“Ignooraaminen voi olla iisiä mut miksi avoidaa” – Code-switching and the use of “Finglish” on Twitter

Marika Valkamaa

MA Thesis

English, Language Specialist Path

School of Languages and Translation Studies

Faculty of Humanities

University of Turku

March 2020

UNIVERSITY OF TURKU
School of Languages and Translation Studies/Faculty of Humanities

VALKAMAA, MARIKA: “Ignooraaminen voi olla iisiä mut miksi avoidaa” – Code-switching and the use of “Finglish” on Twitter

MA Thesis, 54 p., 10 app.
English, Language Specialist Degree Program
March 2020

The Internet and social media have changed people’s language use, and the effect of English language can be seen in the language use of Finnish people. The aim of this study is to examine written code-switching and the nature of language use of Finnish users on social medium Twitter. Several approaches were utilized. The Matrix Language Frame model (by Myers-Scotton) was adjusted to be suitable for the present study. Myers-Scotton’s approaches were combined with Montes-Alcalá’s, Poplack’s, and Muysken’s ideas. With addressed research questions the goal was to see the types and levels of code-switching on Twitter and explain the possible reasons for code-switching; determine and examine the Matrix Language (dominant), the Embedded Language (subordinate), and the Composite Matrix Language (ambiguous) in intra-sentential CS; and examine the code-switched words’ word class and inflection.

The data was collected from Finnish twitter accounts and different keywords to find code-switched content were used. The data consisted of 88 tweets and it was used as a micro corpus for this study. All the tweets were from public accounts that are available for anyone. Levels and types of code-switching were presented with tables and various amount of examples. The emphasis in intra-sentential CS was in the Matrix Language, Embedded Language, and Composite Matrix Language which were determined and presented with examples. Besides levels and types, the focus was on the inflection habits of code-switched lexicons and sentences. The two languages’ code-switched words’ grammatical structures were examined with various amount of examples.

The results revealed several cases of code-switching both from English to Finnish and from Finnish to English. Even though all the tweeters are presumably L1 Finnish speakers, tweets with code-switching in both languages were found. However, Finnish was the Matrix Language in the majority of tweets, whereas English was the Embedded Language. Both, different types and levels of code-switching were found. Code-switching patterns occurred mainly consistently and the inflection would mostly follow the Matrix Language’s grammar structure. However, findings revealed exceptions to this trend as well. According to the present study, written code-switching on social media is not arbitrary but follows similar patterns than that of spoken code-switching patterns. Further research ideas for the future would be exploring social media platforms and use other approaches, i.e. Conversational Analysis, to understand social interaction online better. Also, the nature of language use could be compared between different social media.

Key words: code-switching, matrix language frame model, social media, Twitter, Finglish

Table of contents

List of Tables

List of abbreviations

CL = Composite Matrix Language

CMC = Computer-mediated Communication

CMD = Computer-mediated Discourse

CS = Code-switching

EL = Embedded Language

L1 = First language or Native language

L2 = Second language

MLF = Matrix Language Frame model

ML = Matrix Language

1 Introduction

In today's world people are communicating on the Internet in several different languages. Still, if willing to communicate with a person from a foreign country, English is usually seen as the lingua franca, and used the most for online communication. Gumberz defines *communication* as follows: one person talks and another person adds a response (Gumperz 1982, 1). This kind of action can happen between two or several people. To participate in such verbal interaction, the interlocutors should know the rules of the spoken language, understand the semantics behind it, and have the knowledge beyond grammatical competence of the language (Gumperz 1982, 1). Peter Gärdenfors (1999, 19) says: "cognitive semantics identifies meaning of expressions with mental entities". A person can often do this not only in one but also in two languages, sometimes even many more. The phenomenon of *bilingualism* is a common occurrence around the world, including Finland (Montes-Alcalá 2005, 6). A bilingual is a person who has the knowledge of two vocabulary and grammatical systems, and hence can fluently speak two languages (ibid.). Bilingualism can be defined various ways and some require that a person needs to be native in two languages in order to call themselves bilinguals. However, in this study, Montes-Alcala´s definition will be followed, and hence anyone who has the fluency in two languages can be considered as a bilingual. When a bilingual alternates between two languages in the same conversation, we talk about a phenomenon called *code-switching* (Montes-Alcalá 2005, 6). Code-switching has been broadly studied, and researchers have focused on its form, meaning, and grammatical patterns, but mostly in the spoken discourse (Sebba 2012).

Online communication has changed the way people use language. Communication through online networks involves adopting same words and phrases that we use in real life conversations (Wood & Smith 2004). People 'surf', 'meet', and 'send' online even though literally they are sitting next to their computers and not actually doing those things (ibid.). Wood and Smith claim that online communication is less personal and less powerful than traditional modes of discourse because the computer-mediated communication (CMC) is exchanging information through machines (ibid.). Some people are worried that CMC is dangerous, because one can never know who they are really talking to online. On the other hand, Internet has enabled new channels of communication which makes interaction possible without the need for physical presence (ibid.).

The present study focuses on online writing and code-switching in written discourse. The focus will be on written online code-switching, in particular, I will study Finnish-English bilinguals and their use of code-switching in social medium, Twitter. Even though online interaction is often seen in-between mode (between spoken and written) (Foertsch 1995), in this study Twitter is seen as a written discourse because the focus is on the language structure of the tweets, rather than on the interaction. The aim is to study the levels on which code-switching occurs, what types of code-switching are common, and how code-switching affects inflection of the words. Inflection is studied whether it follows the grammatical structure of the dominant language or the subordinate language. The variation between different word classes is also examined. The present study explains the nature of the language use on social media adding perspectives from the phenomenon of code-switching, and considers its pragmatic use: what are the possible functions of people using code-switching in their tweets.

Sometimes, speakers flip from one language to another without even acknowledging that they are doing so. Unintentional L1 code-switching into L2 speech is usually a sign from low proficiency in L2 (Lipski 2016, 144). Written code-switching is usually more intentional because written language is thought more thoroughly than spoken language. This is true in the present study as well even though tweets are usually less thought through than, for example, legal texts or formal e-mails. Rather often people who write on Twitter use words and code-switching in order to achieve something. It can be a sign of language of the Young, solution for forgetting a word in another language, or even to express a sense of humor. There are a number of possible reasons for code-switching to happen. The British linguist David Crystal (1987) mentions a few: code-switching is seen as a sign of solidarity within a group (assuming that all speakers in a conversation are bilingual). Rapport is established between the speakers when the interlocutor might respond with a similar switch (ibid.). In many situations code-switching is done deliberately to both create unity or to exclude someone from a conversation (ibid.). A speaker may not be able to express themselves in one language so they switch to the other to compensate. The final reason by Crystal is the alteration that occurs when the speaker wants to convey their attitude. This kind of switch can be seen as a socio-linguistic tool by bilinguals (Crystal 1987).

The use of code-switching is very common online, especially on social media. It occurs in various ways and often non-native English people add English to their text to convey

different things, for example a bigger audience to reach the post. Guzman et al. (2016) say: "Multilingual documents may comprise more than one language for various reasons, including translation, change of author/speaker, use of loanwords, and code-switching." Even if the text passage itself is written on Twitter in the writer's own mother tongue, the tags marked by a hashtag (#) are often in English. As Cambridge English Dictionary states: a hashtag is "used on social media for describing the general subject of a tweet or other post (= message)" (*CED*, s.v. "hashtag" n.). The reason could be that the writer wants to achieve a bigger audience and thus more attention to what they have written, or reach people who have the same object of interest.

The present thesis is a descriptive qualitative research about the nature of language mixing on social media. The focus is how native Finnish people use mixed Finnish and English (also known as Finglish) in their short text passages, tweets. In this thesis, Finglish means text which has code-switched content in Finnish and English. Some may say that English text where code-switched Finnish words occur could be called "Engnish", but since the word is not commonly used, merely the word Finglish will be used. The study focuses on types of code-switching that are, according to Poplack (1980): *intra-sentential* (within a sentence), *inter-sentential* (between sentences), and *extra-sentential* also known as *tag-switching.* Besides these types, the focus will also be on which levels code-switching happens as the linguist Cecilia Montes-Alcalá, a specialist in bilingualism and language contact, divides them: word-level, clause-level, and sentence-level. After types and levels the focus will be on code-switching within sentences. For intra-sentential code-switching parts of the linguist Carol Myers-Scotton's, expert of code-switching, Matrix Language Frame model will be used to examine what kind of code-switching occurs within a sentence and how the dominant and *Embedded Language* work with each other. Lastly, the present study concentrates on whether code-switching happens by insertion, alternation or congruent lexicalization, and how these differ between different word classes. This division is taken from a Dutch linguist Pieter Muysken. All the theories and concepts will be introduced and explained in the section 3 Theoretical Framework. The theories support each other and overlap in some ways, and in order to fulfill the aim of the study the addressed research questions are the following:

1. Which types and levels of code-switching occur in tweets by Finnish Twitter users, and what possibly motivates people using code-switching?

With this questions the interest is to examine the types and levels where code-switching occurs. The tweets will be divided according to their types and levels and a numeric data of them will be presented. The focus will also be on tweets that have code-switching mixed in different levels and types within a tweet. Montes-Alcala's and Muysken's theories will be combined to analyze the code-switched content.

2. What type of intra-sentential code-switching occurs online by Finnish Twitter users, according to the Matrix Language Frame model?

Intra-sentential code-switching is taken as a separate analyzable unit and Myers-Scotton's Matrix Language Frame model will be used in the analysis. The model is rather vast so that is why the model will be adjusted suitable for the present study. The emphasis will be on Matrix Language and Embedded Language, ambiguous combination of the languages (=Composite Matrix Language), and the grammatical inflection of the code-switched words. In the analysis Myers-Scotton's theory are combined with Muysken's ideas.

3. What kinds of words are being code-switched in tweets and how does it affect inflection?

The code-switched words will be categorized according to their word class, and a table is presented to show the variation. The analysis will also be on how the word class affects the inflection: is the word inflected according to the original language or the switched language? Also, it will be discussed how the written form changes when the word is inflected. Muysken's division of insertion, alternation, and congruent lexicalization will be utilized in the analysis.

In connection to these questions, possible reasons for **why** code-switching occurs will be examined and discussed. The hypotheses are that written code-switching has patterns and despite the tweeter (writer who posts the tweet) similar patterns can be found, and written code-switching is somewhat similar to spoken code-switching. The hypotheses rely on previous studies that have been made utilizing the same framework. For example, Callahan (2002) analyzed a corpus of bilingual novels using the Matrix Language Frame model. Her findings showed that CS followed the same patterns that had been found in spoken CS (Callahan 2002).

This study will start by explaining general background of English language use in Finland, including bilingualism and English use among native Finnish people. In addition, social

network, Twitter and language use online will be presented and explained. After this the meaning of code-switching and its occurrence in written and spoken texts will be explained. This follows by theoretical framework section where all the theories and researchers behind them will be introduced and explained. Grammatical terminology is also introduced briefly. After the theoretical part the collected data will be shown and the methodology will be explained. This follows by analysis, results & discussion, and finally, conclusion with further research ideas related to the present topic.

2 Background

This section discusses the general proficiency of English language use in Finland, the meaning of social network, and language use and communication online

2.1 The English language in Finland

English and Finnish are rather different languages. One of the most obvious differences between them is the inflectional morphology (Halmari 1997, 33). According to Halmari, English has only eight inflectional morphemes, whereas in Finnish both verbs and nouns acquire a huge amount of inflectional suffixes (ibid.). Other differences between the languages are for example articles, prepositions, and pronunciation. In English articles and prepositions are in use whereas Finnish inflects words by adding suffixes at the end of the words. These terms will be thoroughly explained in the section 3.2. However, like in many countries, despite of the linguistic differences, people in Finland increasingly use English in their lives.

Like many bilinguals (or multilinguals), Finnish people mix Finnish language with other languages. The most common languages to mix Finnish with are Swedish and English. Swedish has always had a remarkable role in Finland since Swedish is one of Finland's official languages, and in 2018 approximately 5.2 % of people living in Finland spoke it as their mother tongue (Statistics Finland). However, the role of English has increased tremendously, and even more ever since social media became a part of people's everyday living. People write in English more than before and use it to communicate with other people all around the world. Finnish people are interested in the English language and the motivation to study it is bigger than, for example, 50 years ago.

Finnish are tend to be rather shy to speak English because the pronunciation has been a joke amongst journalists. Hence, a term "rally English" (=rallienglanti) was born. This term was used as a keyword to gather material on Twitter amongst other terms that will be introduced later. Rally English means speaking English with a strong Finnish, monotone accent and sometimes translating Finnish phrases literally that do not necessarily mean anything in English. Nowadays the term refers not only to rally drivers but to any Finn who speaks English with a strong Finnish accent. There has been a lot of discussion how accents are a positive thing because it tells that the speaker speaks another language too and hence is a bilingual. Research shows that the role of English in Finland has changed (Leppänen 2007), and people

are braver to use it more than before. Especially young people tend to use English as a lingua franca, alongside with Finnish. Nowadays people may use English for constructing identity and communality (ibid.). Because Finnish people use and hear English more than ever, thanks to TV, e-sports, and popular culture, just to mention a few, many Finnish people can be considered bilinguals. Social network is an important tool for communication, and the next section will explain the general concept of social network.

2.2 Social network

Like mentioned before, social media, as a phenomenon has spread into our lives and it has affected tremendously how we use language online. Social network is an online community of individuals who exchange messages, share information, and cooperate on joint activities. Cambridge dictionary gives the following explanation: "a website or computer program that allows people to communicate and share information on the internet using a computer or mobile phone" (*CED*, s.v. "social network" n.). Many people have joined to social networking sites where members create and maintain personal profiles that they link with other people. The resulting network of "friends", "contacts", or "followers" who have similar interests, business goals, or academic courses have replaced older concepts of community for many people, especially for young people (britannica.com). Social network includes many different kinds of services or platforms, for example Facebook, Twitter, and YouTube just to mention a few. Services require registration after which the users are free to use the service communicating with each other. Many services have specialized in some certain function, for example Instagram for sharing pictures to friends and LinkedIn to use as a mobile reference for employers. Social network may have started as a young people's use of communication but nowadays it is a part of many people's lives, despite the age.

Like mentioned above, social media platforms serve different functions, and the present study focuses on discourse written on Twitter. Twitter is an online social networking and microblogging service. It was founded in 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams. In the launching, Dorsey underlined that Twitter wants its users to "shape the platform into a generic infrastructure for online communication and social inter-action". In 2019, Twitter was estimated to have 330 million active users in which 134 million use it on a daily basis. Twitter users are tend to be more mature than for example Snapchat or Instagram, and in 2018, The Video Advertising Bureau (theVAB) said that 63 % of the Twitter users were

between 35–65 years old. Users communicate in Twitter by “tweeting”. A tweet is a 280-character-long brief message. A user types a tweet via mobile phone keypad or computer and sends it to Twitter’s server, which relays it to a list of other users (also known as followers). In addition, users can choose to follow specific topics, creating a dialogue of a certain topic and pushing the number of followers in a given Twitter feed into the millions (britannica.com). It is also common that a tweeter wants to share another tweeter’s text passage to their own followers. They can do this by re-tweeting the original tweet which means that the original tweet appears in the re-tweeters list of posts. Nowadays, Twitter is used not only by ordinary people, but also journalists, politicians, and many influencers. Twitter has affected people’s language use, and the next section will take a closer look at communication and language online.

2.3 Online communication and language

In global communication, the Internet has had a great impact on language use (Warschauer et al. 2002). Language used online can be called interactive written discourse or electronic language, sometimes a term netspeak is used (Baron 2010, 11). Electronic communication can be divided into two dimensions: synchronicity and audience scope. Synchronous means the communication happens in real time, whereas asynchronous means the reader can open the message whenever it is convenient for them. Audience scope means whether the message is intended for a single person or for a larger audience (Baron 2010, 14). Table 1 is an altered version of Barons table made specifically for this study.

Table 1 Dimensions of electronic communication.

	ASYNCHRONOUS	SYNCHRONOUS
ONE-TO-ONE	direct message, text message	instant messaging
ONE-TO-MANY	Facebook, Instagram, Twitter	chat groups, stream chats

As seen from the Table 1, emails and text messages are usually meant from one person to another, and when the message has been sent, the reader can open it whenever they are willing to. One-to-one instant messaging can mean, for example, companies help services, where a customer can use the chat and ask about any problem. Someone from the company replies immediately and thus the communication happens synchronously. One-to-many

messaging can be posting something on Facebook, Twitter, or Instagram, and it is usually posted for anyone who wants to read it whenever they want. Synchronous messaging, however, occurs often for example if a person is live streaming a video and the viewers can interact in real time. Nowadays messaging has developed and many devices show if a person has read the message or if they are typing a message. Also, deleting messages is easier than before. Years ago, a text message could not be deleted to prevent another person to see it, but nowadays it is possible.

