'Stoefpears run the world'

The use of English code-mixing in Dutch youths' computer-mediated communication

Laura de Weger
s4472624
August 21, 2016
Master's Thesis
Supervisors:
Prof. Dr. Roeland van Hout
Lieke Verheijen, MA

Table of contents

<u>Acknowledgements</u>

The completion of this Master's Thesis would not have been possible without the help and support of several people.

First and foremost, I express my gratitude to my supervisors, Roeland van Hout and Lieke Verheijen, for their guidance, patience and enthusiasm. I always left our meetings motivated to work harder and to improve my research. I could not have asked for better supervision.

I thank my parents for always believing in me and encouraging me to give it my best effort. Their optimism gave me a realistic perspective on my own capabilities and work.

I thank my friends for their feedback, advice, and mental support, and for giving me much-needed distraction from time to time. Délisa and Syania, thank you in particular for your friendship and tremendous support during the past year – it means a lot.

To everyone who contributed to my research directly or indirectly: my sincerest gratitude.

Laura de Weger

Abstract

The aim of this thesis is to analyse and describe the use of English in computer-mediated communication by Dutch youths. The main research question is: how and how much do Dutch youths code-mix and adopt elements from the English language in their Dutch written computer-mediated communication? This question has been answered through corpus research, using a CMC corpus consisting of messages by male and female youths between the ages of 12 and 23 on MSN, SMS, Twitter and WhatsApp. Based on previous research on code-mixing, youth language and computer-mediated communication, various language-internal and -external factors that contribute to these topics have been analysed. The following ten factors are taken into account: length of switches, number of switches per CMC item, lexical category of switches, integration of switches, semantic fields of switches, intentionality of switches, frequency of switches, CMC mode, gender, and age. A quantitative and qualitative analysis of how these factors influence the use of English and interact with each other has been conducted. A total of 8619 switches to English by youths on the four different CMC modes was collected and analysed. The main conclusion from the analysis is that 2.19% of the words in the corpus were English elements, in itself a considerable amount. However, the results suggest that the Dutch youths do not communicate in English with a near-native proficiency level: although they exhibit a certain level of creativity in code-mixing, the English elements are mostly conversational words and phrases such as greetings, affective language, swear words and fixed expressions. The results imply that Dutch youths mainly use English as a part of 'teenage talk': to boost their expressivity and to distinguish themselves from older speakers of Dutch.

1 <u>Introduction</u>

A well-known, funny Dutch television commercial by coffee company Douwe Egberts shows two elderly ladies drinking a cup of coffee together, while having a conversation in stereotypical Dutch youth language.[1] This short commercial gives an impression of the language used by Dutch youths, and how it differs from Standard Dutch: hearing two elderly ladies use this type of language is highly unusual and thus generates a comical effect. Notable is the number of English words they use; some examples are *check, speaker, chill,* and *bitch.* Although this commercial displays a stereotype, it is very spot-on; English is in fact a part of many Dutch youths' everyday speech.

Even though Dutch people grow up largely monolingual, the English language plays a large role in the life of the average Dutch citizen; not a day goes by without being exposed to it. Given the power over the western world of the United States of America and Great Britain, countries where English is the official language, and given the fact that language contact is bound to lead to code-mixing, it is no wonder that the Dutch tend to code-mix with English now and then.

The internet plays a large role in most Dutch people's everyday lives as well. Especially with the recent popularity of the smartphone, people can go online wherever and whenever they want. Youths communicate via the internet in a unique way, using features such as emojis, abbreviations and non-standard spelling variants and punctuation, to convey pragmatic and prosodic parts of language which cannot be conveyed through text the same way as in face-to-face conversations.

Considering the fact that both the English language and the internet play a role in the way Dutch youths communicate nowadays, it is not surprising that English is used in their online communication as well. It is not uncommon to see a teenager sending a message to their friend on the mobile chat application WhatsApp, calling something 'cute' or 'awkward', instead of their Dutch translations ('schattig' and 'ongemakkelijk' respectively), or send a response using abbreviations such as 'lol', 'omg' or 'wtf'.

Both code-mixing and computer-mediated communication are recent popular topics for study and their unique features may cause them to interact in interesting ways. This is why I combine the two here. The general research question I aim to answer in my thesis is as follows:

> How and how much do Dutch youths code-mix and insert elements from the English language in Dutch written computer-mediated communication?

Because this is a very broad question, I have divided it into a number of sub-questions, which indicate the various aspects I have analysed. The sub-questions can be found in chapter 4.

I have attempted to quantitatively and systematically analyse many different factors that affect code-mixing, as well as provide a qualitative analysis of especially interesting and notable cases of code-mixing. I have studied the use of English in Dutch CMC, using a corpus that consists of a large collection of Dutch tweets, text messages, WhatsApp messages and MSN

---

[1] https://www.youtube.com/watch?v=ATJ-DqKyRxQ

chat messages, written by males and females between the ages of 12 and 23. This corpus provides sufficient data to analyse the various aspects of code-mixing in Dutch CMC.

In chapter 2, the societal background of code-mixing and computer-mediated communication are discussed. In chapter 3, the theoretical framework on which I have based my research and hypotheses is discussed, followed by the hypotheses in chapter 4. Chapter 5 describes the methodology used. In chapter 6, the results are given and discussed in detail, after which the conclusions are summarised in chapter 7.

2 <u>Background</u>

The English language is very present in Dutch people's daily lives. This is not limited to adults; it goes for teens and children too. From a very young age, Dutch children are exposed to English. Not only do they learn English at primary and secondary school, but they are also surrounded by the language in their free time (Cenoz & Jessner, 2000). The radio plays music with English lyrics, many English-spoken programmes on TV air with original audio and Dutch subtitles, and video games targeted at teen and adult audiences are often not translated at all. These are just a few examples of how the Dutch are exposed to the English language on a daily basis. It can be debated to which extent English is still a foreign language to the people growing up in the Netherlands at all, and to which extent it has become a second mother tongue. Be that as it may, bilingualism cannot but lead to code-mixing and lexical borrowing between the speakers' languages, although they are not always conscious of that (Myers-Scotton, 2002). Because of the huge amount of English they are exposed to throughout their lives, it is not a surprise that the Dutch have adopted many English words and phrases into their Dutch vocabulary.

