker 41	65
Speaker 42	19
Speaker 45	9
Speaker 46	47
Speaker 47	21
Speaker 48	9
Speaker 49	71
Speaker 50	22
Speaker 51	14
Speaker 52	8
Speaker 53	43
Speaker 54	31
Speaker 55	43
Total	**1,347**

RESÜMEE

TARTU ÜLIKOOL
ANGLISTIKA OSAKOND

Johanna Konso

A corpus-based study of *like* in Estonian EFL learners' speech

Korpuspõhine uurimus Eesti inglise keelt võõrkeelena õppijate *like*'i kasutusest kõnes

bakalaureusetöö
2021
Lehekülgede arv: 38

Annotatsioon:

Käesolev bakalaureusetöö uurib Eesti inglise keelt võõrkeelena (EFL) õppijate *like*'i kasutust korpusuuringu abil. Töö eesmärgiks on kindlaks teha kuidas ning mil määral erineb Eesti EFL õppijate *like*'i kasutus võrreldes inglise keelt emakeelena kõnelejatega, kuid põhirõhk on siiski *like*'i pragmaatilistel funktsioonidel.

Töö esimeses osas defineeritakse pragmaatilised markerid, õppijakeel ja -korpus ning antakse ülevaade varasematest uurimustest nii Eestis kui ka mujal maailmas. Teises osas tutvustatakse metodoloogiat, kasutatud korpuseid (LINDSEI-EST ja LOCNEC) ning defineeritakse kõik juhuvalimis analüüsitud *like*'i funktsioonid. Järgneb andmeanalüüs ja tulemuste tutvustamine ning arutelu.

Kokku oli LINDSEI-EST korpuses 672 *like*'i kasutust ja LOCNEC korpuses oli see vastavalt 1347. Eesti EFL õppijate *like*'i kasutamise suhteline sagedus oli 1985 sõna 100 000 kohta ja inglise keele emakeelena kõnelejate puhul 1139 sõna 100 000 kohta, mis näitab, et Eesti EFL õppijad ülekasutavad *like*'i. Uuriti ka *like*'i erinevaid funktsioone lähemalt, millest ilmnes, et eestlaste *like*'i kasutus on üldjoones inglise keelt emakeelena kõnelejatega sarnane. Kõige suuremad erinevused olid näha nelja kõige enam kasutatavate funktsioonid seas (tegusõna, asesõna, ning markerina kõhkluse või fookuse väljendamiseks), kus tuli välja, et Eesti EFL õppijad kasutavad vastavaid grammatilisi funktsioone liiga vähe ning pragmaatilisi liiga palju. Kõhkluse väljendamine *like*'i abil oli eestlaste seas tunduvamalt populaarsem, mis viitab sellele, et nad ei oska end veel päris emakeelena kõnelejate tasemel väljendada.

Märksõnad: Inglise keel ja keeleteadus, õppijakeel, õppijakorpus, korpusuuring, pragmaatilised markerid, *like*.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Johanna Konso,

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

   A corpus-based study of *like* in Estonian EFL learners' speech,

   mille juhendaja on Jane Klavan,

   reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Johanna Konso

25.05.2021

Autorsuse kinnitus

Kinnitan, et olen koostanud käesoleva bakalaureusetöö ise ning toonud korrektselt välja teiste autorite panuse. Töö on koostatud lähtudes Tartu Ülikooli maailma keelte ja kultuuride kolledži anglistika osakonna bakalaureusetöö nõuetest ning on kooskõlas heade akadeemiliste tavadega.

Johanna Konso

25.05.2021

Lõputöö on lubatud kaitsmisele.

Jane Klavan

25.05.2021