Susan C. Herring introduced a term *computer-mediated discourse* (CMD) (2001). The study developed alongside of interactive networking, and the scholars saw communication in the new medium (Herring 2001, 612). Computer-mediated discourse can be seen in different media and it can be suited with the Table 1 above. Herring claims that online writing has changed our language use and the Internet has created new fractured language (ibid.). When observing online one can notice that this is true and it can be acknowledged when reading texts on the Internet. This also means that languages change all the time, and Internet jargon has become its own category of discourse.

Language use on the Internet occurred a critical change, at the latest, when the use of smart phones started to spread. Mobile phones are not just for calling anymore, but they have many function: People talk with each other, send messages, play games, shop online, listen to music, and watch videos. A big use of mobile phone is, of course, text messaging with other people. Instant texting has caused the fact that language used online is often narrower and more diminished than spoken language (Herring 2001, 612). Online communication, such as tweeting, has also created a new form of language, its own dialects. Mobile phones introduced us the abbreviations (fab=fabulous), acronyms (brb=be right back), and emoticons (smileys, such as ":)") (Baron 2010, 27–28). In addition, people have various ways to express different activities online. For example, laughing can be expressed in different ways, depending on the spoken language. English speaking community uses LOL-acronym that comes from the words "laughing out loud" or just simply "HAHA". In Russia the same thing is expressed by using "XAXAXA", Spanish "JAJAJA", Japanese "wwww", and in Thailand "5555". These characters are usually connected to the language's pronunciation.

Language used on Twitter is often informal and less planned as it is when writing a formal text, like an essay. On Twitter, writing happens impulsively, it can be a quick thought that comes to one's mind, or a reply to someone one strongly agrees or disagrees.

Typographical and grammatical errors are common, because the main function is to get to share the information or opinion. Because Twitter has limited amount of characters to use, this affects the length of the written passages. The length limitation encourages people to use abbreviations, or other ways express themselves within the limited character amount. Hashtags (#) are one way people communicate on Twitter. They are used to call attention to something. As mentioned before, with hashtags all the posts with the certain subject can be found when using the hashtag as a keyword. For example, if someone posts a picture of their cat and captions is #cat, and another person searches by using #cat, all the pictures with the hashtag appear on the list on the screen. Dialect words spread rapidly on Twitter as well. People encounter new words that they would not have otherwise encountered. Terms like 'lush' or 'lit' have spread nationwide because of social networking. In addition, the term 'trolling' means more than it did before. Online trolling means that a person deliberately makes controversial comment to provoke emotional reaction (Fichman & Sanfilipo 2016, 6). This means Twitter has created new terminology as well. The next section will present the theoretical framework of the present study.

3 Theoretical framework

In this section, the theoretical background will be introduced and key concepts and theories related to the present study will be explained. First, code-switching and its functions are introduced and explained with examples. Then, code-switching as a sociolinguistic phenomenon will be explained briefly. At the end of this section, the theory behind the used methods will be explained and some of used terms in the analysis will be presented.

3.1 Code-switching

As mentioned before, bilinguals or multilinguals speak fluently two or more languages. It is common that when a person knows many languages, they start to switch parts of different languages in their speech. Code-switching is a phenomenon where a speaker does that within a discourse. It can be done consciously or unconsciously, but the switches usually happen fluently between two or more different languages. CS can have grammatical and pragmatic perspectives (Blom and Gumperz 1972).

John Gumperz was an American linguist and a pioneer in the study of code-switching. His approach to code-switching can divided into four concepts: *contextualization, contextualization cue, situational vs. metaphorical code-switching,* and *we vs. they code* (Mäkilähde 2019, 75). Contextualization means how speakers and listeners use verbal and non-verbal signs to relate what is said. Contextualization cue means "any feature of linguistic form that contributes to the signaling of contextual presuppositions" (Gumperz 1982, 131). Gumperz explains how code-switching has different functions: *situational and metaphorical code-switching* (1982, 60). The alternate choice of languages controlled by components such as topic, setting and participants is said to be *situational switching,* and usually this kind of code-switching has a pragmatic perspective (ibid.). The choice of language in which it takes place within a single conversation without any change of major factor in the interaction is called *metaphorical* or *conversational code-switching* (ibid.). Gumperz (1977, 1) says: "By conversational code-switching, I refer to the juxtaposition of passages of speech belonging to two different grammatical systems or sub-systems, within the same exchange." This means that a person starts talking in one language but at some point they change the language to another and continue speaking with that: "Hei äiti*, look at me*!" ("Hey mom, look at me!") Usually, the switch occurs fluently without pauses, hesitations, or changes in the rhythm or

pitch level to mark the switch of a code (Gumperz 1982, 59–60). Still, all the interlocutors have no difficulties in understanding each other. This kind of code-switching is very common, for example, in bilingual families, where a child has learnt two languages at the same time from their parents. Gumperz continues that the conversation has all the same aspects compared to a single-language-conversation (ibid.). Situational code-switching happens in a certain setting or the code is changed due to the addressee or activity whereas metaphorical code-switching is usually less planned and more intuitive (ibid.). Gumperz uses the term word choice, which means that the speaker chooses a word/phrase from several options. More about word and language choice will be told in the section 3.1.2. By *we code and they code* Gumperz (1982, 66) explained as follows: we code is a minority language used in informal situations whereas they code is the majority language that is less personal and formal. An example would be Finnish vs. Swedish language when Finland was under the Swedish rule. The majority of people living in Finland spoke Finnish with each other but all the formal situations happened in Swedish. Another example would be Blom's and Gumperz's (1972, 425) notion of clerks and residents in the bilingual community. In the example, greetings take place in the local dialect, but business is talked in the standard language (ibid.). Myers-Scotton (1979) agrees that language choice involves an individual's identification with a social role or status relative to the interlocutor's status. For this kind of behavior, she uses a term *transactional code-switching* (ibid.).

Present-day code-switching is predominantly seen as a phenomenon of spoken language and that is why research has more focused on speech. However, code-switching in the earlier periods is studied in written sources but it cannot be said whether code-switching used to be a phenomenon of the written medium only or if it was a spoken phenomenon as well (Schendl & Wright 2011, 27). Nowadays social media offer excellent platform to study code-switching in written text, and hence written code-switching can be noted equivalent phenomenon to spoken code-switching. This is one of the reasons why this topic was chosen for the present study, and it makes this study contemporary and important. Social factors are the source of variation in CS patterns (Gardner-Chloros, 2009). The same language pairings can be combined in various ways and social media platforms offer plenty of data to examine this phenomenon. However, one problematic aspect is that the researchers have sometimes found the term code-switching problematic. Different fields of research have studied the phenomenon (psychology, linguistics, socio-linguistics, ethnography etc.) and they have used

the same term and hence it has been understood different ways. Some researchers have offered the term language choice to be used, but since code-switching is a stabilized term in the field of linguistics, that will be used in the present study.

There are a number of possible reasons for switching from one language to another. It may be to redefine the interaction as appropriate to a different social arena, or to avoid through continual code-switching, defining the interaction in terms of any social arena (Romaine 1995). It commonly occurs when an individual wants to express solidarity with a certain group (ibid.). Avoidance is also important to notice because code-switching often serves as a strategy of neutrality and a speaker takes into consideration which code is the most appropriate in a particular situation (ibid.). As mentioned before, Crystal (1987) gives many reasons for code-switching to happen. The first one is that a speaker may not be able to express themselves in one language, and hence they switch to another one to compensate the deficiency (ibid.). The second reason may be to express solidarity with a certain group (ibid.). The third reason is to convey attitude (ibid.). All these reasons are notable in the present study. People who code-switch their language on Twitter usually do it to express solidarity, attitude or just cannot find the right expression in Finnish. This will be discussed more in the analysis and discussion sections. Gumperz's explanation for reason of code-switching is: "Depending on such factors as region of origin, local residence, social class and occupational niche, each communicating subgroup tends to establish its own conventions with respect to both borrowing and code switching." (Gumperz 1982, 68).

Nowadays people interact in many ways in the Internet and hence code-switching occurs when talking to friends online. For example if two Finnish people talk about a game that is played in English, the terms are automatically easier to talk about in the original language. This kind of jargon talk is really common in specific game or hobby groups. Online interaction and social media have increased code-switching in writing. My data consists of many examples of this, for example:

(1) "…kotona lapset *evolvaa, transferraa* ja *poweruppaa* Pokemoneja"
(…"at home children evolve, transfer and power-up Pokémon").

The mobile game Pokemon Go! is completely in English, so Finnish children obviously use those terms fluently in English. Children might not even know the Finnish equivalent to the words, or even the literal meaning, but they know the function of the words when they are

used for the mutual knowledge of the game. This same kind of jargon talk may occur when a sport is reported on TV. One example in my data is from snowboarding:

(2) "*Svitsillä* ekaan *hittiin* ja *gräbi* ennen *ländäystä*. Kova *rani*."
("With switch to the first hit and a grab before landing. What a run!")

For someone who does not follow this sport actively, the Finnish sentence may not make any sense. However, in the snowboarding community these terms are used frequently and it would be weirder to use their Finnish equivalents. For those people the sentence is completely understandable. A third example of such jargon from my data is the following:

(3) "Aaaaah. Sain maalattua 8 *goblinia* valmiiks. Voittaja fiilis! Enää ois 2 *perytonia* viellä.." ("Uh! I just finished painting 8 of the goblins. What a feeling! Still, 2 perytons left..")

To be honest, the meaning of a peryton was not clear at the beginning. After reading several online game forums it became clear that a peryton is a "mythological hybrid animal, half stag and half bird". The tweeter together with their friends must understand the tweet without hesitations, but by choosing to write goblin (and not its Finnish equivalent "menninkäinen") and peryton, the tweeter excluded all of those who were outside of the target audience of the tweet.

Previously introduced examples were about how Finnish people use code-switching in writing when the terminology is originally from an English source. This same kind of code-switching would presumably appear in their speech. However, because Twitter allows to write only 280-characer-long texts, it is common that no such code-switched words are explained in any further way. One reason for code-switching, which was explained before, was to use a word that is more suitable in the situation than the same word in another language. In speech it is easy to use a word and then explain it if someone looks confused but in writing the writer makes the decision once writing it, and chooses to include or exclude people. Online writing allows this and if someone wants to, they could ask for the explanation. When collecting the data, some comments were noticed under the tweets that were criticizing a tweet where code-switching happened:

(4) "oh you are soo international and cannot even speak Finnish anymore".

When choosing to use an English word, it might actually irritate people. However, Twitter does not provide the space to explain a code-switched word because of the space limitation. Of course, it is possible to continue and write another tweet but Twitter is meant to be a service where one can quickly read and move on. Hence, writing two tweets in a row breaks the social media "rules".

When searching the information about previous studies about Finnish-English code-switching, it was found that Finnish-English code-switching has been examined in the nineteen sixties by Lehtinen (1966). Her study was a descriptive study of the English dominant Finnish speaker. Also the role of code-switching has been examined among the bilinguals in Canada, and borrowing in the language of American Finns Martin 1988 and Lauttamus (1991). Finnish-English code-switching has also been studied by Poplack, Wheeler, and Westwood (1989). According to them: "the social role of language mixing is propitious to the smooth integration of elements of both codes, typological considerations are predictive of the types of mixture." The Matrix Language Frame model and Muysken's categorization (insertion, alternation, and congruent lexicalization) has been recently used in Finnish study, when studying code-switching between different Saami languages. Finnish-English code-switching is not a new phenomenon but as the present study focuses on code-switching on social media, it brings a new perspective to the linguistic field.

3.1.1 Forms of code-switching in linguistics

Sociolinguist Ho-Dac Tuc (2003) explains the different approaches that have studied code-switching: **The structural** approach postulates language-specific or universal models of patterns in code-switching. **The psychological** approach examines the interaction of the two grammars involved in a code-switched text (spoken or written). The speaker's intentions are studied in **the ethnographic** approach. **The attitudinal** approach emphasizes attitudes toward code-switching, and **the functional** approach posits the conversational functions of code-switching. As Ho-Dac (2003) says:

> In general, a range of approaches to code-switching has led researchers to an agreement that this linguistic phenomenon cannot be dismissed as arbitrary, and that code-switching can be approached from different angles concurrently with one another (Ho-Dac 2003).

The present study has aspects of all the five approaches: The structure and grammar are studied. Intentions and attitudes are in minor role of the study but functions are discussed as well.

The forms of code-switching in linguistics vary and next, some of them will be introduced. Code-switching is, of course*, a syntactic phenomenon* because researchers want to understand how languages form structural wholes even though two or more languages appear in the same spoken or written text. Researchers have taken different points of view when studying code-switching and parts of several researchers' ideas and models will be utilized in the present study. One of them is Professor Shana Poplack's (1980) used division. She divides code-switching into three types*: inter-sentential, intra-sentential, and extra-sentential*, the latter also known as *tag-switching*. In inter-sentential code switching, the language switch is done at sentence boundaries—words or phrases at the beginning or end of a sentence are code-switched (ibid). This type is seen the most often in fluent bilingual speakers. Finnish-English example from my data:

(5) "ytl. Ole suopea mulle. I deserve this."
("ytl. Be good to me. I deserve this."

In intra-sentential code-switching the shift is done in the middle of a sentence, with no interruptions, hesitations, or pauses to indicate a shift (Poplack 1980). The speaker is usually unaware of the shift (ibid.). Intra-sentential code-switching occurs in the clause-level and in the word-level. English-Finnish example from my data:

(6) "Coffee and *pulla* is available at 13:00."
("Coffee and bun is available at 13:00.")

In Extra-Sentential or Tag Switching the switching of either a single word or a tag phrase (or both) switch from one language to another. This type is common in intra-sentential switches. It involves the alternation of a tag from one language into an utterance in another language. For example, Finnish people use some boundary words like "niin kuin" or "niinku" (like) or "tuota" (well) in the middle of a sentence while speaking English. Tags can also be settled phrases, sayings, or acronyms. Acronyms were found in the collected data:

(7) "Joku pitkä elokuva ja teevee sanoo stop liian jännää *wtf*"
("A long movie and the TV says stop too exciting wtf")

Like mentioned earlier, code-switching has been studied in various fields and a lot of researchers have tried to form a model that could be used when code-switching is studied. In the early efforts to describe tendencies e.g. Gumperz (1976) offered where code-switching could and could not occur. These were for example between pronominal subjects and verbs. Poplack (1980) explained her *Equivalence Constraint:* "switched sentences are made up of concatenated fragments of alternating languages, each of which is grammatical in the language of its provenance". The boundary between adjacent fragments occurs between two constituents that are ordered in the same way in both languages, the linear coherence of sentence structure is ensured without duplicating the content (Poplack 1980).

According to many other researchers later on, code-switching has a pattern and it rarely happens randomly (Muysken 1995, 177). Still, according to Muysken (1995): "various models have been proposed and tested with the result that some cases appear to fall under one constraint, and others under another." This may be why many researchers combine different models when studying code-switching, in order to find consistency for the phenomenon. In the present study, the same action will be done because one complete model does not suit for the aim of the present study. Muysken (1995, 180) himself divides code-switching into *insertion, alternation, and congruent lexicalization*. By *insertion* he means that an element from one language to another is embedded within the structure of another (ibid.). For example, a code-switched English word is also inflected like a Finnish word. The code-switched word follows Finnish structure. The phrase structure, including the order and type of constituents, is determined by the base language (Lipski 2009, 2). This kind of code switching usually happens spontaneously and it is similar to lexical borrowing (Muysken 1995). An example from the collected data:

(8) "Kohta tuo *turnaa* zombieksi."
("Soon he will turn into a zombie.")

Alternation Muysken explains to be two separate structures from different languages that are juxtaposed. A true switch from one language to another happens here involving both languages' grammar and lexicon. First segment is not embedded in the second one nor vice versa. This kind of code switching occurs between utterances in a turn or between turns (ibid.)

(9) "*Look out,* talvi tulee!"
("Look out, the winter is coming!")

Congruent lexicalization, however, means that two languages share a grammatical structure but CS happens in word-level (Juutinen 2019, 166). Both of the languages share a grammatical structure which can be filled with words from either language (ibid.). The going back and forth suggests that the elements from the two languages are inserted, as constituents or as words, into a shared structure. This means that the two languages should be structurally to a very high degree (Lipski 2016, 152). This is rare between Finnish and English, since the two languages are rather different from each other. The following example would be an example of congruent lexicalization:

(10) "Tiedätkö sinä where Laura is?"
(Do you know where Laura is?"

The word order would be the same in both languages. The CS happens intra-sententially but only the Finnish words have been replaced by the English equivalents. Finnish and English grammatical structure and vocabulary are completely different, so the congruent lexicalization happens here merely with the word order.