2.1 Word borrowing and code-mixing

Word borrowing and code-mixing are very common phenomena which occur in languages all over the world. But the quantity in which the Dutch borrow from and code-switch to English is noteworthy. This is a relatively recent development, as Before the Second World War in the 1940s, Dutch mostly borrowed from French and German (Van der Sijs, 2009). What often leads to word borrowing is code-mixing. When a speaker alternates between two or more languages within a single conversation, it is called code-switching or code-mixing. Nowadays, code-mixing happens mostly in certain areas, such as commercials, job adverts and business communication (Zenner, Speelman & Geeraerts, 2013; Van Meurs, Korzilius & Den Hollander, 2006; Van Meurs, Korzilius & Hermans, 2004; Hornikx, van Meurs & De Boer, 2010; Gerritsen et al., 2000). But code-mixing can be found in the speech of the average Dutch citizen (Zenner, Speelman & Geeraerts, 2015). Dutch speakers mainly code-mix on an intrasentential level (sometimes using an English word even though a Dutch equivalent exists), but they also utter longer phrases and occasionally even complete sentences in English (Zenner & Geeraerts, 2015). This manner of code-mixing is notable, because the Dutch are, of course, generally no native speakers of English and their proficiency in English is usually not as high as their proficiency in Dutch (Van Onna & Jansen, 2006). This raises the question what motivates the switches to English. It is a well-known folk linguistic belief that the Dutch use English commonly in their daily life. However, most research so far has focused on specific contexts where English is used quite frequently in Dutch, such as jobs ads (Zenner, Speelman & Geeraerts, 2013; Van Meurs, Korzilius & Den Hollander, 2006; Van Meurs, Korzilius & Hermans, 2004) and other types of advertising, e.g. commercials (Hornikx, van Meurs & De Boer, 2010; Gerritsen et al., 2000). Little research has studied the use of English by native speakers of Dutch in their natural, spontaneous language production.

2.2 Globalisation and anglicisation

Globalisation plays an important role in the domain of language change (Meyerhoff, 2006). The contact between different cultures that is caused by globalisation leads to language contact (Meyerhoff, 2006), which in its turn causes code-mixing and word borrowing (Myers-Scotton, 2002). Because of the worldwide British colonial power in the nineteenth and early twentieth centuries and the recent North American dominance, Great Britain and the United

States of America have had a large influence on a global scale, making English a dominant language in modern western society, used as a lingua franca all over the world (Cenoz & Jessner, 2000). This dominant position of the English language has brought about a great influence on many languages worldwide (Cenoz & Jessner, 2000; Görlach, 2002) – Dutch is no exception.

The frequent use of English words and phrases in the Dutch language has sparked off much criticism. There are organisations of language purists, such as *Stichting Nederlands, Stichting LOUT* and *De Bond tegen leenwoorden*, who claim that English is a threat to Dutch. They are against the borrowing of words from English into Dutch and want to stop anglicisation as much as they can. The many organisations and individuals claiming that the influence of English threatens the Dutch language make it relevant to research how much English is really used by speakers of Dutch on a day-to-day basis in their Dutch speech and writing.

2.3 Computer-mediated communication

Another recent development in language use is computer-mediated communication (henceforth abbreviated to CMC). CMC is generally defined as 'any human communication achieved through, or with the help of, computer technology' (Thurlow, Lengel & Tomic, 2004). Examples of this type of communication are emails, text messages and posts on social networking sites. More and more people communicate with each other via the internet (Thurlow, Lengel & Tomic, 2004), for example through social networking sites such as Facebook, microblogs such as Twitter or instant messaging providers such as WhatsApp. Because CMC is quite recent, much less is known about it than about spoken language or other forms of written language. What we do know is that the often informal, spontaneous, conversational nature of CMC makes it different from the more formal written language that has been around for ages; it is a completely new way of interacting, one that contains elements of both written and oral communication (Herring, 2010). What is more, it also contains new elements which are not present in standard spoken or written language, such as emoticons, textisms (non-standard spelling variants and abbreviations) and the addition of other media such as pictures or videos (Verheijen, 2015). In addition, youths are the ones who appear to communicate via social networking sites and chat messages the most (Hargittai & Hinnant, 2008). All this makes the conversations youths have on social networking sites and in online chat very interesting to analyse. Moreover, the results from this paper can ultimately be compared to those on spoken youth language which may give insights on the differences and similarities between the two.

2.4 Lingua franca on the internet

The internet is a multinational and multilingual space to which people from all over the world can have access and add their own content (Danet & Herring, 2007). Because of the global identity of the internet, a need for a lingua franca has arisen. Since English already has the position of an important lingua franca, it is an obvious choice for a lingua franca on the internet; many second-language learners of English use both English and their native language online (Danet & Herring, 2007). Even though the non-English speaking part of the internet is growing fast, English is still the most used language online (Dor, 2004; Warschauer, Said & Zohry, 2002). Because English is such an important language online and because the internet plays such a large role in the daily lives of most people in the western world, the influence of English on other languages has increased due to the internet. Also, many internet/computer/technology-related terms are originally from English. These terms are

not often translated to other languages, because people are confronted with them in English on the internet and are unaware of whether a translation of the word in their own language exists (and if so, what it is). This makes English an influential language in the semantic field of computers, internet and technology.

3 <u>Theoretical framework</u>

3.1 Terminology

Because there are many terms related to code-switching, code-mixing and word borrowing, and various ways in which these terms are used, it is important to specify how these terms are defined in this thesis. First of all, there is a distinction between code-mixing and word borrowing. Code-mixing is a synchronous phenomenon, while word borrowing is asynchronous.

There is much debate in the field of language contact over what the term 'code-switching' exactly refers to. Some use it to refer to the alternative use of two or more languages in one conversation; others use it to refer to the alternative use of two or more languages within a single sentence. The less frequently used term 'code-mixing' usually functions as an umbrella term for the alternation between multiple languages within a single conversation.

To avoid any confusion, I follow Muysken (2000) in his use of the terms code-mixing, switch and switching:

> "I am using the term code-mixing to refer to all cases where lexical items and grammatical features from two languages appear in one sentence. […] sometimes the terms switch, switch point, or switching will be used informally while referring to the co-occurrence of fragments from different languages in a sentence." (Muysken, 2000: 1)

Because the current study analyses CMC, chat abbreviations such as 'lol' and 'omg' are also discussed. These are commonly called 'textisms'. Because my focus lies on code-mixing with English by Dutch youths, the textisms discussed all originate from English, but are used here in Dutch CMC. This is why they will be referred to as 'textism switches'.

Another term that I use throughout this thesis is 'English element(s)'. The written CMC by Dutch youths includes, as discussed in more detail in later chapters, switches to English of various lengths: single words, phrases, sentences, and textisms. To be inclusive, the term 'English element' will be used to refer to a switch to English in a CMC item regardless of its length.

The focus of this thesis lies on code-mixing/code-switching, not on lexical borrowing. Yet lexical borrowing is also mentioned, because of the close relation between the two linguistic phenomena, which are elaborated on in section 3.2. However, the English elements that are discussed in the results and discussions chapters are only referred to as switches and not as 'loans', 'loanwords' or 'borrowings'. Still, the distinction between code-mixing (or code-switching) and lexical borrowing must be clarified.