The process of alternation is frequent in communities with a tradition of language separation but occurs in other kinds of communities as well. Researchers have said that it might be "the most frequent and least structurally intrusive type of code-switching" (Muysken 1997, 364). Insertion is frequent in areas, for example, in previous colonial settings where two or more languages are in an asymmetric position. The shift may be from insertion into the language of the origin to insertion into the language of the host country (ibid.). Congruent lexicalization may happen when a bilingual speaker speaks two languages that have equal prestige (ibid.). Muysken says:

> When several constituents in a row are switched, which together do not form one unique constituent, insertion is not plausible and congruent lexicalization is a serious possibility. A number of elements form a unique constituent if that constituent contains no other elements. With several constituents, we would have to assume multiple contiguous insertions. When the switched element is a single, well-defined constituent, e.g. a noun phrase or a prepositional phrase, insertion is a plausible option; this holds a fortiori for single words. When several words are switched which do not form one or more constituents together, congruent lexicalization is plausible. (Muysken 1997, 365)

In the present study the tweets are considered by utilizing Muysken's division and more examples will be given in the analysis section.

3.1.2 Levels of code-switching

Besides Muyskens' ideas, another approach of the present study is to examine in what kinds of levels code-switching occurs in written language. Cecilia Montes-Alcalá divides the levels into word-level, clause-level, and sentence-level. Word-level means that only a word is code-switched in the sentence that is otherwise spoken/written in another language. Clause-level means that a certain clause is code-switched, and sentence-level means that sentences in two languages are spoken/written sequentially.

As already mentioned before, code-switching is traditionally divided into situational or metaphoric. Situational code-switching is used if one of the speakers does not understand, and hence code-switching is necessary. Other term to describe such action is *language choice* (Montes-Alcalá). Language choice is, of course, always present on Twitter because the tweeter has time to think what to write. Sometimes the situation may prompt the tweeter to make a certain choice, but it is possible that code-switching may occur unconsciously on tweets also. Metaphoric code-switching fulfills empathic functions, and this is concerned in the present study, since the text passages examined have no physical parties. Metaphoric code-switching is usually used when it helps to make the point clearer, or to semantically elaborate its counterpart. An example of metaphoric, word-level code-switching would be:

(11) "Tää on niin hienoo, oon niin *excited*!"
("This is so great, I am so excited!")

It is important to understand that language contact and its levels are present whenever code-switching happens, because two or more languages are in contact (Montes-Alcala 2005, 12). Language contact can be divided in different categories. *Phonological levels* are also known as accents that occur in speech. It is shown in intonation, stress, and pronunciation (Montes-Alcala 2005, 11). Another interference happens in *syntactic level.* This means that the interference is seen in word order, for example, Finnish is spoken but in English word order. (ibid.) *Morphological level* affects for example numbers markers (singular, plural) (ibid.). *Lexical levels* include borrowing. Borrowing happens when a word is taken from one language to fill a gap in

the other language's lexicon. These words are usually adjusted to the other language's form (ibid.) Examples of borrowing in Finnish would be "banaani" and "prinsessa" (banana, princess). This means that a word may have been a switch many years ago, but nowadays the word is a part of both languages' lexicon and hence we talk about borrowing, not switching. Borrowing and code-switching are different processes as Gankoff, Poplack, and Vanniarajan explain:

> Code-switching within the confines of a single sentence requires access to the syntactic apparatus of both languages, because, as is generally observed, each of the monolingual fragments making up a codeswitched sentence is internally grammatical by the rules of its language. Borrowing on the other hand operates independently of the grammar of the donor language, though it may involve lexical items from that language that are not yet incorporated into the monolingual vocabulary of the host language, and these items may retain aspects of the donor language phonology. (Gankoff, Poplack, and Vanniarajan 1990, 72).

They do admit, though, that code-switching and borrowing have the same outcome: "For example, a code-switch consisting of a single noun in one language within a sentence entirely in the other language may be superficially indistinguishable from a borrowing." (ibid.) Myers-Scotton says that there are two kinds of borrowing: cultural and core borrowing. Cultural borrowing is taking a word from a culture that does not have an equivalent in the other culture (Myers-Scotton 2002, 41). Core borrowing is a word that duplicates already existing word in the L1 (ibid.). In the present study the focus on borrowing will be diminished. The main focus in on code-switching and because Finnish and English are such different languages, code-switching should be easily separated from borrowing.

3.1.3 The Matrix Language Frame Model

Myers-Scotton has formed many models for studying code-switching. Her Markedness Model states that people choose language based on the relation with other interlocutors which they wish to have in place. Because the present study is not mainly pragmatic, Myers-Scotton's Matrix Language Frame Model (MLF) will be used. Matrix Language Frame Model is one of the most popular models when it comes to insertional code-switching (Winford 2003). Myers-Scotton (2001) claims that when we speak (or write), one language is the *Matrix Language* (ML) that is the base language. However, elements of the *Embedded Language* (EL) are inserted into the frame of the Matrix Language. According to the MLF, the grammatical morphemes are from the

base-language. Myers-Scotton says: “When two or more varieties come together within a single bilingual constituent, they do not participate equally.”(Myers-Scotton 2001, 23). This means that ML is the source of the grammatical frame and EL can contribute limited material, meaning that languages have a hierarchical relationship. An example from the data:

(12) “Teidän podcastin influenssit näkyy jo.”
(”The influence of your podcast is showing already.”)

Here, Finnish is the Matrix Language and English is the Embedded Language. English words “podcast” and “influence” have taken grammatical form from Finnish and they are inflected like Finnish words by adding suffix –n to mark the genitive and –t to mark the plurality.

The model predicts which utterances containing code-switching can be grammatically well-formed and therefore may occur in the speech. The ungrammatical utterances are not supposed to occur, unless they are stylistically marked or have, for example, emphasizing function (Myers-Scotton 1993a, 75). Hence, even though one language is the main source of the grammatical frame, the other language can contribute grammatical aspects as well. (Myers-Scotton 2001, 25). This happens when a language’s position in a speaker’s repertoire is diminishing and that of another language is gaining ground. (ibid.) This is called Composite Matrix Language and here a sentence can have aspects from both grammars (ibid.) For example:

(13) “Tän kirjan lukeminen *is so calming*.”
(“Reading this book *is so calming.”*)

The beginning of the sentence is in Finnish with Finnish grammar but the ending is in English and follows English grammar instead of Finnish grammar. The main assumption of Matrix Language Frame Model is that languages cannot have symmetrical relationship. One is always dominant and supplies the morphosyntactic frame (Myers-Scotton 1993b, 35). This means that the interaction of morphology and syntax is generally from the Matrix Language. The Embedded Language has an auxiliary function and it supplies *content morphemes* which are embedded into the Matrix Language (ibid.). These will be explained more in the next section.

Myers-Scotton says that there are several criteria to determine the Matrix Language. It is not necessarily always the speakers L1. Linguistically, the Matrix Language is the language which supplies more morphemes in a discourse of minimum of two sentences (Myers-Scotton

1993c, 486). However, in the present study the minimum of two sentences that Myers-Scotton provides is not taken into consideration because tweets have character-limit of 280. This means tweets are rather short passages of text, and very often only one-sentence-long. Myers-Scotton (2000) says that Matrix Language should provide *system morphemes* whereas content morphemes can be from either, the Matrix Language or the Embedded Language. The distinction between content and system morphemes is crucial in identifying the Matrix Language. Content morphemes, e.g. nouns, verbs, adjectives and some prepositions, express semantic and pragmatic aspects and assign or receive thematic roles. These are essential to convey messages in communication. System morphemes, e.g. function words and inflections, express the relation between content morphemes and do not assign or receive thematic roles. They are essential in building grammatical frames (Myers-Scotton & Jake 2000, 100). In this study, the Matrix Language and the Embedded Language were determined according to content and system morphemes. Sometimes the tweets were short and had morphemes equally from both languages (Finnish and English), so those would marked as Composite Matrix Language in the analysis.

Part of MLF Model is to divide words into three constituent types: Matrix Language islands, Embedded Language islands, and mixed constituents (Myers-Scotton 2002, 57–58). An island can be a single lexeme but it always has a minimum of two morphemes (ibid.). EL islands consist of morphemes from one language are grammatically well formed (ibid.). Mixed constituents consist of morphemes from more than one language (ibid.). In the present study several approaches will be combined and the focus on the islands was decided to left brief because of the limitation of the Master's thesis.

Matrix Language Frame Model was chosen for the present study because it gives an interesting point of view how determining words and phrases dominant or non-dominant gives us the understanding how code-switching works. For example, one would think that Finnish tweeters write their tweets in Finnish and add English words to emphasize, or to prove a point. However, later on in the analysis it was noted that Finnish people tweet also in English and sometimes it was rather challenging to determine which one of the languages was the Matrix Language. The model also proves that code-switching is not arbitrary but after noting the Matrix Language and the Embedded Language, the words usually act according to the model. However, attempts to identify the Matrix Language have sometimes been unfruitful, which has resulted in the formulation of the "matrix language" as an abstract construct (Myers-Scotton 2002). Still, as researchers have stated, the MLF model is one of the most influential in this field, and has been

successfully applied to several language pairs (Deuchar 2006) and hence it is used in this study as well. The next section explains some of the linguistic terminology used in the analysis.

3.2 Linguistic terminology

In the present thesis a few linguistic terms are used especially in the section 5.4 which is the analysis of the CS between word classes and how the code-switched words are inflected. Some of the terms were explained already in the previous section, but the most central ones in this study will briefly explained in this section.

The term *syntax* means how sentences are structured (Carnie 2013, 4). Syntax focuses on the level of language that is between words and meaning of utterances (ibid.) It is a cover term for many different things happening within a sentence. In the present thesis many syntactic features are examined, both in English language and in Finnish language. The focus will be on the sounds of both languages. The study of acoustics and articulation is called *phonetics* (ibid.). Because Finnish is pronounced as it is written, it is a *phonetic* language whereas English is not. If a sound has a meaningful unit, or it is its own separate word, it is called a *morpheme*. (ibid.) In the present study morphemes, and *suffixes,* to be precise, will be at the center of the focus. Suffix is a morpheme located at the end of the word. For example the word 'teacher' has two morphemes: teach and suffix –er. The next section will introduce and explain my data and chosen methodologies in more detail.

4 Data and Methodology

In this section the data used for the present study will be introduced. Used methodological approaches will be also explained. The sub-section 4.1 will present how the chosen tweets on Twitter were collected and how they form a micro corpus for the present study. Sub-section 4.2 explains how the chosen methods are used in the present study.

4.1 Chosen tweets

Millions of tweets are posted on Twitter every day. In order to find tweets with possible code-switched content key words were used: “Finglish”, “rallienglanti” (=rally English), “anglismi” and “anglisismi” (=aglicism), “sekakieli” (=mixed language), “englanniksi” (in English), “englanti”, “enkku” (English), “iisi” (=easy), and “fiilis” (=feeling). There was no thematic or contextual interest of what the tweets are concerning about so if suitable tweets were found by chance, they were took as a part of the data. The only condition was that the tweet included code-switching and hence merely such tweets were selected. It is acknowledged that the keywords produced data that is language-orientated, and the code-switching is probably intentional in many of the tweets. Because the code-switched content was rather difficult to find without using any keywords, these were decided to use. The keywords are good to keep in mind when thinking about the ‘naturalness’ of the data. After skimming through the tweets on Twitter, it became clear that “Finglish” in a finance world means “financial English”. However, it did not disturb the material gathering as those tweets were rather easy to limit out. In general, Finnish key words were used to find tweets with Finnish and English content. In addition, tweets were searched by using English keywords as well, but those tweets had always merely English content. Tweets that were used for this research were posted between 2010 and 2020. Since the present study focuses on the structural and grammatical aspects, the posting date is not in the center of attention here and thus older tweets were also chosen to be a part of the data. All the tweets are from public accounts which means anyone can read them and comment on them. The tweets were gathered in the fall and winter of 2019–2020 and the total amount of the chosen tweets is 88. It is possible that people delete or edit their posts after posting, so it cannot be said that the tweets can be found later on as they were at the collecting moment. However, screenshots were taken of each tweet. In the present study the date or the tweeter’s name will be not mentioned, since those details are irrelevant and I

wish to keep the tweeters anonymous for research ethical reasons. However, all the tweets have been written in a chronological order, and the list of them is attached to this thesis as Appendix 1. It is left implied that the data could be consulted upon request.

After collecting the tweets, they were divided depending on which kind of code-switching occurred in the tweet. The first division was under three sections according to their types: intra-sentential, inter-sentential, and tag-switching. The next division was according to their levels: word-level, clause-level, or sentence-level code-switching. Some of the tweets fit in more than just one category. Because the tweets were put in each suitable category, it means that after the division more than 88 tweets were in total under the titles. Later on, a numeric data will be presented about each category, and those explain why the total amount is sometimes a bigger number than 88.

4.2 Methods

Like mentioned before, code-switching can be analyzed several ways. This thesis combines various methodological approaches which will be presented next. The first approach was taken from Poplack, and according to her ideas, the tweets were divided into intra-sentential, inter-sentential, and extra-sentential (tag-switching) code-switching. After determining the types, intra-sentential CS was examined in detail. The second approach used on intra-sentential CS was Myers-Scotton's Matrix-Language Frame Model, but since the theory is rather vast, the model was adjusted to be suitable for the present study. In the intra-sentential CS the Matrix Language, the Embedded Language, and Composite Matrix Language (ambiguous combination of the languages) were determined according to Myers-Scotton's approach, and these will be demonstrated by examples in the analysis section. The analyzed aspects were the relation of the Matrix Language, the Embedded Language, and Composite Matrix Language. Montes-Alcalá's theory about levels of code-switching (word-level, clause-level, sentence-level) was utilized the third. The focus was to find and see the level in which code-switching occurs the most. The fourth approach that this study is focused on is Muysken's approach of insertion, alternation, and congruent lexicalization. This is the last part of the analysis and when studying those, grammatical aspect were analyzed as well as inflection of the code-switched words. The thesis is a qualitative research about the nature of code-switching used on Twitter, and the research questions are addressed here again:

1. Which types and levels of code-switching occur in tweets by Finnish Twitter users, and what possibly motivates people using code-switching?

2. What type of intra-sentential code-switching occurs online by Finnish Twitter users, according to the Matrix Language Frame model?

3. What kinds of words are being code-switched in tweets and how does it affect inflection?

The next section provides the collected data with numerous examples. The analysis will be made throughout the section and reflections to the theory are presented as well.

5 Analysis

In this section the data is analyzed section by section and various examples related to the research questions are given. Each sub-section will start with a table where numeric information is given about the findings. Section 5.1 will start by giving examples and examining the reasons for types of CS which are: intra-sentential CS, then inter-sentential CS, and finally extra-sentential CS. The numeric data of Matrix Language, Embedded Language, and Composite Matrix Language will be provided. After this, examples will be presented. The examples are all from the gathered data, and the translations will be in brackets. After the types, the focus is on the levels in section 5.2: word-level, clause-level, and sentence-level. These are again presented with examples and possible reasons will be examined as well. The third sub-section 5.3 presents tweets that had mixed CS content, so this means for example word-level and clause-level within one tweet. The last sub-section 5.4 will be about word classes and inflection, and in that section the closer look at the syntactic elements on tweets will be examined. Possible reasons for Finglish will be reflected through the examples throughout the analysis.

5.1 Different types of CS in the tweets

The total amount of tweets was 88. The analysis started by dividing the tweets into intra-sentential code-switching, inter-sentential code-switching, and extra-sentential code-switching (=tag-switching). Table 2 shows numeric data of the types of code-switching:

Table 2 Types of code-switching in the tweets

intra-sentential	73% (65)
inter-sentential	15% (13)
extra-sentential (tag-switching)	12% (11)

The result is that intra-sentential code-switching occurred in 65 tweets (73 %), inter-sentential code-switching in 13 tweets (15 %), and tag-switching in 11 (12 %). The total number is 89 (higher than the total amount of tweets) because in some tweets several types of code-switching occurred. There was variation between the types, but as assumed, intra-sentential code-switching occurs the most often. Intra-sentential is thought to be the most common code-switching in speech as well, so this finding was expected.

In the tweets where code-switching happened intra-sententially, the Matrix Language (Finnish, English or composite) was determined. All the tweets were also divided according to their Matrix Language and results are shown in the Table 3 as follows:

Table 3 Matrix Language in the tweets

Matrix Language is Finnish	57% (51)
Matrix Language is English	22% (19)
Composite Matrix Language	6% (5)
no Matrix Language	15% (13)

As can be seen from the Table 3, the ML is Finnish in 51 tweets (57%), the ML is English in 19 tweets (22 %) and composite ML is in 5 tweets (6%). As mentioned before, Matrix Language Frame model is studied in intra-sentential code-switching, and the amount of inter-sentential tweets was 13. This is why 13 tweets (15 %) do not have ML at all. It is expected that Finnish is the Matrix Language the most often because the tweets were gathered from Finnish Twitter users. Hence, the majority uses their mother tongue Finnish on their tweets, which makes the number of Finnish as the Matrix Language tremendously high.

Next, types of code-switching will be gone through with examples. The first example of type of code-switching is intra-sentential code-switching where code-switching happens within a sentence:

(1) "Yhyy itken ku oon niin *awkward* puhuu enkkuu mut se oli niin sulonen ja halus musta kuvankin."
(= "I cry because I am so awkward to use English but he was so cute and wanted to take my picture too.")

Code-switching happens inside a sentence and hence the type of code-switching is intra-sentential. The level of code-switching is word-level because one word in a Finnish sentence has been switched to English. The sentence is mostly written in Finnish which makes Finnish the Matrix Language and English the Embedded Language. Code-switched word has not been inflected but inserted in the middle of a Finnish sentence. This means that the sentence follows Finnish grammar rules but the English word follows English grammar structure. Content and system morphemes are from the ML, Finnish. The word "awkward" is commonly used among Finnish (young) people so the reason why the tweeter wanted to use it here may be because they want to use the young style of language. It can also be a sign of bragging

because rather often the word is misspelled (akward). One reason could be also that the tweeter explains about something that happened in an English setting: they had talked English with someone, and this is probably why they chose to use English word in their tweet. It can also be possible that the word is commonly used in the tweeter's life, so that they have forgotten the Finnish equivalent.

The second type is an example of inter-sentential code-switching where code-switching happens between sentences:

(2) "…*wtf that's annoying of ur dad*. Ennen ei ollu enkus läheskää nii kova taso ku nykyää ku suomalaisnuoret käyttää enkkuu ihan hitosti"
("…wtf, that's annoying of your dad. Before, the level was not as high as nowadays because Finnish young people use English freaking a lot")

In the example, the first sentence is in English that follows English grammar, and then the tweeter changes to Finnish in the second sentence. English sentence has English morphemes, and Finnish sentence has Finnish morphemes. Code-switching happens with alternation because two separate structures are juxtaposed. Both sentences follow their own language structure, word order, and grammar rules. Additionally, a tag "wtf" has been put in the beginning of the sentence, but this is not a tag-switch since the English tag is in an English sentence. The reason why the first sentence is code-switched could be a sign of solidarity. Tweeter clearly talks about speaking English already, and to make themselves more convincing, they start by using English. Since they have started the tweet with a tag "wtf", it is possible that the rest of the sentence is "accidentally" written in English as well. However, the rest of the tweet is written in Finnish, possibly because they want to use their own mother tongue to express themselves accurately. Neither Matrix Language nor Embedded Language can be determined because code-switching happens between sentences.

The third example is about tag-switching which means that a short expression, or saying has been inserted in the sentence that has been written in another language:

(3) "Suosikkireaktoni Ylen henkilöjuttuun musta *so far*: …" (="My favorite reaction for YLE's article about me so far: …")

A short tag, or expression in English has been inserted at the end of the Finnish utterance. The reason why tag phrases code-switch, in general, may be that a speaker or a writer temporarily

forgets the equivalent term in the other language. This observation comes from my own life: (especially) English students often speak Finglish. They tend to forget words in Finnish but can say them immediately in English. To make the speech more flowing and fluent, the English expression is often chosen instead. The reason for Finglish in this tweet could be forgetiing the correct term in Finnish, or playing with the words, and hence the English tag has been chosen.

5.2 Different levels of CS in the tweets

After dividing the tweets into different types, the focus turned to the levels of code-switching. According to Montes-Alcalá, three levels of code-switching exist: word-level, clause level, and sentence-level. Table 4 shows the levels of code-switching found in tweets:

Table 4 Levels of code-switching in the tweets

word-level	89% (74)
clause-level	11% (9)
sentence-level	11% (9)

All the tweets were divided according to mentioned levels, and the result is: 89 % (74) word-level, 11 % (9) clause-level, and 11 % (9) sentence-level. Again, as presumed, word-level code-switching occurs the most often. In written and spoken code-switching it is usual that a speaker switches merely one word, and that often happens even unconsciously. This was already mentioned in the theoretical framework section. Of course, tweets are written discourse so it could have been possible that sentence-level occurred even more since writing is always more conscious than speaking.

The first example given is word-level code-switching:

(4) “There is no way I am going to end up having a *hirvikolari* in central Helsinki”… #Finglish. (hirvikolari=traffic accident involving a moose)

In the example, the whole sentence is in English apart from one word, “hirvikolari”. This is an example of intra-sentential CS where English is the Matrix Language and Finnish is the Embedded Language. The code-switched word is common in Finnish but needs a little explanation when said in English. This is probably the reason why the tweeter has used a Finnish word in the middle of an English sentence. This is also, what Myers-Scotton (2002)

calls *cultural borrowing*. If there was an English equivalent, the tweeter would have probably used that one. In the example the Finnish word has been inserted in the middle of an English sentence and the tweeter even added an article "a" in front of the Finnish word, even though articles do not exist in Finnish language. This means that the Finnish word is treated like an English word. The reason why the tweet is in English at all is interesting. Why not just write in Finnish? A probable reason could be to achieve a bigger audience, or maybe the tweeter has more English speaking followers than Finnish speaking. Sometimes code-switching can be done even on purpose to leave the English audience wondering the meaning of the Finnish word. Another example would be the following:

(5) "*Deadline* on Syyskuun vika päivä! Sen jälkeen ei enää pääse mukaan."
("The deadline is on the last day of September. After that one cannot join.")

Sometimes the English word is so familiar that Finnish people use it as such without even realizing that they are using a non-Finnish word. The code-switching happens unintentionally. "Deadline" as such word is very common and often used maybe because in Finnish similar word lacks. In Finnish one can explain "the final date of…" but because in English this handy word exists, Finnish people frequently use it.

The next example is a clause-level CS:

(5) "*It made sense* että mä julkaisen niitä biisejä mitä mulla on. Oon *thank god finally* löytänyt itselleni oikean tiimin."
("It made sense that I publish the songs that I have. I have, thank god, finally found the right team for me.")

A clause is a part of a sentence that contains a subject and a predicate. In the tweet "It made sense" is a clause that is followed by another clause "että mä julkaisen niitä biisejä mitä mulla on", but the language changes between them. The latter clause with the main content is said in Finnish which means Finnish is the Matrix Language and English is the Embedded Language, which adds just little details to complete. Both parts follow their own language structure which means that the sentence has alternation code-switching that happens intra-sententially. The latter sentence "Oon *thank god finally* löytänyt itselleni oikean tiimin" has two tag-switched impressions "thank god" and "finally" in the middle of a Finnish sentence. The tags follow their

own grammar structure within a Finnish structured sentence which means this is intra-sentential alternation. Again, Finnish is the Matrix Language and English is Embedded. When reading this tweet, it feels like a young person has written this. It seems that an effect of popular culture is the reason for Finglish in the tweet. Young people who watch TV or live abroad are under the influence of this kind of language use: some expressions are said in English because many people in English speaking environment use them. When a person hears them all the time, it is natural that the expressions become parts of their own language, mixing the two languages and creating Finglish phrases.

The third level is sentence-level code-switching and an example of that would be the following:

(6) "Hm it doesn't seem like she's eating :D *ok osaan enkkuu*"
("Hm it doesn't seem like she's eating :D ok I know English")

Here, the tweeter has written two separate sentences in different languages. It cannot be determined which one is the Matrix Language and which one is the Embedded Language because code-switching happens inter-sententially. It seems that the tweeter has doubts on their English, and they probably think that the English sentence is written wrong. To say "ok osaan enkkuu" in Finnish is a way of saving themselves from the possible error made in the first sentence. On the other hand, this tweet could be a reply to an originally English tweet and that is why English is chosen in the first sentence. Maybe the tweeter had difficulties in writing the English sentence and, hence, tried to save the situation by adding the Finnish sentence that ironically says that "I know this language". This kind of behavior on social media seems to be rather common amongst Finnish people. Maybe that is related to the shame that one would speak English accidentally wrong. The whole utterance is turned to be a joke in order to avoid the shame. This is related to Finnish people's shyness of speaking English which was explained in the section 2.1.

5.3 Types and levels of CS mixed in the tweets

Not all the tweets have merely either one type or level of CS in them. It is very common that a tweet includes several types and levels:

(7) “Silitin ruskeaa kiharakarvasta labradoria ja ku sen omistaja puhu enkkuu olin sillee *I just wanted to say you have a really nice dog* ja osotin koiraa *nice dog*!!! Molemmat ilahtu!! *Profit* 100%”
(”I was petting a brown curly haired Labrador, and when its owner spoke English, I was like I just wanted to say you have a really nice dog and pointed at the dog nice dog!!! We both became happy!! Profit 100 %”)

The sentence starts in Finnish, and the first code-switching happens in: “I just wanted to say you have a really nice dog”. This is a type of inter-sentential CS, because the first sentence is in Finnish and the second one is in English. Its level of CS would be sentence-level. CS happens by alternation because two sentences with different languages are juxtaposed. Secondly, the latter “nice dog” is a type of intra-sentential and the CS level is word-level. English is the Embedded Language because the rest of the sentence is in Finnish: “osotin koiraa *nice dog”*. It is also an alternation because both “osotin koiraa” and “nice dog” follow their own grammatical structures. Thirdly, “Profit 100%” would be a type of tag-switching and the CS level is again word-level, since there is not a sentence around it. The tweet started in Finnish and Finnish seems to be the dominant language, and hence “Profit 100%” would be an English tag-switch. The reason for code-switching here is clear: the tweeter narrates what had actually happened and how they had spoken in the situation. The code-switching happens when the spoken parts are told as they did in the real setting: in English. The explanation itself happens in Finnish because the tweeter presumably always tweets in Finnish and hence the outcome is a Finglish tweet.

(8) “*lol* mut *real talk*, siit tulee niin syyllinen fiilis et *why cant i be happy and excited about this suggestion to meet up, what am i a total fucking dumbass loser*”
(“lol but real talk, it makes me feel guilty, why cant i be happy and excited about this suggestion to meet up, what am i a total fucking dumbass loser”)

The sentence starts with an English acronym “lol” and then has one Finnish word-level CS insertion “mut”, then the tweeter continues in English “real talk”. In this part English is the Matrix Language and Finnish is the Embedded Language, because the short utterance has a Finnish word “mut” within an English “sentence”. The sentence is not actually complete because it is lacking a verb, but in online writing this happens rather often. The tweet continues with a Finnish sentence-level CS alternation “siit tulee niin syyllinen fiilis et” and

then the rest of the sentence is in English: “why cant i be happy and excited…” . If we look at the first sentence as a whole: “lol mut real talk, siit tulee niin syyllinen fiilis et…” that changes things: Finnish is the Matrix Language and English is the Embedded Language. Acronym “lol” and a phrase “real talk” would be English tag-switching in a Finnish sentence. The rest of the tweet is completely in English, which means the switching happens on sentence-level. Two sentences in different languages are juxtaposed. The reason why “lol” and “real talk” have been written in English here could be the impact of popular culture again. However, why the tweeter changes the language completely into English can have many reasons. The tweeter could be a bilingual whose English is more dominant and the description of their feelings naturally happens by using that language. Usually the tweets are addressed to someone, so here the tweeter might want to achieve their English or bilingual followers.

Sometimes determining the code-switched syntactic parts is rather complicating:

(9) “On the road! Seo menoo ny. Helsinki here we come, vapiskaa.”
(”On the road! Here we go now. Helsinki here we come, shiver.”)

The sentence starts with an English saying “on the road”. Then inter-sentential code-switching happens, since the next sentence is completely in Finnish “seo menoo ny”. The first two sentences are juxtaposed which means CS happens in sentence-level. The last sentence is in Finglish which makes this intra-sentential CS but with Composite Matrix Language, because neither is dominant over another. On the other hand, “Helsinki here we come, vapiskaa” could be also seen as a Finnish sentence with “here we come” English tag-switch in the middle of the Finnish sentence. Examples like this are sometimes rather difficult to analyze because there are more than one clear option. It depends how we divide the syntactic parts and what is seen equal. Especially with tags, the interpretation of CS can be ambiguous.

5.4 Code-switching between word classes and inflection

Different word classes act differently when it comes to inflecting the code-switched words. It is common that when a person switches an *adjective* from one language to another, it does not take the Matrix Language’s grammar:

(10) “Joo, onhan se negatiivinen *asset*, mut kiihdyttäminen on *satisfying***.** Se on hauskaa!”
(“Yeah, it is a negative asset, but the speeding up is satisfying. It is fun!”)

In the first sentence the Matrix Language is Finnish and English is the Embedded Language. CS happens in a word-level and the inserted word "satisfying" is inflected like it would be in an English sentence. Adjectives are code-switched rather often and the reason could be that sometimes an English word might express the speakers feeling better than a Finnish word, resulting a Finglish outcome. Our feelings affect our speech so it is common that the person uses the exact word which describes their feeling the best at that specific moment. If the word in English describes it better, it is normal that a speaker chooses that language and not their mother tongue. In the example, the other English word in the sentence "asset" is a noun, but nouns do not tend to have a certain pattern in their inflections. Various examples were found with different use of nouns:

(11) "If I can just get some *makaronilaatikko*, I will be just okay"
("If I can just get some macaroni casserole, I will be just okay")

In this example we have a word-level CS and the inserted word "makaronilaatikko" follows English grammar because that is the Matrix Language. In English, the noun after "get" should be in its basic form. Finnish grammar, however, would take a partitive case makaronilaatikko**a**, as Finnish grammar a suffix –a/-ä is added at the end of nouns that are objects in a sentence. But like Myers-Scotton said, code-switched words are tend to inflect according to the Matrix Language. It would be an odd way to say: "If I can just get some makaronilaatikko**a**". In another words: insertion rarely happens if only one Finnish noun is added as an object in an English sentence. The reason for Finglish here is cultural: "makaronilaatikko" is a traditional Finnish food that do not have equivalent word in English. It could be "macaroni casserole" but that is similar to Americans' "mac and cheese". However, the product is not exactly the same in Finland and hence the tweeter has used the Finnish word. Again, the tweeter might have followers who understand both languages and this is why the Finnish word is left unexplained further. Different example of the inflection of a noun would be the following:

(12) "the label above some flat's mail slot reading "no mainokse***s***""
(mainokset=advertisements).

In this example code-switching happens in word-level with a rather odd inflection. The Finnish word "mainokset" has taken English suffix –s to mark its plurality. The writer of the label

probably does not know the correct word in English but still wants to write the post in English, and for some reason they have used English grammar in a Finnish word. They are using the English suffix instead of Finnish suffix, that would be –t, marking the plurality (=mainokse**t**). The writer has taken the Finnish word "mainokset" and replaced the Finnish suffix –t with and English suffix –s. The writer is on the right path, meaning, they have chosen the correct suffix to impress plurality. However, in this case Finnish and English are too different languages, and the replacement makes the word peculiar. Insertion is attempted but the result is odd. The reason for doing this could be a sign of low proficiency in L2 but also a sign of a sense of humor: the person does not know English well and shows it by writing the word completely wrong on purpose. As, mentioned before, this often happens amongst Finnish people who are insecure of their English skills. In order to try their best, they say the word/sentence completely wrong and acknowledge that their English is bad.

Sometimes the English noun can be inflected like a Finnish word, and still seem "normal" to a Finnish speaker.

(13) "…puhuin mun uuden *roomie*n kaa vissii 15 min enkkuu ennen ku esiteltii toisemme ja tajuttii et ollaa suomalaisia".
("…I talked with my new roomie for about 15 minutes in English before we introduced ourselves and realized we are both Finnish").

The tweet has a word-level CS, and the word "roomie" has taken Finnish genitive form that suffix –n symbolizes. Finnish is the Matrix Language and the Embedded English word is inflected like Finnish, which makes this insertion. The setting has probably affected the use of Finglish here because presumably this incident happened abroad, in an English environment. "Roomie" is an easy word to inflect as a Finnish genitive because only a letter *n* can be added at the end of the word. This is probably why the tweeter has chosen to use the original English word, and not translated it to Finnish. This kind of code-switching causes sometimes double plurality. If an English word is common for the user, they might use the word part of a Finnish sentence and also inflect it in the plural form in Finnish, even though the word is plural already in English:

(14) "Poika on meillä kattonut tota. Mulla vasta Breaking Bad vaihe. *Thingssit* tulossa."
("My son has watched it. I am still in the Breaking Bad phase. Things are coming.")

The tweeter uses an English word “things” on their tweet, but adds Finnish plural suffix –t at the end of the word to mark the plurality repetitively. CS here happens word-level and is intra-sentential, and the English word has been inserted in a Finnish sentence. The Matrix Language Finnish is dominant and hence the English plural word takes Finnish plural form as well. This kind of code-switching tends to be more unconscious because the tweeter probably does not think about the word “things” as an English plural, but just uses it as its own unit. Another similar example would the following:

(15) “Tuo oli niin tuore asia ettei ollut ehtinyt *newsseihin*, mutta nyt on.”
(“That thing was so new that it hadn’t even reached the news yet but now it has.”)