> "Code-switching is the use of two languages in one clause or utterance. As such code-switching is different from lexical borrowing, which involves the incorporation of lexical elements from one language in the lexicon of another language." (Muysken, 1995: 189)

In section 5.2.1, it is explained which criteria have been used to decide whether words are counted as a switch or not in the data of the present study.

3.2 Code-mixing

For decades, code-mixing has been a fruitful research topic. Many researchers have attempted to describe the various types of code-mixing and the restrictions under which it is possible (e.g. Poplack, 1980; Joshi, 1985). A distinction is made between intersentential code-mixing, where speakers alternate between sentences, and intrasentential code-mixing, where speakers switch within a sentence (Myers-Scotton, 1993).

Another important distinction is between alternational and insertional code-mixing (Muysken, 1995). Alternational code-mixing means that all languages involved in the code-mixing are alternated evenly. Insertional code-mixing means that the speakers mainly speak in language A (called the matrix language) and insert elements from language B (the embedded language) into their speech here and there. Because this study focuses on the code mixing of Dutch youths who speak predominantly Dutch, but switch to English occasionally, it is a case insertional code-mixing. The matrix language in this study is Dutch and the embedded language is English.

Many classic studies have approached code-mixing from a structural viewpoint (e.g. Pfaff, 1979; Poplack, 1980; Joshi, 1985; Di Sciullo, Muysken & Singh, 1986). They have mostly focused on whether there are universal and/or language-specific rules and constraints for (intrasentential) code-mixing. Most of the constraints that have been found are of a syntactic nature. A constraint that may be relevant for the present study is the constraint on switchability of closed-class items (Joshi, 1985). This constraint states that closed-class items (function words, such as determiners, prepositions, pronouns) are not subject to code-mixing, which is in line with the borrowability hierarchy.

Another constraint that may be relevant is the size of constituent constraint (Poplack, 1980), which states that constituents of a higher level such as phrases, clauses and sentences (meaning the position of the constituent is higher in the syntactic structure of the sentence) are switched more frequently than lower-level or smaller constituents, with the exception of nouns. This means that, when code-mixing, words of categories other than nouns (e.g. verbs, adjectives, adverbs) are expected to appear in a switch consisting of multiple words more often than by themselves. Because nouns are not subject to this constraint, they are expected to appear by themselves relatively more than words of other word categories.

The equivalence of structure constraint (Pfaff, 1979) states that where the switch occurs, the grammars of the two languages must overlap. This means that when two languages have many grammatical differences, it can be difficult to code-mix. Since English and Dutch are closely related and have many similarities (Millar, 2007), code-mixing should be relatively easy. One notable difference, though, is the standard word order, which is SVO (subject-verb-object) for English and SOV (subject-object-verb) with V2 (verb second) for Dutch (Fromkin, 2000); this may make code-mixing syntactically more problematic.

In recent years, the sociolinguistic side of code-mixing has been studied more and more (e.g. Hornikx, van Meurs & De Boer, 2010; De Decker & Vandekerckhove, 2012; Kytölä, 2013; Zenner & Geeraerts, 2015). The present study focuses on sociolinguistic aspects of code-mixing as well.

Many studies (e.g. Zenner, Speelman & Geeraerts, 2015) that focus on code-mixing have studied spoken code-mixing; there has not been much research into written code-mixing

(Sebba, 2012), with some exceptions (e.g. Hassan & Hashim, 2009; De Decker & Vandekerckhove, 2012; Kytölä, 2013, Vandekerckhove, Cuvelier & De Decker, 2015). Not only the linguistic properties of code-mixing are analysed, but the language-external factors as well. The focus here lies on the switches themselves rather than the syntax and grammar of the surrounding sentences. Therefore, the research on constraints is only minimally relevant to my research.

Previous research has identified various reasons for code-mixing and lexical borrowing. A common reason is when a word describes an object or concept for which there is no term in the recipient language (Millar, 2007). This causes such words to be adopted frequently. Another reason is prestige. Because English is currently one of the most prestigious languages in the world (Millar, 2007), it makes sense that people would want to switch to English.

Borrowability refers to the likelihood that a word can be adopted into another language. There is a certain hierarchy to this borrowability; some words are more likely to be borrowed than others. Many factors play a role in the borrowability hierarchy. Nouns are borrowed cross-linguistically more often than other word categories (Matras, 2007), because, as mentioned above, one of the important reasons why words are borrowed is to refer to new objects and concepts, which can be done by borrowing nouns. Also, lexical words are typically part of open classes (word classes to which new words can be added rather easily, e.g., besides nouns, also verbs, adjectives, adverbs), whereas function words are usually part of closed classes (word classes to which new words can practically not be added) (Fromkin, 2000), which makes it easier for lexical words to be borrowed cross-linguistically than function words. Because code-mixing and word borrowing are both ways of adopting words into another language, whether haphazardly or more permanently, borrowability can also be applied to code-mixing.

Taking into account the borrowability hierarchy, the size of constituent constraint by Poplack (1980) and the constraint on switchability of closed-class items by Joshi (1985), it is to be expected that nouns will make up the largest category in the single-word (and partial-word) switches in the corpus under investigation here.

Another theory states that interjections are often subject to borrowing and code-mixing, because of their unique status in a sentence: they are generally not part of the grammatical makeup of the sentence, they are neither lexical nor function words, they act as a 'satellite' in the sentence (Muysken, 1999). Because interjections stand apart from the grammar of a sentence, it is difficult to compare them to other word categories. This is why it is important to take into account what kind of influence it has on code-mixing.

These two reasons for borrowing lead to a different type of code-mixing. When nouns and verbs are adopted into another language out of lexical need – because that language does not have an equivalent for the term – it is called unintentional code-mixing. In the case of interjections, they are not adopted into the recipient language because of lexical need, but because they are easy to implement in the sentence. It can be assumed that equivalents for most interjections exist in recipient languages, which makes the code-mixing intentional.

Another type of switch that is generally not adopted into another language due to the absence of an equivalent in the recipient language is a switch consisting of multiple words. Unless there is a specific idiomatic meaning behind a phrase or sentence which would be lost in

translation, most phrases and sentences are straightforwardly translatable. If a speaker then decides to use a multiple-word switch, it is intentional.

Zenner and Geeraerts (2015) analysed all switches to English consisting of more than one word in a corpus of Dutch speakers. They found that most of those switches were (semi-)fixed expressions. Various methods were used to determine their 'fixedness', for example, looking them up in multiple dictionaries and on google. They suggest that these fixed expressions are copied as whole from the source language and inserted into the recipient language, which makes them more similar to traditional loanwords than to creative code-mixing.