The English word “news” is treated as singular but technically it is an uncountable mass noun. Because in Finnish “news” has its own words for singular and plural, the plural inflection has been operated on the English word “news”, with a suffix -eihin. This is probably because the tweeter has wanted to put the word “news” on its genitive form, and the Finnish suffix –n to mark the genitive would be impossible to add at the end of the word “news” (=newsn*).

Sometimes the inflection is done the other way around, meaning, that a Finnish word is inflected like an English word. The following example is understandable to anyone who knows both Finnish and English. Unfortunately it would be impossible to understand to anyone who does not know Finnish, even though the sentence seems to be written completely in English:

(16) “Oh how bittersweet thou are, the vitutation of the spring Saturday imuration.”

The tweeter probably has a sense of humor and this tweet is meant to be a joke. Finnish expression “vitutus” means “pissed off”, that something annoys someone but with a strong negative tone. “Imurointi” however means vacuuming. These two Finnish expression have been turned into English nouns by adding suffix –ation at the end of a Finnish word. To be precise, these Finnish words in the tweet mean nothing in either language, but they have been modified or adjusted to be more suitable for the English inflection. Still, a bilingual person would understand that the tweeter is pissed off because they need to vacuum on a Saturday. English is the Matrix language, and even though the Finnish words are not correct Finnish, this would be an example of word-level intra-sentential code-switching.

When thinking about the previous examples, it becomes clear that the inflection varies and the most frequently follows the Matrix Language's grammatical structure. When Finnish is the Matrix Language, system morphemes like prepositions are left out and Finnish suffixes are put at the end of the English words. This is notable because Finnish does not have articles or prepositions but suffixes determine the words' case. This proves what was suggested before: code-switching is not random, and for bilinguals the code-switched words are meaningful, even though they take parts of each other's grammar (even if Finnish and English's grammars are totally different). Sometimes CS happens because the other language is better for describing what is meant and that is why a word from the other language is used.

When code-switched word is a verb, the CS occurs differently from nouns and adjectives:

(17) ”Niin, sitä tarkoitinkin. :D Minä *downgreidaan* toinen *upgreidaa*. Sori #anglismi”
(”Yes, that is what I meant. :D I downgrade and the other upgrades. Sorry #anglismi”.)

In this sentence a word-level CS occurs and the tweeter has code-switched verbs “downgrade” and “upgrade”. The Matrix Language is Finnish and the Embedded Language is English, and the verbs are inflected like Finnish words which makes it insertion. The code-switched words might seem weird for non-Finnish speakers and there is a reason for that. Finnish is a *phonetic language* which means, it is pronounced as it is written. This is why the tweeter has partly changed the written form of the English words and changed them to their phonetic form. The first verb “I downgrade” (in English I downgrade) is spelled “downgreidaan”, in International Phonetic Alphabet (IPA): [‘daʊngreɪda:n]. According to Finnish grammar, suffix –n is used to express that a singular 1st person is doing something. The same happens with the word “upgrade”: “upgreidaa”, IPA: [‘ʌpgreɪda:]. Again, the Finnish suffix –aa means that a singular 3rd person is doing something. The words may look peculiar to English speaking person, but anyone who knows a phonetic language might not even notice the change in the spelling. When going through the tweets, it was noticed that this kind of change in spelling is common when a Finnish person is writing in English. The spelling changes to its phonetic form how a Finn would say it according to Finnish pronunciation rules. Among the tweets several examples of such phenomenon were found and they will be presented next to see how

different ways tweeters use this kind of code-switching, and hence how diverse Finglish talk can be:

(18) "Puhun suomee ja enkkuu myös sekasin kun haluun *inkluudaa* ihmisii jotka ei puhu suomee mut välillä en jaksa ajatell mite tää ilmaistaan enkuks"
("I mix Finnish and English when I want to include people who do not speak Finnish, but sometimes I am not bothered to think how this is expressed in English")

(19) "Missä mun luistimet? Mä *niidaan* niitä koulussa." … #anglismi
("Where are my skates? I need them at school." … #anglismi)

(20) "…Cheek sanoi "yritin *kapturoida* sen tunnelman." … #suomi #anglismi"
("…Cheek said "I tried to capture the feeling" … #suomi #anglismi")

(21) "*Svitsillä* ekaan *hittiin* ja *gräbi* ennen *ländäystä*. Kova *rani*! #halfpipe #sotshi #anglismi"
("With switch to the first hit and a grab before landing. What a run!") (In this example we have an example of sports jargon that was mentioned earlier.)

(22) "Ignooraaminen voi olla iisiä mut miksi avoidaa."
("Ignoring can be easy but why avoid?")

Verbs "include, need, capture, switch, grab, land, ignore, and avoid" are originally English words, but their written forms have been changed to fit Finnish language better. Because Finnish is a phonetic language and Finnish is pronounced as written, people who do not know English and are native in Finnish, would think that a word "need" is written "niid" [ni:d]. In the previous examples, the writer probably knows the word's correct written form, but to keep the sentence coherent, they change the written form to be consistent with Finnish words. After the written form has been changed to be suitable with Finnish language, the inflection is easier to do according to Finnish grammatical structure. Insertion in Finnish is usually easier to form when the English word has been changed to its phonetic form. Finnish people who have moved to an English-speaking country tend to do this to words rather often, and not only with verbs. The reason for Finglish in the tweets could be including certain people and excluding others. Another reason could be that people are speaking about a topic they are more familiar with in English. One reason could also be that they temporarily forget the word in Finnish and writing it in English is faster. The idea in tweets is that one quickly responses or one wants to share something without even deeply pondering it. Sometimes it is clear that the Finnish equivalent is forgotten, so the English word replaces the forgotten word:

(23) “Missä on *paatiruuma*?” (”Where is the bathroom?”)

(24) “Pane astiat *sinkiin*.” (“Put the dishes into the sink.”)

Some tweets had the English word written as an English word but inflected like a Finnish word which makes the word looking peculiar again, but for bilinguals the word is somewhat understandable:

(25) ”fiilis kun opettaja callouttaa siitä että on kolmatta kertaa samalla kurssilla” (”feeling when the teacher calls out that one is taking the same course for the third time”)

The tweeter has used English phrasal verb “to call out” in the Finnish sentence. This kind of word-level intra-sentential CS is common, but how the tweeter has inflected the English word makes it look a bit odd. “Call out” has been turned into one word callout, and suffix –aa in Finnish means that a third person singular is doing something. This is the result of insertion again, because the grammar of one language is fitted into another language’s word. The English phrasal verb is treated like a normal Finnish word but for some reason the English written form has been maintained. This kind of Finglish, where merely a verb is changed in English is rather common on Twitter, and this can be noted when looking at the previous examples 18–22, and 25.

In the tweets it often occurred that people would like to say something in one language, but certain word does not have an equivalent in the other language. One example of the “makaronilaatikko= macaroni casserole” was mentioned before. This is where word-level code-switching often happens and it can be even said to be cultural borrowing as mentioned before (Myers-Scotton 2002). Such words are often related to holidays or events that do not happen in English speaking countries:

(26) “First *munkkis* and *sima* of this *vappu* enjoyed with the team at the office. Happy *Vappu*!”

“Vappu” is celebrated on the first of May, and this celebration has its own word in Finnish. “Munkki” is a sweet pastry eaten on the 1st of May (=Vappu). Here, the tweeter has inflected “munkkis” in the similar way than “mainokses” before by inserting English plural suffix –s at

the end of a Finnish word. Again, English is the Matrix Language and the Embedded Language's words follow English grammatical structure. "Sima" is a traditional drink that Finnish tend to drink on the 1st of May. Because tweets are short text passages, it is easier to use the Finnish words because they do not have English equivalents. The writer probably assumes that those who it involves, will understand the Finnish words in the middle of an English sentence and hence they do not open the words any further. Several similar examples were found:

(27) "Can't wait for ya'll to see my costume for *penkkarit*! *Penkkarit*? Umm. Party for *lukuloma*. I mean, studying holiday(?)…"

(28) "Haha, "We fight back with *sisu"*

Here again, it is presumable that a Finn understands what is meant on the examples. "Penkkarit" (originally penkinpainajaiset) is a day when senior high schoolers celebrate their last day before their study break (=lukuloma) begins. The writer has tried to explain this but neither of the words translate easily into English. *Sisu,* on the other hand, is a Finnish concept that means determination, courage, bravery, and a mindset which enables people to reach beyond their limitations of an action. The word lacks an equivalent in other languages but in Finland the word is commonly known. The reason why the tweeters have used code-switching is simply cultural. Cultural aspects affect how we use language and it also explains the use of code-switching in certain situations.

Composite Matrix Language was explained before, and its characteristics are that the levels and structures could be split and recombined. Lexical and grammatical structure can be abstract and neither of the languages is dominant (Myers-Scotton 2001, 53). This kind of example was found as well among the tweets:

(29) "Making of Salibandyliigan puolivälierä, featuring kapteenit erotuomarikopissa! #sekakieli"
("Making of the quarterfinal of the floorball league, featuring captains in the referee booth")

In this sentence we can see prepositions, Finnish suffixes, and tag-switching. The sentence starts with one language and then continues with the other. This kind of sentence would be Composite Matrix Language because both grammars and vocabularies occur but neither is dominant over another. Composite Matrix Language proves that code-switching has its own

rules and words cannot be randomly switched. Two grammars can collide and we still understand completely what the sentence is about, IF we are bilingual and fluent in English and Finnish. Finglish is in a way its own understandable language for Finnish-English bilinguals.

Many idiomatic phrases show congruent lexicalization in the text. Even though English and Finnish are rather different and congruent lexicalization rarely occurs, an example was found among the tweets. The next example has CS with congruent lexicalization:

(30) "There is some wifi in my *bussi*"
("There is some wifi on my bus")

Congruent lexicalization happens here because the tweeter is using a Finnish loan word "bussi" that is understandable in both languages. Code-switched Finnish word suits for the English sentence and is totally understandable even though an English word is replaced with a Finnish word. Even though Finnish and English are rather different, congruent lexicalization is possible in some cases.

Code-switching has its own grammar rules and this may be the reason why *pronouns* are rarely switched. The following example is not from the data, but it shows how code-switching can be done "wrong" (nobody switches language like that):

(31) "*Minä* said that you lied."* ("I said that you lied.") or "*I* kaaduin ulkona."*("I fell outside.")

These examples support the fact that code-switching pronouns is against CS rules. It makes the sentence peculiar, too complicated and difficult to understand.

In the following, Table 5 shows how code-switching varied between different word classes.

Table 5 Variation of CS between word classes

Nouns	63% (60)
Verbs	23% (22)
Adjectives	13% (12)

The result of variation of CS between different word classes is nouns: 63 % (60), verbs: 23 % (22), and adjectives: 13 % (12). Other word-classes had CS as well but the three were the largest groups where CS happened. Nouns are the most common word class to appear in

sentences so it is expected that they are the most code-switched as well. It was a little surprising that verbs are the second largest group and are code-switched more often than adjectives. The reason may be that verbs occur more often and nowadays many young people code-switch even common verbs in their speech (previously mentioned “need” and “include” etc.).

The three subsections showed that different types and levels are found on twitter, and, as assumed, the most frequently CS happens intra-sententially and word-level. The Matrix Language in intra-sentential CS is Finnish in the majority of the tweets, and hence it can be presumed that people tend to write in their own mother tongue and add code-switched words in another language. According to the data, it was also seen that code-switched words will the most likely inflect according to the Matrix Language and sometimes Finnish people tend to write code-switched English words phonetically (as they are pronounced) because Finnish is written like that. The data showed that CS happens in many word classes but the most often nouns were code-switched, possibly because nouns occur the most often in speech and writing. It can be said that Finglish is written on tweets consistently even though examples of different kinds of CS were found in the data.

6 Results and Discussion

This section interprets the findings of the study by reflecting them to the given theories. The aim of the present study was to examine the nature of code-switching on social medium Twitter. The tweets were all from native Finnish speakers who mixed their code between English and Finnish. Tweets that had code-switched content were collected in the fall of 2019 and the winter of 2020, and the total amount of the tweets was 88. The tweets were written either in English or Finnish and the topic was not the priority and hence all found tweets with code-switched content were collected. In order to find tweets with code-switched content various, mainly Finnish key words were used, such as: Finglish, sekakieli, anglismi, enkku, filmi, iisi, and rallienglanti. Finnish key words offered suitable tweets that had mixed Finnish and English. Key words were necessary obtain is order to find code-switched content but at the same time using them rather language-orientated tweets were found. This needs to be taken into account when considering the results. Even though the content of the tweets was not at the center of the attention in this study, it is important to acknowledge that the tweeters may have been more conscious of language use when writing their tweets.

The first research question was interested in different levels and types of CS, and possible reasons for CS were examined. Levels for this thesis were taken from Montes-Alcalá's theory and they mean whether CS happened word-level, clause-level, or sentence-level. The results were that CS happens in all levels and from the data the division was the following: word-level 74, clause-level 9, and sentence-level 9. Code-switching seems to happen the most often in word-level, and this was presumed before because the same kind of results have been found when spoken code-switching has been studied. As it was presented in the Theoretical framework section, code-switching is usually spontaneous and might happen unconsciously. Even though written code-switching is more a conscious decision, it still occurs similarly to spoken code-switching, on word-level. Twitter as a social media platform is informal, and people tend to write there rapidly and without planning the content. This may be one of the reasons why code-switching happens on word-level: a word that first comes to mind is used and all the people who are involved will (presumably) understand it. At the same time it excludes people: if you did not understand the tweet, it was not meant for you. This kind of CS occurs often in tweets which are about hobbies, TV-series et cetera. Code-switching can be used as a tool: hobby-related words can be code-switched in order to address the tweet

merely those who belong to the same hobby community. Such example in the data was, for example, where the tweeter explained about painting goblins and perytons. Sometimes code-switching happened on word-level because the equivalent word was missing in the other language. These kinds of cultural borrowings happened for example related to holiday-words or foods and drinks. It is common that different cultures have different words for celebrations, and if that is not celebrated in an English-speaking country, English obviously lacks a word for that celebration. This is where a speaker often uses code-switching: it is easier to use another language's word than describe the food or drink thoroughly. This kind of behavior is common in Twitter where the amount of characters in the platform is limited. In general this kind of behavior is more common in written CS than in spoken CS, since describing and explaining in written form is more complicating and takes more time than in speech. The last findings on levels were idiomatic expressions that were results of a cultural-based code-switching. These kinds of expressions are common in one language that are taken as a whole into another language, for example, "oh my god" and thank god". It was interesting to see how clause-level CS was usually located either in the beginning or at the end of the tweet. This could be the tweeter's way to emphasize the rest of the tweet.

The other part of the first research question was to examine the types of CS. Different types of CS mean whether CS happens intra-sentential (within a sentence), inter-sentential (between sentences) or extra-sentential (tags). The results were the following: intra-sentential CS occurred 65 times (73 %), inter-sentential CS occurred 13 times (15 %), and tag-switching occurred 11 times (12 %). In the previous section it was mentioned that word-level CS occurred the most frequently. As a continuum for that, intra-sentential CS occurred the most often as well. This finding supports the reliability of the present study: CS was found the most within a sentence which means that only independent words were code-switched the most as well. Inter-sentential and tag-switching occurred almost the same amount and this can be thought rather surprising. Nowadays people tend to use tags, such as acronyms (such as lol and diy), rather often, but still, in the data merely 15 % included such code-switching. It was decided not to search the tweets using an acronym as a keyword for two reasons: firstly, such tweets are usually written merely in English and lack code-switched content. The other reason was to avoid manipulation of the data. If tags were used as keywords to find such tweets, the overall findings would have been tremendously different and would give biased information about the nature of language and code-switching on

Twitter. According to the present study it can be said that, rather surprisingly, acronyms do not increase code-switching in written language in Twitter. The reason could be the average age of Twitter users: older people tend to use Twitter but not acronyms. The result would be rather different if examining the same thing in another social media platform, such as Instagram or Snapchat, which are more popular among young people.

After examining levels and types of CS, intra-sentential CS was then examined more thoroughly. The second research question was interested in the Matrix Language Frame model in intra-sentential tweets. The model was taken from Myers-Scotton Matrix and Muysken's ideas of code-switching were combined to that. The model was not used directly as Myers-Scotton uses it, but it was altered to be suitable for the present study. According to the model, Matrix Language and Embedded Language were determined on found tweets including intra-sentential CS and also proved by explanations why either language was Matrix or Embedded. The found ML and EL were presented with examples and explained along with the reasons why such phenomenon was happening in a certain tweet. The numeric finding was that Finnish was the ML in 51 of the tweets, English was ML in 19 of the tweets and no ML in 13 of the tweets. If ML was not determined, it means that CS did not happen intra-sententially in the tweets. According to the Matrix Language Frame model if a sentence has ambiguous combination of the two languages, they are called composite matrix languages. This means that neither of the languages is dominant over the other nor are called Matrix or Embedded. The amount of composite matrix languages among the tweets was 5. Composite matrix languages were sometimes rather difficult to determine, because it was not always clear whether one of the languages is dominant or not. In the Theoretical framework section it was already presented that people tend to speak their mother tongue and include code-switched content with another, not dominant, language. The present study supports this fact. It was assumed that Finnish would be the Matrix Language more than English because the tweets were from Finnish tweeters' accounts.

When examining the tweets, the focus was also on grammatical inflection of the code-switched words, and this was presented with the third research question. In this section, the analysis took parts from Myers-Scotton's theory and additionally Muysken's ideas. Like mentioned before, this study showed that the code-switched words often act according to the Matrix Language's grammar rules, and consistency was found to support this claim. The switched words were the most frequently inflected after the ML and this may be due to the

fluency of language. Examples were found in both ways: when Finnish was the ML, the code-switched English words were inflected like Finnish words, when English was the ML, Finnish code-switched words were inflected like English words. In addition, when Finnish was the ML, the English words were written without the system morphemes, such as articles and prepositions in the sentence. Whereas English was the ML, Finnish words would not have typical Finnish suffixes at the end of the words to mark the tense, plurality, or possession. In written code-switching, EL words mainly functioned according to the ML's grammar rules. Muysken's ideas were examined whether CS was a result of insertion, alternation or congruent lexicalization. Both insertion and alternation focus on structural constraints on mixing. Congruent lexicalization means that the two languages share the grammatical structure (Muysken 1997). Examples of insertion and alternation were found and presented with examples. Congruent lexicalization was rather difficult to find since Finnish and English are alike. This study was not meant to be a syntactic study, and hence another approach besides Muysken's theory would have given more detailed information and more points of view for this study. However, inflection in code-switching on social media is a rather new field to study and this thesis offers something new to the field and further research would be interesting to see in the future.

At the end of the analysis code-switched words were separated according to their word classes to see the variation between them. The results were the following: nouns were code-switched 60 times, verbs 22 times, and adjectives 12 times. Nouns appeared the most in the sentences so it was expected that they were code-switched the most often as well. They were also inflected according to the dominant language: if a Finnish word was put in an English sentence, an article was added in front of the noun even though articles are not in use in Finnish. Verbs were code-switch rather easily, especially if explaining a hobby or sport. English code-switched verbs were often inflected like Finnish words, by adding suffix at the end to mark who is doing something. Adjectives code-switched the most frequently on word-level, a person would write a sentence in Finnish and merely add English adjective in the sentence without inflecting it. It was presumed that pronouns would never be code-switched, and such code-switching was not found either. Word classes were examined in order to see whether a pattern can be found in their use. This was found in the data, since CS happened according to the CS "rules". Certain word classes, such as particles and pronouns were not code-switched because it would make the language too odd. Even though code-switching breaks and

combines grammar structures, it still follows certain patterns to make the language understandable. It is important to remember the main function of language: to understand and to become understood.

This thesis aimed to show that written code-switching is not arbitrary or random, but CS has rules according to which it is formed. Writing on social media is often seen somewhere between spoken and written, but in the present study tweets were analyzed as written texts and hence code-switching was seen as written as well. Written CS is more conscious than spoken, because the writer has more time to think how to write the words. Also, speech is a more rapid way to express oneself, so describing and explaining happens more in speech than in writing. Still, written CS occurs similarly to spoken CS and identical aspects were found on the written discourse. Social media as a platform is thought to be informal, so people do not think thoroughly what they write there. This is why it offered an interesting source of written passages that were examined in the present study. The amount of tweets was, like already mentioned, 88 and with a bigger amount of sampling the results would have been more reliable. Also, tweets could have been gathered among English speakers who use Finnish code-switching, but such tweets would have been rather difficult to find. The Matrix Language Frame model could have been used differently, including the part of islands, explained in the Theoretical framework section. However, the present thesis is enough to show how Finnish speakers use their language on social media, and how switching between English and Finnish functions. All the types according to Montes-Alcalá were found, and also MLF model was a successful model to use with the tweets.

According to the results, it can be said that code-switching is a frequent phenomenon and it has consistency on written discourse. Finnish people tend to use code-switching in similar ways. Even though people write on social media without always planning the content, code-switching is used rather often and people use it in order to express themselves better. Finnish people might be shy to use spoken English but written English might be easier to use, because the accent cannot be heard when English is written. The truth is that, in theory, Finnish people should be rather good in English, because the education starts early and Finnish education is thought to be one of the best in the world. In the latest PISA (The Program for International Student Assessment) study, Finland was ranked in the 7th place in the world (oecd.org). However, as the data showed, young people are braver and more confident to use English and I think that the confidence among Finnish people to use English

will increase when the confident generation is gradually growing. Even though some might say that English spreading to Finnish people language use is violating Finnish language, I do not see the spreading only as a negative thing. Languages help us to understand other cultures, and unfortunately exploring them with only Finnish is rather hard. Code-switching as a phenomenon is sometimes seen irritating, and as shown in the data, some people get bullied for using that ("you are soo international…"). In my opinion Finnish people should use English with pride both in speech and in writing, and the language use on Twitter shows that gradually Finnish people are starting to do so. The next section will conclude my findings of the present study and offer further research ideas for the future.

7 Conclusion

The aim of the present study was to study the nature of code-switching among Finnish twitter users. CS was analyzed according to their types (intra-sentential, inter-sentential, and extra-sentential) and levels (word, clause, and sentence). Intra-sentential code-switching was examined separately using Myers-Scotton's Matrix Language Frame model. The focus was also on the word classes and inflection: the aim was to see how code-switched words behave grammatically in the sentence. The analysis was based on linguists Myers-Scotton's, Muysken's, Poplack's, and Montes-Alcalá's theories. The hypotheses was that written code-switching would act similarly to spoken code-switching, and those patterns were examined according to the data, that was tweets from the online platform Twitter. All the tweets were from Finnish-English bilinguals, and presumably all of them speak Finnish as their mother tongue.

Many kinds of Finglish tweets were discovered. CS occurred in word-level, clause-level, and sentence-level. It also occurred within sentences, between sentenced, and by tags added to the tweets. Within sentences the Matrix Language (the dominant language) and the Embedded Language were determined. It was noticed that content and system morphemes, and grammar rules followed mainly the Matrix Language's grammar rules and inflection habits. In some cases, the sentence had CS equally from both languages where neither ML nor EL could had been determined. These cases are called to be Composite Matrix Languages, because two languages are written but neither is dominant over another. Inflection and CS in word classes were also examined. Even though some weird inflections were found among the tweets, the majority followed the ML's grammar rules. When Finnish was the Matrix Language, system morphemes like prepositions were left out and Finnish suffixes were put at the end of the English words. CS in different word classes was put in a table in the analysis (Table 5), and the result was that CS occurred the most often with nouns (63 %), the second frequently with verbs (23 %), and the third frequently with adjectives (12 %). It was also notable that pronouns would never be code-switched.

As a conclusion can be hence said that written code-switching on social media is similar to spoken code-switching. Plenty of further research ideas exist. Written and spoken CS have similarities and it would be interesting to compare them more. Further research idea would be to examine people's natural speech and compare it to their written text. In addition,

people who write, for example, on Twitter could be interviewed how they think they use CS and then record how they actually use it in speech. Syntactic research could be done, like mentioned already in the section 6 Results and Discussion. Another idea for the future would be exploring social media platforms and use other approaches, i.e. Conversational Analysis, to understand social interaction online better. Also, social media as a phenomenon has spread rapidly into our lives and a lot of language studying on that field can be done. Chosen methods for this study could be used when studying CS in blog writings, Instagram captions or even song lyrics. Especially in Finland, people tend to use English more and it is a fact that we cannot escape its influence towards language use in Finland in the future.

References

Primary sources

Finglish-Twitter Micro Corpus

Secondary sources

Baron, N. 2010. Language in an Online and Mobile World. Oxford University Press.

Blom, J. & Gumperz, J. 1972. Social Meaning in Linguistic Structures: Code-switching in Norway. In *Directions in Sociolinguistics,* edited by J. Gumperz & J. Hyme. New York: Holt, Rinehart and Winston.

Callahan, L. 2002. The Matrix Language Frame Model and Spanish/English Codeswitching in Fiction. Language & Communication, Vol 22, No. 1, pp. 1–16. Berkeley: University of California.

Cambridge English Online Dictionary. 2020. Accessed 9 January 2020. https://dictionary.cambridge.org/dictionary/english/hashtag.

Carnie, A. 2013. Syntax: A Generative Introduction. Oxford: Blackwell Publishing.

Crystal, D. 1987. The Cambridge Encyclopedia of Language. Cambridge: Cambridge University Press.

Deuchar, M. 2006. Welsh-English Code-switching and the Matrix Language Frame Model. Lingua, November 2006. Vol 116, No. 11, pp. 1986–2011.

Encyclopedia Britannica. 2020. Accessed 4 July 2019. https://www.britannica.com/technology/social-network.

Encyclopedia Britannica. 2020. Accessed 4 July 2019. https://www.britannica.com/topic/Twitter.

Fichman, P. & Sanfilipo, M. 2016. Online Trolling and Its Perpetrators. Maryland: Rowman & Littlefield.

Foertsch, J. 1995. The Impact of Electronic Networks on Scholarly Communication: Avenues to Research. Discourse Processes: A Multidisciplinary Journal, Vol 19, No.2, pp. 301–28.

Gardner-Chloros, P. 2009. Sociolinguistic Factors in Code-switching. In *The Cambridge Handbook of Linguistic Code-switching*, edited by B. Bullock and A. Toribio, pp. 97–113. Cambridge: Cambridge University Press.

Gumperz, J. 1977. The Sociolinguistic Significance of Conversational Code-Switching. December 1977, Vol. 8, No. 2, pp. 1–34. Berkely: University of California.

———. 1982. Discourse Strategies. Cambridge: Cambridge University Press.

Guzman, G.; Serigos, J.; Bullock, B., and Toribio, A. 2016. Simple Tools for Exploring Variation in Code-Switching for Linguistics. In *Proceedings of the Second Workshop on Computational Approaches to Code Switching,* pp. 12–20. Austin, Texas: The Association for Computational Linguistics.

Gärdenfors, P. 1999. Some Tenets of Cognitive Semantics. In *Cognitive Semantics, Meaning and Cognition,* edited by J. Allwood and P. Gärdenfors, pp. 19–36. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Halmari, H. 1997. Government and Codeswitching: Explaining American Finnish. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Herring, S. 2001. Computer-Mediated Discourse. In *The handbook of discourse analysis,* edited by D. Schriffin, D. Tannen and H. Hamilton, pp. 612–634. Cornwall: Blackwell.
Juutinen, M. 2019. Koltan- ja merisaamen koodien sekoitus. [Code-mixing between Kolt Saami Neiden Dialect and Sea Saami]. In *Speech and Language,* Vol. 39, No. 2, pp. 165–185. Helsinki: Punamusta.
Lauttamus, T. 1991. Borrowing, Code-switching, and Shift in Language Contact: Evidence from Finnish-English bilingualism. In *Language Contacts East and West,* edited by M. Ojanen & M. Palander, pp. 32–53. University of Joensuu.
Lehtinen, M. 1966. An Analysis of a Finnish-English Bilingual Corpus. Indiana University.
Leppänen, S. 2007. Youth Language in Media Contexts: Insights into the Functions of English in Finland. In *World Englishes,* Vol. 26, No. 2, pp. 149–169. Wiley.
Lipski, J. 2016. The Role of Unintentional/involuntary Code-switching: Did I Really Say That? In Spanish-English Codeswitching in the Caribbean and the US, edited by R. Guzzardo Tamargo, C. Mazak and M. Carmen Parafita Couto, pp. 139–170. Amsterdam/Philadelphia: John Benjamins Publishing Company.
———. 2009. "Fluent Dysfluency" as Congruent Lexicalization: A Special Case of Radical Code-Mixing. In *Journal of Language Contact.* January 2009, Vol. 2, No. 2, pp. 1–39. Online Publication. Accessed 18 February 2020. http://test.scripts.psu.edu/users/j/m/jml34/congruent.pdf.
Martin, M. 1988. The Role of Code-switching Among Canadian Finnish-English Bilinguals. In *Finns in North America: Proceeding of Finn Forum III,* edited by M. Karni, O. Koivukangas, and E. Laine. Turku: Institute of Migration.
Montes-Alcalá, C. 2005. Two Languages, One Pen: Socio-Pragmatic Functions in Written Spanish-English Code-Switching. Santa Barbara: University of California.
———. 2007. Blogging in Two Languages: Code-Switching in Bilingual Blogs. In *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics*, edited by J. Holmquist, A. Lorenzino, and L. Sayahi, pp. 162–170. Somerville, MA: Cascadilla Proceedings Project.
Muysken, P. 1995. Code-switching and grammatical theory. In *One Speaker, Two Languages: Cross-Disciplinary Perspectives on Code-Switching,* edited by L. Milroy & P. Muysken, pp. 177–198. Cambridge: Cambridge University Press.
———. 1997. Code-Switching Processes: Alternation, Insertion, Congruent Lexicalization. In *Language Choices: Conditions, Constraints, and Consequences,* edited by M. Pütz, pp. 361–380. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Myers-Scotton, C. 1979. Explaining Linguistic Choices as Identity Negotiations. In *Language: Social Psychological Perspectives,* edited by H. Giles, W. Robinson, and P. Smith. New York: Pergamon Press.
———. 1993a. Duelling Languages. Grammatical Structure in Codeswitching. Oxford & New York: Oxford University Press.
———. 1993b. Social Motivations for Codeswitching: Evidence from Africa. Oxford & New York: Oxford University Press.
———. 1993c. Common and uncommon ground: Social and structural factors in codeswitching. In *Language in Society*. December 1993, Vol. 22, No. 4, pp. 475–503. Cambridge: Cambridge University Press.

———. 2001. The Matrix Language Frame Model: Development and responses. In *Codeswitching Worldwide II*, edited by R. Jacobson, pp. 23–59. Berlin & New York: Mouton de Gruyter.
———. 2002. Contact Linguistics: Bilingual Encounters and Grammatical Outcomes. Oxford University Press.
Myers-Scotton, C. & Jake, J. 2000. Explaining Aspects of Codeswitching and their Implications. In *One mind, two languages: Bilingual language processing*, edited by J. Nicol, pp. 91–125. Oxford: Blackwell.
Mäkilähde, A. 2019. The Philological-Pragmatic Approach: A Study of Language Choice and Code-Switching in Early Modern English School Performances. Turku: University of Turku.
PISA, Program for International Student Assessment. 2018. Accessed 24.3.2020 https://www.oecd.org/pisa/publications/pisa-2018-results.htm
Poplack, S. 1980. Language: Sometimes I'll Start a Sentence in Spanish Y TERMINO EN ESPANOL: Toward a Typology of Code-switching. January 1980, Vol 18, No. 7–8.
Poplack, S; Wheeler, S. & Westwood, A. 1989. World Englishes: Distinguishing language contact phenomena: evidence from Finnish-English bilingualism. November 1989, Vol. 8, No. 3, pp. 389–406.
Romaine, S. 1995. Language in Society. Oxford: Blackwell.
Sankoff, D.; Poplack, S. & Vanniarajan, S. 1990. Language Variation and Change, 2: The case of the nonce loan in Tamil. March 1990, Vol 2, No. 1. Cambridge University Press.
Schendl, H. & Wright, L. 2011. Code-switching in Early English: Historical Background and Methodological and Theoretical Issues. In *Code-switching in Early English,* pp. 15–46. Berlin: The Deutche Nationalbibliographie.
Sebba, M. 2012. Researching and Theorising Multilingual Texts. In *Language Mixing and Code-Switching in Writing*, edited by Sebba, M.; Mahootian, S & Jonsson, C. New York: Routledge.
Statistics Finland. 2020. Accessed 14 March 2020. https://www.tilastokeskus.fi/tup/suoluk/suoluk_vaesto_en.html.
Tuc, H. 2003. Vietnamese-English Bilingualism. Patterns of Code-Switching. New York: Routledge.
Video Advertising Bureau (VAB). 2020. Accessed 9 January 2020. https://theVAB.com.
Warschauer, M; El Said, G. R. and Zohry, A. G. 2002. Language Choice Online: Globalization and Identity in Egypt. July 2002, Vol. 7, No. 4.
Winford, D. 2003. Code Switching: Linguistic Aspects. An Introduction to Contact Linguistics. Malden, MA: Blackwell Pub.
Wood, A. & Smith, M. 2004. Online Communication. Linking Technology, Identity & Culture. New York: Psychology Press.