There are many contextual factors that contribute to code-mixing, such as the conversational partners and the conversation topic. Sociolinguistic factors such as age and gender also have an influence on the way people code-mix. Zenner, Speelman and Geeraerts (2015) conducted a study into Dutch-English code-mixing by Dutch and Flemish contestants in a reality TV show, and found that males switched to English somewhat more than females and younger contestants somewhat more than older contestants. A way to study the influence of the conversation topic on code-mixing, is by examining the semantic fields of the switches. Semantic fields are defined as structured parts of the lexicon, in which words are related in meaning, for example, pronouns, numerals, colour terms and cooking terms (Millar, 2007). Another semantic field is that of technology and computer terms. Because digital devices and modern technology have been around for a relatively short time and are developing at a rapid pace, many words in this semantic field have not existed for a long time either and new words are added constantly. Because English is one of the most used lingua franca, especially in the western world (Cenoz & Jessner, 2000), many of these terms originate from English. As mentioned earlier in this section, when a language does not have a word for a concept (yet), it tends to borrow the word from another language, which does have a word for it. This is why it is likely for many languages to use terms from the semantic field of computer and technology from English.

When words are adopted into a recipient language, they can undergo various types of integration. Integration means that the switch is altered in such a way that its fit into the recipient language is improved. Millar (2007) describes how foreign words are integrated into the recipient language. For spoken language, this can happen phonologically and morphologically. Because every language has its own phonological system, speakers often (knowingly or unknowingly) alter the pronunciation of a word or phrase from another language to fit their own familiar phonological system. A similar thing can happen with the morphology of switches. Every language has its own grammatical rules and inflections. When lexical items such as nouns, verbs and adjectives are adopted into another language, they may be in need of being inflected to fit the grammar of the recipient language, in which case inflectional morphemes of the recipient language are added to switches. Of course, in written language, there can be no phonological integration. However, there can be graphemic integration instead: the spelling of a switch may be altered to fit the orthographic rules of the recipient language, or to imitate phonological integration by spelling the switch in such a way that it reflects how it would be pronounced in the recipient language. Vandekerckhove, Cuvelier and De Decker (2015) found and discussed such graphemic integration of English in Flemish, South African, Kenyan, Nigerian, Ghanaian and Sierra Leonean. They concluded that young people make use of graphemic appropriation and integration to show that they are skilful chatters and texters.

3.3 English in youth language

Because this study focuses on code-mixing by youths, it is important to take into account the way youths speak. Youths are noted for making use of 'youth language' in their everyday speech (Schoonen & Appel, 2005). Nortier (2016) adds that this type of language is also found in written CMC. As a consequence, it is to be expected that some form of youth language is present in the CMC corpus used in the present study.

'Youth language' or 'street language' is a language variety spoken among young people. It differs from the standard language mainly in its constantly changing vocabulary (Schoonen & Appel, 2005). Verheijen (2016) found a clear influence of age on the use of youth language in CMC: teenaged youths, in particular 15- and 16-year-olds, used more non-standard language than the slightly older young adults.

Schoonen and Appel (2005) studied the use of youth language by Dutch secondary school students. The large majority of participants said that they used youth language in informal situations on a regular basis – on the streets with peers, or at school with other students. Youths speak youth language with other youths, but generally do not with adults, such as teachers or parents: they use different registers depending on their conversational partner, called communication accommodation. When asked why they used it, most of them said that they did it 'automatically', without thinking: it is their standard way of speaking to each other. Some admitted using it to distinguish themselves from others, in other words, to help create a personal identity. They also studied what constitutes youth language in the Netherlands and found that it often includes words adopted from other languages. The language most foreign elements came from is Sranan, a language spoken by Surinamese immigrants in the Netherland, but English comes second. This shows that using English is definitely a central part of the language youths use when speaking to each other. While this gives an insight into *why* youths use English as a part of their youth language, it does not clarify *how* they use it. Furthermore, it is not straightforward if English plays a similar role when writing/typing to each other.

A way in which English is used in Dutch youth language is literally adopting English slang words into their speech, such as *chill*, *dope* and *the bomb* (Braak, 2002). Cornips (2004) adds that it is a hallmark of youth language that English lexical items, such as verbs and nouns, are adopted, which may be inflected with Dutch affixes (morphological integration).

In short, youths seem to integrate English words into their youth language to convey coolness, to express their identity and to boost their expressivity. They do this by implementing English slang and adopting other English elements into their speech. This thesis will reveal whether that also goes for online youth language, i.e. CMC.

3.4 Computer-mediated communication

As mentioned above, CMC is a way of communicating that has emerged only in recent decades. Many people communicate via the internet and people are having full-on conversations, similar to face-to-face conversations, through instant messaging providers (Herring, 2010). Because of a lack of prosody and body language, people use other ways to express intonation and emotion in CMC, for example through unconventional spelling and emoticons (Georgakopoulou, 2011). CMC occurs in a variety of ways. Video and audio chats are also types of CMC, but the present study only focuses on written CMC. But even written

CMC comes in various types. A few examples are blog posts, emails and instant messaging. There are clear differences between these types. In some types, such as instant messaging, conversations typically take place in real time, where the speakers usually reply as fast as possible after receiving a message, resulting in a (quasi-) synchronous conversation between two or more speakers, similar to a face-to-face conversation. This is not the case for emails, where instant replies are not as common. There are also differences in the type of language people use. When having an informal conversation on an instant messaging service, people tend to use casual language, rather similar to spontaneous speech (Herring, 2010). Though conversations via instant messaging services are much more direct and spontaneous than, for example, written letters or emails, the digital medium makes them not entirely as direct and spontaneous as spoken conversations: they can be positioned somewhere in between spoken and written language (Georgakopoulou, 2011). This is why CMC is referred to as semi-spontaneous.

The four CMC modes that are analysed in this thesis are MSN chat, SMS, Twitter and WhatsApp. All four CMC modes are generally used in a casual, informal way, which makes them appropriate to compare to each other when it comes to language use. Still, these modes each have their own characteristics and constraints, which contribute to the way people use language when communicating via these media. These characteristics and constraints, as analysed by Verheijen (2016), are displayed in Table 1.

CMC MODES				
Characteristics	**MSN chat**	**SMS**	**Twitter**	**WhatsApp**
Message size limit	No	Yes (max. 160 characters)	Yes (max. 140 characters)	No
Synchronicity of communication	Synchronous (real time)	Asynchronous (deferred time)	Asynchronous (deferred time)	Synchronous (real time)
Visibility	Private	Private	Public, sometimes private	Private
Level of interactivity	One-to-one, sometimes many-to-many	One-to-one, sometimes one-to-many	Mostly one-to-many, sometimes one-to-one	One-to-one, sometimes many-to-many
Technology	Computer	Mobile phone	Computer or mobile phone	Computer or mobile phone
Channel of communication	Multimodal	Textual	Multimodal	Multimodal

Table 1. CMC modes characteristics from Verheijen (2016).