Appendix 1 Tweets

2010:
-cool. i just noticed that in my last reply to you joka toinen sana oli enkkuu ja joka toine suomee :D (11.04.2010)

2012:
-What rymes with Englanniksi… (18.07.2012)
-I speak always like that. suomee ja enkkuu sekasin. I will tell you when I watch it. (03.12.2012)
"Hm it doesn't seem like she's eating :D ok osaan enkkuu" (04.12.2012)

2013:
-Better gamebird in pivo than ten at branch (21.10.2013)
-On the road! Seo menoo ny. Helsinki here we come, vapiskaa. (17.10.2013)

2014:
-Everytime I think of school I get very anxious my whole last year went miten sanot "päin vittua" enkuksi? (10.07.2014)
-Svitsillä ekaan hittiin ja gräbi ennen ländäystä. Kova rani! #halfpipe #sotshi #anglismi (18.02.2014)
-Haha, "We fight back with sisu" –JM Latvala. (03.08.2014)
-Best #Finglish I saw last week: the label above some flat's mail slot reading "no mainokses" (cf. En "no adverts" / Fi "ei mainoksia") (19.01.2014)
-#ios8 and I have something in common: we both speak #Finglish "… Right now it would take you about 7 minuuttia to drive home" (18.09.2014)
-"There is no way I am going to end up having a hirvikolari in central Helsinki", tells Matti Murtoniemi #Finglish (10.6.2014)
-There's no place like home. Unless there's homevaurio. #finglish (29.01.2014)

2015:
-its because if some ppl follow us they understand what we talk about so… puhu vaan enkkuu. (26.03.2015)
-TWD:tä katsellessa kuultua: Kohta tuo turnaa zombieksi. Turnaa? #anglismi vailla vertaa. (19.10.2015)
-En mä nyt lähtis linkkaamaan #stubb #anglismi (15.04.2015)
-Making of Salibandyliigan puolivälierä, featuring kapteenit erotuomarikopissa! #sekakieli (12.03.2015)
-Paluu to the front of the sorvi. (07.01.2015)
-There is some wifi in my bussi! (28.10.2015)
-Hey fuck what kello on? (30.05.2015)
-Can't wait for ya'll to see my costume for penkkarit! Penkkarit? Umm. Party for lukuloma. I mean, studying holiday(?) Wtf anyway #Finglish (26.01.2015)

2016:
-Messukeskus on enkuksi exhibition center, pieni ekshibitionismi siis kuuluu asiaan! (28.02.2016)

-Petteri Orpo lupasi ottaa ”partien” kanssa keinot käyttöön. Eikä ministeri enää osaa suomea? (02.09.2016)
-Yhyy itken ku oon niin awkward puhuu enkkuu mut se oli niin sulonen ja halus musta kuvankin. (17.01.2016)
-siis lmao puhuin mun uuden roomien kaa vissii 15 min enkkuu ennen ku esiteltii toisemme ja tajuttii et ollaa suomalaisia (25.10.2016)
-Niin, sitä tarkoitinkin. :D Minä downgreidaan toinen upgreidaa. Sori #anglismi (02.10.2016)
-Nextille levelille. Uudelle levelille. Mitenkäs olisi ihan jos: uudelle tasolle. #Kiitos suomen kieli #anglismi (06.02.2016)
-Ja viran puolesta vähän kirpasee, kun kotona lapset ”evolvaa”, ”transferraa” ja ”poweruppaa” Pokemoneja. #anglismi #hyväkieli (06.10.2016)
-TPS aloittaa kautensa isosti ”opening gamella”. Jopas nyt, ei hypätä suoraan toiseen kotiotteluun? #anglismi (18.05.2016)
-”Missä mun luistimet? Mä niidaan niitä koulussa.” Siis merkityksessä ”I need them”. #anglismi (14.01.2016)
-Sä venaat semioottist ouppenii lyriikalle ja bonjaat et rest in peace suomen kieli #semiootiikka #lyriikka #anglismi (17.02.2016)
-third, or strongly in the #rallienglanti moodissa (29.07.2016)

2017:
-”Puhun suomee ja enkkuu myös sekasin kun haluun inkluudaa ihmisii jotka ei puhu suomee mut välillä en jaksa ajatell mite tää ilmaistaan enkuks” (29.04.2017)
-Selasin kanavia. Eteeni tuli #vainelämää. Cheek sanoi ”yritin kapturoida sen tunnelman.” Vaihdoin välittömästi kanavaa. #suomi #anglismi (08.09.2017)
-Eikö keynotelle ole toimivaa suomenkielistä vastinetta? Vai loppuiko merkit kesken? ;) #anglismi (28.04.2017)

2018:
-Ennen kuin kukaan lähtee kiivailemaan ”Suomi polkee ihmisoikeuksia”, kannattaa huomata, että jutut ovat fake news! (17.04.2018)
-Joku kysy Jodelissa mikä on kuitti enkuksi ja siihen oli vastattu cuith. (26.01.2018)
-Deadline on Syyskuun vika päivä! Sen jälkeen ei enää pääse mukaan. (24.09.2018)
-Silitin ruskeaa kiharakarvasta labradoria ja ku sen omistaja puhu enkkuu olin sillee I just wanted to say you have a really nice dog ja osotin koiraa nice dog!!! Molemmat ilahtu!! Profit 100% (16.06.2018)
-…Onneksi tittelit eivät sentään ole äidinkielellä. Commercial Director on tosi paljon enemmän awesome kuin kaupallinen johtaja yms. Just sayin’ #anglismi (10.10.2018)
-Dronelle ehdotettu kekseliästä suomenkielistä vaihtoehtoa: ilmuri.. (18.09.2018)
-It made sense että mä julkaisen niitä biisejä mitä mulla on. Oon thank god finally löytänyt itselleni oikean tiimin. (10.10.2018)
-We coined the term #greedaus – being greedy for the 60% products from #smarket. --- “te olette greedaajia” (14.04.2018)
-LMAO Petteri! Ylä-mummo doesn’t translate so well into English. But yeah, I have some feeling Paris doesn’t even know what hockey is… (02.07.2018)
-Myös hot tip: jos sua vituttaa, ettei lapsesi eväsleipään voi laittaa kinkkua ja juustoa, älä laita kinkkua ja juustoa. Laita vaikka kasvipohjaista levitettä (esim. tartex), tai vegejuustoa ja kasviksia. Fuck yeah miten iisiä. (22.10.2018)

-Tuo oli niin tuore asia ettei ollut ehtinyt newsseihin, mutta nyt on. (5.10.2018)

2019:
-First munkkis and sima of this vappu enjoyed with the team at the office. Happy Vappu! (30.04.2019)
-Tuli sähköpostilla enkuksi tieto henkilökunnan tapaamisesta. Lisäksi oli kirjattu loppuun ”Coffee and pulla is available at 13:00.” Arvostan. (18.03.2019)
-Joo jes! ja englanniksi bongi on bong. Hits from da bong kuten Cypress Hills laulaa luritteli aikoinaan. (09.11.2019)
-ytl. Ole suopea mulle I deserve this. (21.03.2019)
-Don’t say like that oon aina miettiny kuinka huonoo mun enkku on verrattua sun kaikkiin päivityksiin ja millasta enkkuu sä käytät arjessa. Oot oikeesti hyvä. (11.09.2019)
-dlsjkdlskfld rip us then – and yeah it was but it brings me STRESS like I don’t have time for school and stanning bts and skz at the same time ja en tiiä miks me puhutaan enkkuu (10.04.2019)
-puhuin tänää mun pomolle enkkuu and after like 5 mins I realized that she probably can’t speak English that good and didn’t understand half of it so that’s my day at work then vittu (13.03.2019)
-joo siis samaistun, toi olis omast mielest paras mahollisuus saada mistään L. Mut wtf that’s annoying of ur dad, enne ei ollu enkus läheskää nii kova taso ku nykyää ku suomalaisnuoret käyttää enkkuu ihan hitosti (05.10.2019)
-ole englannin kielen asiantuntija & opettaja. >”haha lol mut kaikkihan enkkuu osaa” >ok cool jep näinpä >”mut hei voisitsä lukee tän mun esseen/hakemuksen/CV:n läpi ku se pitää tehä enkuks?” >… (11.08.2019)
-Go Niinistö! ”PERKELE” NICE Show President Trump, Creetings From Finland Suomi Komeeta Niinistö (02.10.2019)
-Suosikkireaktoni Ylen henkilöjuttuun musta so far: --- (22.09.2019)
-Podcastia pukkaa, coming soon! kiitos vierailusta viime vuoden Battlen voittaja (tweeter’s name), tulevan päivän moderaattori (tweeter’s name), Verohallinnon viestintäpäällikkö (tweeter’s name) ja loistava haastattelijamme (tweeter’s name) #kojufilm #luovuus #podcast#businessbattle (19.09.2019)
-kaksi nuorta jonnea juttelee ruuhkabussissa autoista. Toinen epäilee, toinen intoilee. ”Joo, onhan se negatiivinen asset, mut kiihdyttäminen on satisfying. Se on hauskaa!” ”Ketä se kiinnostaa? Miksi hankkia auto Länsiväylällä ajeluun? Pelkkä egoboostii. Metro ja bussi riittää.” (18.9.2019)
-”Kuinkahan vahvasti valokuvamuistin omaavat photoshoppaa muistojaan?” (19.09.2019)
-”First we give them siima then we pull the matto alta” (19.09.2019)
-Teidän podcastin influenssit näkyy jo. Ei ihan Hs kulttuuripalsta, mutta lähellä #anglismi (27.02.2019)
-Kielen bastardisoituminen. (a picture where the one says in Finnish: Jou, gubbella coolit lainit, meitsi semisti diggaa) (09.04.2019)
-“If I can just get some makaronilaatikko, I will be just okay” (01.06.2019)
-Tuohino:D, “even every 3ilometric” said at the end of stage #rallienglanti (15.02.2019)
-“Pane astiat sinkiin” Put the dishes into the sink. (08.01.2019)
-“Missä on paatiruuma?” Where is the bathroom? (08.01.2019)
-Oh how bittersweet thou are, the vitutation of the spring Saturday imuration. #Finglish (23.3.2019)

-lol mut real talk, siit tulee niin syyllinen fiilis et why can't I be happy and excited about this suggestion to meet up, what am I a total fucking dumbass loser (11.11.2019)
-Tänään pari afterwork olutta, huomenna iisiä ja vähä videopelejä, sunnuntaina aamulla kuudelta ylös -> #F1fi kauden avaus, parin tunnin nokoset ja lentokentälle kohti Meksikoa. Nään siis sittenkin kauden avauksen, WOOP WOOP! (15.03.2019)
-Ignooraaminen voi olla iisiä mut miksi avoidaa. Nautitaan kielen rikkaudesta ja erityisesti internetissä mulle näyttäytyy enemmän kielen rajoitteellisuudelta kuin päinvastoin jos välttää sanomasta "cringe" twiitissä, vaikka sopisi kontekstiin paremmin kuin myötähäpeä. (24.12.2019)
-Tämä(kin) tiedon lisäämä tuska on arvokasta. ilmeisesti vain FB:n feediä tarkkailemalla arvioidaan, mikä poliittinen puoli ketäkin targetoi. Laajennusvaraa muille alustoille ja kaupallisiin mainoksiin. Kuhunkin kohdistetun euro määrän arviointi olisi iisiä. (12.02.2019)
-Merry Christmas to y'all my beloved twitter friends. You've changed my life. Love you. KIPPIS FROM FLORIDA. (24.12.2019)
-Mun bestis otti itsestään niin kauniin kuvan, että laitoin sen mun kännykän taustakuvaks. Nyt ihmiset: "kukas hän on" semmoisella äänensävyllä, kuin kyseessä olisi mun gf. Kuulkaas hetero- ja parisuhdenormatiiviset: myös ystävyyssuhteet voivat olla rakkaita ja tärkeitä! Friends <3 (11.12.2019)
-Onko sinulla nirso lapsi? Millä keinoin olet saanut lapsen maistamaan uusia ruoka-aineita ja onko nirsoilusta päästy eroon? Kyllä, tämä on taustatyötä. Kuulette myöhemmin lisää siitä, mihin. But it's awesome. (12.12.2019)
-Haluan lisää hehkuttaa cheerleadingiä. (09.12.2019)
-Poika on meillä kattonut tota. Mulla vasta Breaking Bad vaihe. Thingssit tulossa. (12.07.2019)

2020:
-Sosioekonominen asemani ei ole ihan topissa (06.01.2020)
-Aaaaaaaah. Sain maalattua 8 goblinia valmiiks. Voittaja fiilis! Enää 2 perytonia viellä.. (09.01.2020)
-Vähän sellanen fiilis, että voisi lyötä heti aluksi trigger warning (/sisältövaroitus) leiman koko vuoteen 2020.. jos kuitenkin kaikesta huolimatta tulisi iloinen 20-luku. (06.01.2020)
-Kiitos (tweeter's name) kun toit tämän tietoisuuteen. Juuri kun oli Sabbath Assemblyn repeatilla kuuntelu on hellittänyt, tuli tästä uusi repeat-tason biisi. Ja jos muut on yhtä hyviä, niin sama hoito niille. (11.01.2020)
-OMG! Se tapahtui! Kuulun nyt SaunaVihdan Stream Teamiin. Aivan mahtava fiilis! (11.01.2020)
-Eilen closattu Nokian bull sertit 11% voitolla viime dipistä. Eilen myös avattu Finnair longi ja samaten Tesla shortti. Tesla aivan törkeästi ylihinnoiteltu, suhteellisen varma fiilis tästä treidistä. #sijoittaminen (11.01.2020)
-Taas sellainen fiilis et vois vaan vetäytyä siihen omaan kuoreen ja laittaa muu maailma ignoreen. (11.01.2020)
-Se fiilis kun very-ex wannabe kestävyysurheilija, nykyinen koirankusetuslenkittäjä saa huikeen idiksen lähtee #crossfit alkeiskurssille… Aivan hitokseen kivaa… 3 treenikertaa takana ja sattuu aivan prk*sti (10.01.2020)
-fiilis kun opettaja callouttaa siitä että on kolmatta kertaa samalla kurssilla (09.01.2020)
-Joku pitkä elokuva ja teevee sanoo stop liian jännää wtf (3.1.2020)

Appendix 2 Finnish summary

Johdanto

Tämä tutkielma tutkii suomen ja englannin kielen välistä koodinvaihtoa (code-switching) sosiaalisessa mediassa, tarkemmin sanottuna Twitterissä. Tutkielman avulla halutaan tuoda lisätietoja kielen luonteesta verkkoympäristössä sekä nähdä millä lailla suomalaiset kaksikieliset käyttävät koodinvaihtoa nettikirjoituksessaan. Koodinvaihdolla tarkoitetaan kahden tai useamman kielen käyttämistä yhdessä diskurssissa, ja näiden kielten ominaisuuksien sekoittamista kommunikaatiossa (Gumperz 1982). Koodinvaihtoa on tutkittu paljon, mutta useimmiten se mielletään puheen ilmiöksi. Nettikirjoitus on ajateltu olevan jotain puheen ja kirjoitetun kielen välillä, mutta tässä tutkimuksessa sitä kohdellaan kirjoitettuna kielenä, sillä tutkimuksen keskiössä on kieli ja sen rakenne. Kirjoitettua koodinvaihtoa on tutkittu paljon vähemmän kuin puhuttua koodinvaihtoa. Sosiaalisen median käyttö on kasvanut hurjasti viime vuosina ja Twitter yhtenä sosiaalisen median alustana tarjoaakin mielenkiintoisen aineiston tutkia kielenkäyttöä verkossa. Tässä tiivistelmässä kerrotaan ensin tutkimukseen käytetyistä teorioista ja metodeista. Tämän jälkeen siirrytään aineiston keräämiseen ja aineiston esittelemiseen. Lopuksi käydään läpi analyysi ja keskeiset tutkimustulokset.

Teoreettinen viitekehys

Koodinvaihto ilmiönä tarkoittaa kahden tai useamman kielen käyttämistä vuorovaikutustilanteessa. Sitä on tutkittu eri aloilla paljon, ja se on yhdistetty usein enemmänkin puhuttuun kuin kirjoitettuun kieleen. Koodinvaihdossa henkilö voi alkaa puhua yhdellä kielellä ja lennosta vaihtaa kielestä toiseen. Gumperzin mukaan vaihto tapahtuu usein ilman epäröintiä tai tauotusta. Joskus henkilö saattaa vaihtaa kieltä ainoastaan yhden sanan verran, joskus jopa kokonaiseen lausekkeen verran. Koodinvaihto voi tapahtua tietoisesti tai tiedottomasti, ikään kuin puhujan huomaamatta. Syitä koodinvaihtoon voi olla sanan käyttö harrastukseen liittyen, solidaarisuutena muita puhujia kohtaan tai sanan hetkellinen unohtaminen toisella kielellä. Etenkin nuoret henkilöt suomessa käyttävät englanninkielisiä ilmauksia ja sanontoja osana muuten suomenkielistä puhettaan. Koodinvaihdon

edellytyksenä on se, että keskustelukumppanit ymmärtävät koodinvaihdosta huolimatta puheen merkityksen. Joskus koodinvaihtoa voi käyttää välineenä jättää osa keskustelutovereista keskustelun ulkopuolelle, valitsemalla sanan, jonka tietää vain osan ymmärtävän.