As can be seen in the Table above, the four CMC modes have quite different aspects. What does this mean for the way people write in these social media? Well, for example, let us take a look at the synchronicity of communication. MSN chat and WhatsApp are synchronous, whereas SMS and Twitter are not. When interaction via CMC is synchronous, it is more similar to face-to-face conversations, meaning more conversational terms such as interjections should appear. Interjections are one of the categories which seem to be quite borrowable, meaning that synchronous conversations may contain more code-mixing than asynchronous ones. Likewise, with the level of interactivity – if it is one-to-one, it resembles a real conversation more than if it is one-to-many (such as many posts on Twitter), resulting in more informal communication, again, using words such as interjections.

Many people have a negative attitude towards the way youths write in CMC. Because of the unconventional spelling and grammar they often use on social media, people are worried that

CMC influences youths' literacy skills negatively. However, it has not uncontestedly been proven that this is indeed the case (Verheijen, 2015).

3.5 Code-mixing with English in CMC

Code-mixing in CMC has not been studied extensively yet. The use of English by non-native speakers is an up-and-coming topic. There are studies into, for example, Finnish (Kytölä, 2013), Malaysian (Hassan & Hashim, 2009), Chinese (Bi, 2011), and, last but not least, Flemish. De Decker and Vandekerckhove (2012) conducted a study into a topic similar to that of the current study, analysing the use of English in Flemish youths' CMC, also described by De Decker and Vandekerckhove (2013). They looked at qualitative and quantitative factors, including intentionality, length, word categories and integration of the English switches. One of their main conclusions was that while Flemish youths commonly use English in their everyday online conversations, they do not display an elaborate eloquence in it, based on the findings that most English switches consisted of single words and multiple-word switches were usually of an idiomatic nature. They also found that the youths do not just simply copy and paste English words, but also integrate and adapt them into Flemish. Many different aspects of code-mixing are analysed in this thesis, some adopted from De Decker and Vandekerckhove's (2012) study and some added aspects, both language-internal and -external. I also analysed the interaction between several factors, some of which were analysed by De Decker and Vandekerckhove (2012) as well, but most were not.

First, the factors that have been adopted from De Decker and Vandekerckhove (2012) are described here, including their findings and how I apply them in the current study.

- *Length of switches*: De Decker and Vandekerckhove (2012) found that most switches fell in the single-word switch class, much fewer were textism switches and even fewer multiple-word switches. Where they distinguished single-word switches, multiple-word switches and textism switches, I distinguish single-word switches, phrasal switches, full sentence switches, partial-word switches and textism switches, thus making a more elaborate classification. Single-word switches are separate English lexemes, either embedded in a Dutch sentence or standing alone. Phrasal switches consist of more than one English word, but are not full sentences. Sentence switches are full sentences in English, even though the rest of the conversation is in Dutch. Partial-word switches are words that are partly English, partly Dutch, for example in compounds. Textism switches are English textisms: abbreviations used in CMC (e.g. *lol, omg, btw*).

- *Word categories of switches*: De Decker and Vandekerckhove (2012) distinguished the categories nouns, verbs, adjectives (adjectives used as adverbs), interjections and function words. They found that the majority of the switches were nouns; verbs were the second most frequent word category. The categories that are distinguished in the current study are nouns, verbs, adjectives, adverbs, interjections and an 'other' (miscellaneous) category, which includes all other single-word and partial-word switches, such as function words.

- *Integrating and appropriating switches*: Non-integrated switches are reproductions of the original English words. For some (but not all) switches, such as nouns, verbs and adjectives, it is possible to integrate them in various ways. In integrated switches, the switch has been adapted to the Dutch language in some way. De Decker and Vandekerckhove (2012) distinguished three types of integration: graphemic, morphological and semantic integration. I distinguish two ways in which switches can be integrated in CMC. First, graphemic

integration: changing the spelling of a word to reflect Dutch or English pronunciation. Second, morphological integration: the word form is adapted by adding a Dutch affix to the English switch, or by creating a compound of a Dutch and an English lexical item. The current study adopts the two aforementioned types of integration, but excludes semantic integration, because their definition of semantic integration proved highly problematic to implement objectively. Their definition was as follows (De Decker & Vandekerckhove, 2012: 334): "the English lexeme has received a meaning which seems to be unknown to native adolescent speakers of English". While the existence of this phenomenon cannot be denied, it is difficult for non-native speakers to judge which exact meanings a word has in current English – even when resorting to dictionaries, since language changes constantly, especially among youths. There might be many novel nuances which non-native speakers of English do not know about yet. Also, it is difficult to determine how exactly the meaning of a word is intended in such a limited context.

- *Semantic fields:* While De Decker and Vandekerckhove (2012) did not analyse the semantic fields of the switches extensively, they did analyse the semantic fields of some of the most frequent switches (see the 'frequency of switches' paragraph below). The present study extends the analysis on semantic fields somewhat and will divide the 100 most frequent English elements into semantic fields, to explore whether words from certain semantic fields are more likely to be adopted from English than from other fields.

- *Intentionality of switches*: to analyse the necessity of English elements in Dutch youths' CMC, a distinction is made between unintentional switches versus intentional switches. Unintentional switches are necessary switches, which do not have an equivalent (a word with the same meaning) in Dutch. Intentional switches are luxury switches, which do have an equivalent in Dutch. De Decker and Vandekerckhove (2012) found that the large majority of the switches in their corpus were intentional. They also found that the intentionality of switches interacts with the length of switches: multiple-word switches were relatively less often intentional. The current study also analyses the frequencies of intentionality of switches, as well as its interaction with length of switches, word category, and dictionary status.

- *Frequency of switches*: De Decker and Vandekerckhove (2012) made three separate lists for frequency of switches: for the most frequent intentional single-word switches, for the most frequent unintentional switches, and for the most frequent textism switches. They found an interaction between the intentionalities of the most frequent switches and their semantic fields: the most frequent intentional switches were mainly part of the field of computer and technology, whereas the most frequent unintentional switches were not. To see which English words and phrases are most popular with Dutch youths in CMC, I chart the most frequent switches and make a separate list of the most frequent English textisms, just like De Decker and Vandekerckhove. Additionally, I analyse its interaction with the dictionary status.

In their analysis of the influence of English on Flemish youths' CMC, De Decker and Vandekerckhove (2012) purely focused on the linguistic properties of the switches. In addition to replicating this for Dutch youths, I include the following language-external factors:

- *Gender*: is there a difference between the way males and females code-mix with English in CMC? De Decker and Vandekerckhove (2012) did not study this difference, because practically all their data were from male contributors. It is notable, though, that they mentioned finding switches falling into the semantic field of video games. Because such

games are played more by males than by females (Desai et al., 2010), especially in adolescent age (Griffiths, Davies & Chappell, 2004), this may cause a difference between the semantic fields of male and female contributors.

- *Age*: is there a difference between the way adolescents (between the ages of 12 and 17) and young adults (between 18 and 23 years old) code-mix with English in CMC?

- *CMC mode*: is there a difference in the way code-mixing with English occurs in tweets, WhatsApp messages, MSN messages and text messages?

I also study the interaction between these language-external factors and some of the language-internal factors listed here. Lastly, I add one more factor which De Decker and Vandekerckhove (2012) did not study extensively, but did mention:

- *Multiplicity of switches:* De Decker and Vandekerckhove (2012) mentioned that they found switches to English which appeared in sentences with other switches to English, but did not systematically study this factor. To establish if switches often are accompanied by other switches or mostly occur on their own, I study whether the items contain one or more switch to English (item meaning one MSN chat message, one SMS text message, one tweet or one WhatsApp message).

In the next chapter, I will introduce the research questions and their matching hypotheses, based on the factors I have just described.

4 <u>Hypotheses</u>

The theoretical framework has made clear that many factors play a role in code-mixing, youth language and computer-mediated communication. This is why I have split my research question into eight sub-questions. This chapter introduces them and explains the corresponding hypotheses. The hypotheses deal with the differences between categories and main effects of factors by themselves, and also with interactions between multiple factors. The research questions and hypotheses have been placed in a specific order, making sure that factors are always introduced, before going into their interaction with other factors. With these sub-questions, the most relevant aspects of code-mixing are analysed. Also, anything else notable or worth discussing that was found, is discussed.

4.1 Length: What can be said about the length of the English elements?

Based on the findings of De Decker and Vandekerckhove's (2012) and Zenner and Geeraerts (2015), I formulate the following hypothesis: the large majority are single-word switches. Some of the switches consisting of more than one word are (semi-)fixed expressions.

4.2 Multiplicity: Do most sentences contain just one or more English element(s)?

Based on the fact that Dutch youths are generally not balanced bilinguals and their English proficiency is rarely ever as high as their Dutch proficiency (Van Onna & Jansen, 2006), I formulate the following hypothesis: the large majority of the switches to English are the only switch in the CMC item they in which appear, and only few appear in a CMC item with one or more other switches.

4.3 Word category: To which lexical categories do the English elements belong?

This research question only applies to single-word switches and partial-word switchces, as textisms, phrases and sentences as a whole do not have lexical categories. Based on the borrowability theory, the borrowability theory considering interjections by Muysken (1999) and De Decker and Vandekerckhove's (2012) findings, I formulate the following hypothesis: over half of switches are lexical words (of the categories noun, verb, adjective and adverb) and most of these are nouns. Another category that occurs frequently are interjections, with a larger relative frequency than adjectives and adverbs. Very few switches are functional words. The most frequently used individual switches (apart from the total relative frequency of all adjectives and adverbs) are mainly adjectives and adverbs.

4.4 Integration: How are the English elements integrated into the Dutch language, on a graphemic and morphological level?

Based on De Decker and Vandekerckhove's (2012) findings, I formulate the following hypothesis: most switches are not integrated into the Dutch language in any way. A minority of the switches is integrated in one or, even more rarely, two ways. The switches that *are* integrated, can be integrated as follows: morphologically, by compounding and grammatical inflection, and graphemically, by altering the spelling so it matches the Dutch pronunciation and orthographic rules more.

4.5 Semantic fields: In which semantic fields are English elements the most frequent?

Based on the findings about the most frequent unintentional switches by De Decker and Vandekerckhove (2012), and the research on youth language by Braak (2002), I formulate the following hypothesis: the semantic fields of computer and technology and 'teenage talk' have a larger number of switches than other semantic fields.

4.6 Intentionality: To what extent are the English elements mostly included because of lexical need, and to what extent are they luxury switches?

Based on De Decker and Vandekerckhove's (2012) findings and the borrowability theories by Muysken (1999) and Millar (2007), I formulate the following hypothesis: the large majority of the switches are intentional. The intentionality of the switches interacts with a number of other factors. First, length: single-word switches and partial-word switches have a higher percentage of unintentional switches than textism, phrasal and sentence switches. Second, word category: nouns and verbs have a higher percentage of unintentional switches than adjectives, adverbs and interjections. Third, dictionary status: the majority of unintentional switches are included in the Dutch dictionary, whereas the majority of intentional switches are not.

4.7 Frequency of switches: Which English lexemes and textisms are used most frequently?

Based on De Decker and Vandekerckhove's (2012) findings, I formulate the following hypothesis: there is an interaction between the frequency of switches, the intentionality of switches and the semantic fields to which they belong: the most frequent intentional switches mainly fall in the category of 'teenage talk' and the most frequent unintentional switches are mostly part of the semantic field of computers, internet and technology. There is also a correlation between the frequency of switches and their dictionary status. Many of the most frequent switches are included in the Dutch dictionary. This based on the fact that additions to standard language dictionaries nowadays largely depend on frequency counts.

4.8 Language-external factors: What is the influence of the language-external factors CMC mode, gender and age on the insertion of English elements?

4.8.1 CMC mode

Based on the differences in synchronicity and interactivity between the various CMC modes as described by Verheijen (2016), I formulate the following hypothesis: WhatsApp and MSN chat have relatively the most switches to English. There will also be an interaction between CMC mode and word category: Twitter will have a lower percentage of English interjections than MSN chat, SMS and WhatsApp.

4.8.2 Gender

Based on the findings on code-mixing by Zenner, Speelman and Geeraerts (2015) and the findings on gaming and gender by Desai et al. (2010) and Griffiths, Davies and Chappell (2004), I formulate the following hypothesis: male contributers switch relatively more to English than female contributors. Also, there is an interaction between gender and the semantic fields of the switches: the male youths use more terms inside the semantic field of video games than female youths.

4.8.3 Age group

Based on the findings by Verheijen (2016), I formulate the following hypothesis: age group interacts with a number of other factors. The younger age group is less conforming to the standard language in their written CMC than the older age group. Accordingly, differences may crop up between the two age groups in a number of factors. First of all, an interaction with length: the younger age group uses more textism switches than the older age group, as textisms represent non-standard orthography typical of CMC. Second, word category: the younger age group uses relatively more English interjections and fewer English nouns than the older age group, thus using relatively more 'teenage talk' in English than the older age group. Third, graphemic integration: the younger age group integrates relatively more English elements than the older age group, thus diverging more from the English spelling. Fourth, intentionality: the younger age group uses more intentional switches than the older age group, consciously deviating from Standard Dutch. And last, dictionary status: the younger age group uses fewer English words that have been added to the Dutch dictionary and are thus part of Standard Dutch.

5 __Methodology__

5.1 Materials

The corpus used for this thesis contains various types of written CMC. It has been collected by Lieke Verheijen, who extracted the MSN chat, SMS and Twitter materials from the SoNaR corpus (Treurniet et al., 2012, Treurniet & Sanders, 2012; Oostdijk et al., 2013) and collected the WhatsApp chats herself. The current form of the corpus is a collection of Microsoft Excel files, with one text, WhatsApp, MSN message or tweet per row. The specifications are displayed in Table 2.