Koodinvaihto perustuu siis siihen, miten *kaksikielinen* henkilö yhdistelee osaamiensa kielten osia kommunikaatiossaan. Kaksikielisyys tarkoittaa sitä, että henkilö sujuvasti tuottaa ja ymmärtää kahta kieltä. Tutkijat ovat olleet eri mieltä siitä, pitäisikö henkilön olla natiivi, eli synnynnäinen puhuja, molemmissa kielissä ollakseen kaksikielinen. Tässä tutkimuksessa kaksikieliseksi luetaan kuitenkin kuka tahansa henkilö, joka osaa kahta kieltä sujuvasti. Suomessa kaksikielisyys ei ole uusi ilmiö: Suomessa on vuosikaudet ollut kaksi virallista kieltä, suomi ja ruotsi, ja suomenruotsalaisia onkin maassamme noin 5,2 %. Monet suomalaiset osaavat puhua suomen lisäksi myös englantia sujuvasti, joten kaksikieliset suomalaiset ovat yhä useammin kaksikielisiä suomen ja englannin kielissä. Mikäli henkilö puhuu useampaa kuin kahta kieltä sujuvasti, häntä kutsutaan *monikieliseksi*. Kun henkilö taitaa monia kieliä, hän helposti käyttää useamman kielen sanoja ja rakenteita puheessaan ja kirjoituksessaan. Tässä tutkimuksessa sekoitettua suomen ja englannin kielenkäyttöä kutsutaan sanalla ”Finglish” (Finnish & English). Koodinvaihtoa on aikojen saatossa tutkittu paljonkin, mutta sen on ajateltu olevan enemmän puheessa kuin kirjoituksessa tapahtuva ilmiö. Verkossa esiintyvä kieli mielletään usein olevan puhekielen ja kirjakielen välimalli. Tässä tutkimuksessa sitä käsitellään kuten kirjoitettua kieltä, koska kielen interaktiivisuus jätetään analyysin ulkopuolelle. Tarkemmat tutkimuskysymykset selitetään tuonnempana.

Internetin leviäminen on vaikuttanut paljon ihmisten kielenkäyttöön. Sen ansiosta englannin kielen käyttö kommunikaatiossa on tullut yhä tärkeämmäksi ympäri maailmaa. Sosiaalisessa mediassa keskustellaan usein erikielisten ihmisten kanssa, ja siksi englanti on vakiinnuttanut kielensä internetin yleiskielenä (lingua franca). Internetissä tapahtuva kommunikointi poikkeaa monin tavoin perinteisestä keskustelusta tai perinteisestä kirjoitetusta kommunikaatiosta. Verkossa tapahtuva kommunikaatio on yleensä nopeaa, kirjoitusten sisältöä ei harkita välttämättä tarkkaan, ja myös lyhenteet ovat käytössä merkkien rajallisuuden vuoksi. Esimerkiksi Twitterissä kirjoittajalla on käytössään 280 merkkiä per viesti. Sosiaalisessa mediassa kirjoitukset ovat epävirallisia, joten kirjoitusvirheitä on myös usein enemmän kuin virallisessa tekstissä (esimerkiksi essee tai työhakemus). Eri sosiaalisen median alustat ovat keskittyneet eri asioihin, esimerkiksi, Instagram kuviin, YouTube videoihin ja

Twitter taas verkostoitumiseen. Twitteriä käyttää yli 300 miljoonaa ihmistä aktiivisesti, ja sen käyttäjäkunnan on ajateltu olevan iältään vanhempaa kuin muiden sosiaalisten medioiden (Snapchat, TikTok ym.). Rajallisen merkkimääränsä vuoksi twiitteissä käytetään lyhenteitä, kirjainlyhenteitä (akronyymejä) sekä tunnisteita (hashtag), joilla kategorisoidaan kirjoituksia. Monet poliitikot ja muut vaikutusvaltaiset henkilöt käyttävät usein Twitteriä julkisesti tavallisten ihmisten lisäksi.

Metodit

Tämän tutkimuksen ideana on tutkia millaisia erityyppisiä koodinvaihtoja Twitterissä julkaistuista tekstinpätkistä, eli twiiteistä löytyy, sekä löytyykö niitä eri tasoilta. Tätä tutkimusta varten yhdisteltiin eri metodeita ja teorioita. Tyypeillä tarkoitetaan kielitieteilijä Poplackin tekemää jakoa virkkeensisäiseen, virkkeiden välillä tapahtuvaan sekä irralliseen koodinvaihtoon. Virkkeensisäistä koodinvaihtoa tutkittiin lisäksi kielitieteilijä Myers-Scottonin kehittelemän matriisikielimallin mukaan. Matriisikielimallin mukaan virkkeensisäisessä koodinvaihdossa toinen kielistä on hallitseva matriisikieli (pääkieli A), johon upotetaan toisen kielen B aineksia. Kielen B koodinupotukset ovat yleensä sanan tai lausekkeen mittaisia. Mallia hyödyntämällä pyrittiin selvittämään, kumpi kielistä (suomi vai englanti) on useimmin matriisikieli ja kumpi B kieli. Kielitieteilijä Montes-Alcalá on tutkinut koodinvaihtoa eri tasoilla, ja hänen mukaansa sitä tapahtuu sanatasolla, lausetasolla ja virketasolla. Twiitit jaettiin näihin tasoihin ja pyrittiin selvittämään tasojen vaihtelua määrällisesti. Lisäksi koodinvaihtoa tutkittiin kielitieteilijä Muyskenin teorian mukaan. Twiiteistä etsittiin Muyskenin esittämiä *koodinupotuksia*, jossa yhden kielen aineksia esiintyy toisen kielen määrittämässä lauseessa, *koodien vuorottelua*, jossa eri kielten lauseet seuraavat toisiaan vaikuttamatta toistensa kielioppeihin, sekä *kongruenttia leksikalisaatiota*, jossa kielen käyttäjä täyttää eri kielten sanastolla kielten yhteiset kieliopilliset rakenteet. Analyysi tehtiin teorioiden pohjalta ja monien eri esimerkkien kautta pyrin näyttämään miten koodinvaihtoa esiintyy suomen ja englannin kielen välillä. Keskiössä oli myös selvittää onko kirjoitetulla koodinvaihdolla samankaltaisuuksia puhuttuun koodinvaihtoon. Analyysin lopussa tutkin lisäksi koodinvaihdon vaikutusta sanojen taivutukseen. Myös koodinvaihto sanaluokkien välillä oli mielenkiinnon kohteena. Tutkimuskysymyksiä oli kolme ja niiden avulla pyrittiin selvittämään seuraavat asiat:

1. Mitä koodinvaihdon tyyppejä ja tasoja suomalaiset käyttävät twiiteissään, ja mikä mahdollisesti saa ihmiset käyttämään koodinvaihtoa?
2. Matriisikielimallia hyödyntäen, minkä tyyppistä virkkeensisäistä koodinvaihtoa esiintyy suomalaisten twiiteissä?
3. Mitä sanaluokkia vaihdetaan kieleltä toiselle ja miten koodinvaihto vaikuttaa sanojen taivutukseen?

Analyysiosiossa sekä pohdintakappaleessa tutkimustulokset tullaan esittämään näiden tutkimuskysymysten varjossa.

Datan keräys ja analysointi

Tämä tutkimus keskittyy kirjoitettuun koodinvaihtoon Twitterissä suomen ja englannin kielen välillä. Twitter sosiaalisen median alustana tarjosi mielenkiintoisen tilaisuuden kerätä materiaalia, sillä Twitteriin ihmiset usein kirjoittavat lyhyitä tekstinpätkiä, joiden sisältöä harvemmin kunnolla suunnitellaan etukäteen. Aineisto kerättiin syksyllä ja talvella vuosina 2019–2020 ja yhteensä niitä oli 88 kappaletta. Kaikki tekstinpätkät, eli twiitit, olivat julkisilta tileiltä, joten ne ovat sosiaalisessa mediassa kenen tahansa saatavilla. Oletuksena on, että kaikkien valittujen twiittien kirjoittajat ovat äidinkieleltään suomenkielisiä. Henkilöiden nimiä ei tässä tutkimuksessa mainita, sillä analyysi keskittyy enemmän twiittien kieleen ja kielelliseen rakenteeseen kuin vuorovaikutukseen. Tästä syystä itse twiitin temaattisella sisällölläkään ei tämän tutkimuksen kannalta ole väliä. Twiitit kerättiin käyttämällä eri hakusanoja, kuten ”Finglish, enkku, iisi, fiilis, rallienglanti…”. Aineistoa kerättäessä oli pakko käyttää hakusanoja, sillä Twitteriin julkaistaan miljoonia kirjoituksia joka päivä. Hakusanat olivat myös suomenkielisiä, sillä englanninkielisillä hakusanoilla ei tullut vastaan suomi-englanti -koodinvaihtoa sisältäviä twiittejä. Ainoa ehto twiittin pääsemiseksi osaksi aineistoa oli se, että se sisälsi koodinvaihtoa suomen ja englannin kielen välillä. Hakusanat olivat melko kielipainotteisia, joten tämä on huomioitava aineiston puolueellisuudessa.

Analyysi aloitettiin jakamalla twiitit ensin eri tyyppien mukaan. Tyypit tässä tutkimuksessa ovat: virkkeensisäinen, virkkeidenvälinen ja irrallinen virkkeestä. Virkkeensisäistä koodinvaihtoa oli 73 % twiiteistä (65 kpl), virkkeidenvälistä koodinvaihtoa oli 15 % twiiteistä (13 kpl) ja irrallista koodinvaihtoa oli 12 % twiiteistä (11 kpl). Kävi siis ilmi, että koodinvaihtoa tapahtuu eniten virkkeensisäisesti. Seuraavana keskityttiin koodinvaihtoon

virkkeen sisällä ja pyrittiin määrittelemään matriisikielimallin avulla matriisikieli. Kävi ilmi, että suomen kieli oli matriisi-, eli pääkieli, suurimmassa osassa twiitteistä (57 %). Tämä oli oletettua, sillä tutkimuskirjallisuuden mukaan koodinvaihto tapahtuu yleensä siten, että omaan äidinkieleen otetaan vaikutteita toisesta kielestä. Englanti oli matriisikieli 22 % twiiteistä. Seuraavana twiiteistä määriteltiin koodinvaihdon eri tasot. Tasot tässä tutkimuksessa tarkoittavat sanatasoa, lausetasoa ja virketasoa. Sanatason koodinvaihtoa oli 89 % twiiteistä, lausetason koodinvaihtoa oli 11 % twiiteistä ja sama määrä 11 % tapahtui myös virketasolla. Aineiston perusteella koodinvaihtoa tapahtuu siis eniten sanatasolla, eli yksittäisiä sanoja sanotaan virkkeen sisällä toisella kielellä kuin muu teksti. Aineiston mukaan koodinvaihtoa tapahtuu eniten sekä virkkeen sisällä että sanatasolla, joten tämän tiedon löytäminen on todiste tutkimuksen luotettavuudesta. Kaikki nämä löydökset osoittavat, että kirjoitettu koodinvaihto ei ole mielivaltaista, vaan se tapahtuu järjestelmällisesti ja on verrattavissa suulliseen koodinvaihtoon. Sanojen taivutusta analysoidessa tuli ilmi, että koodinvaihdossa vaihdettu sana taivutetaan useimmiten matriisikielen mukaan. Esimerkiksi, jos lause on suomenkielinen, ja yksittäinen sana vaihdetaan englanniksi, se luultavasti taivutetaan kuitenkin suomen kielioppisääntöjen mukaan. Tapauksia löytyi twiiteistä useita ja molemmilla kielillä: ”puhuin mun uuden *roomien* kans”, ”first *munkkis* and *sima* of the *vappu*”. Kieli pysyy sujuvampana ja ymmärrettävämpänä, mikäli sen osat ovat keskenään rakenteellisesti sopusoinnussa. Tämä lienee syynä siihen, miksi koodinvaihdetut sanat taipuvat matriisikielen mukaan. Sanaluokista tehtiin huomio, että substantiivit vaihtuvat kielestä toiselle useimmiten (63 %), mahdollisesti siitä syystä, että niitä esiintyy kielessä eniten. Verbit olivat toiseksi isoin sanaluokka koodinvaihtumiselle (23 %) ja adjektiivit kolmanneksi suurin (13 %). Huomionarvoista on myös se, ettei pronomineja vaihdettu kertaakaan kieleltä toiselle.

Pohdinta

Ensimmäisen tutkimuskysymyksen avulla pyrittiin selvittämään koodinvaihdon eri tyyppejä ja tasoja. Aineiston perusteella havaittiin, että koodinvaihtoa tapahtuu eniten virkkeensisäisesti ja sanatasolla. Syy tähän on mahdollisesti se, että Twitter sosiaalisena alustana on epämuodollinen ja ihmiset kirjoittavat sinne sen enempää sisältöä suunnittelematta. Ihmiset usein käyttävät ensimmäisiä sanoja, jotka tulevat mieleen ja lopputuloksena on tekstinpätkä,

jossa sekoittuvat eri kielet keskenään. Yksi koodinvaihdon syy on mahdollisesti jo teoriaosuudessa esitelty tekstin suuntaaminen tietylle ryhmälle, jättäen asiaankuulumattomat ymmärryksen ulkopuolelle. Tämä tapahtuu usein esimerkiksi harrastustoiminnassa tai puhuttaessa elokuvista ja tv-sarjoista. Eräs syy koodinvaihdolle voi olla myös kulttuurinen: jokin sana liittyy juhlapyhään jota ei toisenkielisessä maassa vietetä. Näin ollen sanan selittäminen olisi Twitterin mittakaavassa, ja merkkien rajallisuudessa mahdotonta, joten sana otetaan toisesta kielestä käyttöön. Kyseisiä esimerkkejä olivat englanninkieliset twiitet koskien esimerkiksi vappua, penkipainajaisia ja lukulomaa. Kirjoittajan oletuksena lienee, että lukija ymmärtää sen verran molempia kieliä, että tekstin ydin on ymmärrettävissä, tai lukija voi kysyä twiitin kirjoittajalta mitä sana tarkoittaa, jos sen merkitys jäi epäselväksi. Irrallisia koodinvaihtolausahduksia, kuten ”oh my god” tai akronyymejä, esimerkiksi ”lol” (laughing out loud) esiintyi yllättäen vähemmän mitä oletuksena oli. Tämä voi johtua siitä, että Twitterin käyttäjät ovat iäkkäämpiä kuin kyseisten ilmausten käyttäjät yleensä.

Matriisikieli oli useimmissa twiiteissä suomen kieli, ja tämä ei yllättänyt, sillä ihmiset usein puhuvat omalla äidinkielellään, johon lisäävät osia muista kielistä. Suomi oli matriisikieli 51 kertaa 88:sta. Englanti taas oli matriisikieli 19 kertaa, ja yhdistettyinä matriisikielinä (kumpikaan ei dominoi toista) oli 5 kertaa. Koska matriisikielen pystyy määrittämään vain virkkeensisäisestä koodinvaihdosta, 13 twiitissä matriisikieltä ei ollut lainkaan. Matriisikielen määrittäminen oli ajoittain hankalaa, sillä etenkin yhdistetyn matriisikielen määrittäminen ei aina ollut selkeää. Sanojen taivutus taas oli yhteydessä matriisikieleen. Suurimmassa osassa tapauksista koodinvaihdetut sanat taipuivat matriisikielen mukaan. Vaihdettu sana toimi ikään kuin osana toisen kielen kieliopillisia normeja. Jos matriisikieli oli suomi ja koodinvaihdettu sana oli englanninkielinen, englannin sanasta jäi prepositio pois ja sanaa taivutettiin lisäämällä suomen suffiksipääte. Jos taas englanti oli virkkeen matriisikieli ja siihen lisättiin suomenkielinen sana, suomen kielelle ominaiset päätteet jäivät pois ja sanan eteen sijoitettiin prepositio. Kielen sujuvuuden kannalta ilmiö oli odotettu. Lopputulemana voidaan todeta, että koodinvaihdossa pätee tietyt säännöt, eikä sitä voi tehdä täysin mielivaltaisesti. Myöskin analyysin perusteella voidaan todeta, että kirjallinen koodinvaihto sosiaalisessa mediassa on verrattavissa puhuttuun koodinvaihtoon.