OVERVIEW OF CORPUS					
CMC mode	**Year**	**Age group**	**Mean age**	**# of words[2]**	**# of conversations /contributors[3]**
MSN chat	2009-2010	12-17	16.2	45,051	106
		18-23	19.5	4,056	21
		Total		49,107	127
SMS	2011	12-17	15.4	1,009	7
		18-23	20.4	23,790	42
		Total		24,799	49
Twitter	2011	12-17	15.9	2,968	25
		18-23	20.6	99,296	83
		Total		122,264	108
WhatsApp	2015-2016	12-17	14.4	55,865	11
		18-23	20.1	140,134	23
		Total		195,999	34
		Grand total		392,169	

Table 2. Corpus overview.

5.2 Procedure

5.2.1 Data collecting

A first step in collecting the data was to determine which words are counted as switches and which are not. For this we used the online version of the Van Dale's Great Dictionary of the Dutch Language, a recognized authority among the Dutch lexicons. If a word was not included in this dictionary (but it was, of course, in English dictionaries), it was counted as a switch. If it was in the Dutch dictionary, but with an indication that the word has recently been borrowed from English, it was also counted as a switch. Other words which might have been borrowed from English at some point, but did not have this indication, were not counted, because they were likely borrowed such a long time ago that they have been completely integrated into the Dutch language, to such an extent that they are not recognised as English elements anymore. Also, proper names, such as titles of films, books or video games, were not counted as switches either. These criteria provide an objective, systematic judgment about whether an element is a switch to English or not. Whether or not the speakers of these switches consider it code-mixing with English is unclear and irrelevant for the purposes of this research, and thus it is not taken into account.

---

[2] The WhatsApp part of the corpus was so large that not every conversation was used in this research. The limit was maximally 10,000 words per contributor, in order not to skew the data due to an overrepresentation of certain contributors.

[3] Conversations for MSN, contributors for SMS, tweets and WhatsApp.

We were unaware of any way to search the corpus for switches to English automatically, so it had to be done manually. For previous research (Verheijen, 2016) part of the switches had already been found and coded for length (single-word switch, phrasal switch, sentence switch or textism switch). The words that were still left to tag were the words that are in the Dutch Van Dale dictionary. A preliminary list of these words (that had been found in the corpus but not included as switches) was provided and via the Ctrl+F search option, they were systematically sought, included as switches and tagged for length. Though this way of searching the corpus is fast and convenient, it unfortunately does not find every misspelling and graphemic variant of the words searched. Because of practical reasons, it was beyond the scale of this master's thesis to go through the entire corpus and search for every single instance of code-switches, so it has to be taken in account that a few tokens might be missing. To be as complete as realistically possible within the given time frame of this thesis, searches for frequent and predictable spelling variants of the code-switches were added. Also, when variants and other switches were encountered by coincidence, they were included as switches and separately searched for as well.

Image 1. The corpus as displayed in Excel.

As can be seen in image 1, some of the utterances are red. Red sentences indicate utterances that were not written by Dutch youths to speakers of Dutch; for example, automatically generated tweets, retweets (tweets from other Twitter accounts reposted on one's Twitter profile), or messages to non-Dutch speakers. These were not counted as switches and left out of this study.

No way	No way	phrasal switch
Die namen		
Ik ga mee		
7 feb! Om 11 uur		
Nope idd	Nope	single-word switch
Haha		
Whahaha		
Hahaha		
Ik kan me er niets meer van herinneren, eigenlijk		
Is er wat in mijn drankje gedaan		
Hahaha		
Nice	Nice	single-word switch
Met je memes	memes	single-word switch

Image 2. Close-up of the corpus.

Image 2 shows a close-up of part of the corpus in Microsoft Excel, with the full CMC items in the left column, the found switches in the middle column and the length of the switches in the right column.

After the switches had been found, the entire utterances, switches and length were manually copied and pasted into one Excel file and CMC mode, age group and contributor code were added, so the file could be exported to Microsoft Access for data coding. If a sentence contained multiple switches, the sentence would be pasted into the Excel file multiple times, once for each switch.

Image 3. The filtered data in Excel, ready to be exported to Access.

After the file had been exported to Access, the gender information for the WhatsApp data was added and the switches were given an ID. Then the data coding could get commence.

5.2.2 Data coding

The Microsoft Access file, created by Lieke Verheijen, initially consisted of a table and a corresponding form, both of which could be used to code the data. The table consisted of 17 columns, which each contained a piece of relevant information. A description of each of these columns is given below.

ID:	Every CMC item was given a unique ID, numbered from 1-(n).
CMC item:	In this column, the entire CMC item, in which the switch occurred, is displayed.
English borrowing:	This column displays the English element from the CMC item. There is one English element per row, so when there are multiple switches in one item, the item is displayed multiple times.
Lemma:	For every switch, the lemma is displayed here, to make sure every token of the same lemma would be counted as the same word. Lemmas were distinguished by word category, and the English spelling was used.[4] Phrasal and sentence switches were not split into multiple lemmas, but written down as one lemma. It was unnecessary to make a distinction between (words that started with or were fully in) uppercase letters and (words that were all in) lowercase letters, as Access also did not make this distinction.
CMC mode:	MSN / SMS / Twitter / WhatsApp
Age group:	12-17 / 18-23
Gender:	Male / Female[5]
Contributor code:	For each contributor (or conversation in the case of the MSN items), there is a unique code, which is displayed in this column.
Dictionary status:	Yes / No. Dictionary status is the only factor that does not have its own research question, but it is included in the hypotheses for the research questions about intentionality, frequency and age group.
Multiplicity:	One switch per item / Multiple switches per item
Intentionality:	Intentional / Unintentional
Length:	Single-word switch / Phrasal switch[6] / Sentence switch / Partial-word switch / Textism switch. When phrasal and sentence switches contained textism switches as a part of the phrase or sentence, this textism was separately added as a textism switch as well.
Word category:	The single-word switches and partial-word switches were divided into these word categories: Noun / Verb / Adjective / Adverb / Interjection / Pronoun / Other. Phrasal, sentence and textism switches did not get coded in this category.
Integration:	Integrated / Non-integrated. If non-integrated, the next two columns are to be left empty.
- Graphemic integration:	Yes / No
- Morphological integration:	Yes / No
Notes:	If there was anything else to note about the switch, there was a place for it in this column.

Most factors studied in this thesis have their own column in the Excel file, with the exception of semantic fields and frequency of switches. This is because these two factors cannot simply be quantified in such a column, so they have been analysed afterwards.

---

[4] For the lemmas, the English spelling of the words was used, except with verbs, where the Dutch infinitive was used to avoid confusion between verbs and nouns.

[5] Because the gender of the contributors of MSN chats, SMS messages and tweets are unknown, only the WhatsApp data were used for this part.

[6] In our original coding scheme, partial-word switches were not a separate category, but because of the clear distinctions that we came across between these and single-word switches, this category was added later.

Image 4. The data as displayed in the form in Access.

Image 5. The data as displayed in the table in Access.

Here is an example of how a switch to English was coded.

> *heb ik nog steeds niet gedownload:D*
> 'i still haven't downloaded [it] :D'

The unique ID of this CMC item ID is 410; it came from the MSN chat ages 12-17 data, from the conversations with code 1099. The columns were filled in as showed below. Elaboration is included where needed here, but was not included in the Access file.

ID: 410
CMC item: heb ik nog steeds niet gedownload:D
English switch: gedownload

Lemma:	downloaden → *gedownload* is an inflected form of the verb 'to download', so the Dutch infinitive ('downloaden') was entered in the cell
CMC mode:	MSN chat
Age group:	12-17
Gender:	(left empty) → gender data for the MSN chats was unknown
Contributor code:	1099
Dictionary status:	Yes → the verb *downloaden* is part of the online version of the Dutch Van Dale dictionary
Intentionality:	Unintentional → there is no Dutch equivalent for this verb
Length:	Single-word switch
Word category:	Verb
Multiplicity:	One switch per item
Integration:	Integrated
Graphemic integration:	No
Morphological integration:	Yes → the verb has been integrated morphologically by grammatical inflection
Notes:	(left empty)

This gives an insight into how the items were coded. When confronted with issues or ambiguities, these were fixed systematically as much as possible. For example, switches such as *high tea* or *skinny jeans* are made up of two words with a space in between, even though they are fixed combinations used to refer to a single object and can be regarded as compound words. When such switches were present in the English dictionary as a single term, they were tagged as a single-word switch. If not, they were tagged as a phrasal switch. Some unique cases required personal attention, such as the switch *wtf'en* ('to wtf'). As a textism used and inflected as a verb, it could have been tagged either textism switch or single-word switch. Because this was the only one in its kind, it has been tagged as a textism switch. For more elaboration on this particularly interesting switch, see section 6.4.2.

5.2.3 Data analysis

After the tagging of the data, they were ready to be analysed. Based on the hypotheses, queries for tables and cross tables were made in Access to automatically calculate the absolute frequencies of the different categories. Then, the relative frequencies were manually calculated and put into tables. Where appropriate, the data were entered into IBM SPSS Statistics 20, to perform a chi-square test in order to test the significance of the results.

Age group ▾	Total Of ID ▾	no ▾	yes ▾
12-17	3209	1889	1320
18-23	5410	2649	2761

Image 6. An example of a cross table created by Access (age group x dictionary status).

Image 6 shows the cross table that was the result of one of the queries, in this case the interaction between the factors age group and dictionary status.

	agegroup	wordcategory	frequency
1	1,00	1,00	669,00
2	1,00	2,00	312,00
3	1,00	3,00	406,00
4	1,00	4,00	304,00
5	1,00	5,00	875,00
6	1,00	6,00	17,00
7	2,00	1,00	1526,00
8	2,00	2,00	538,00
9	2,00	3,00	609,00
10	2,00	4,00	317,00
11	2,00	5,00	656,00
12	2,00	6,00	34,00

Image 7. The data view in SPSS.

Image 7 shows how the data was entered into SPSS. Next, a cross table was made by weighting cases by frequency, then choosing the option 'crosstabs' in the analyse drop-down menu, selecting one of the variables for the columns and the other one for the rows, and selecting to also calculate percentages and a chi-square test. This resulted in an output such as the one in image 8 below.

Crosstabs

[DataSet0]

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
agegroup * wordcategory	6263	100,0%	0	0,0%	6263	100,0%

agegroup * wordcategory Crosstabulation

			wordcategory						Total
			1,00	2,00	3,00	4,00	5,00	6,00	
agegroup	1,00	Count	669	312	406	304	875	17	2583
		% within agegroup	25,9%	12,1%	15,7%	11,8%	33,9%	0,7%	100,0%
	2,00	Count	1526	538	609	317	656	34	3680
		% within agegroup	41,5%	14,6%	16,5%	8,6%	17,8%	0,9%	100,0%
Total		Count	2195	850	1015	621	1531	51	6263
		% within agegroup	35,0%	13,6%	16,2%	9,9%	24,4%	0,8%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	289,285[a]	5	,000
Likelihood Ratio	289,703	5	,000
Linear-by-Linear Association	267,059	1	,000
N of Valid Cases	6263		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 21,03.

Image 8. The output view in SPSS.

The second table in image 8 is the cross table, including percentages. The third table shows the chi-square tests.

The discussion of the crosstabs and chi-square tests provide the basis of a quantitative analysis of the data. Other interesting cases that were found in the data were noted to provide a qualitative analysis.

For the research questions on semantic fields (questions 5 and 8), word clouds were made instead of Tables on the online word cloud generator tagcrowd.com (Steinbock, 2016), to give a quick overview of the most frequent switches. Tables of the top 100s are included in the Appendices. Because this cloud generator counted every word as a single word (including the words inside switches containing multiple words), the top 100s in the word clouds look slightly different from the top 100s in the Tables. However, because the large majority of the most frequent switches are single words (or textisms) anyway, this does not make much of a difference for the objective of the current study.

6 <u>Results and discussion</u>

The number of switches that were found in the CMC corpus is 8,619. The total number of words in the corpus is 392,169 – this results in a percentage of 2.19%. However, not all words are single words – 1091 of the switches consist of multiple words. This means that at least 9710 of the words are English. Because some of these switches consist of three or more words, say the total number of English words is about 10,000. This results in a percentage of 2.55% – this means that about one in forty words in the corpus are (part of) a switch to English.

The results are discussed in sections corresponding to the eight research questions and hypotheses. At the beginning of each section, the hypothesis is repeated, after which the results are presented and discussed. In the examples of CMC items containing one or more switches, the switches are underlined. If the entire sentence is a switch, it is completely underlined.

6.1 Research question 1: length

Hypothesis:
The large majority are single-word switches. Some of the switches consisting of more than one word are (semi-)fixed expressions.

RQ1: LENGTH		
Length of switch	**Absolute frequency (#)**	**Relative frequency (%)**
Single-word switch	6124	71.1
Textism switch	1265	14.7
Phrasal switch	691	8.0
Sentence switch	400	4.6
Partial-word switch	139	1.6

Table 3. Absolute and relative frequencies of lengths.

As expected, the large majority of the switches are single-word switches. This supports the idea that the Dutch-English code-mixing of Dutch youths consists of a matrix language Dutch, with English as the embedded language. This is in line with what De Decker and Vandekerckhove (2012) found, an order of single-word switch > textism switch > multiple-word switch. It is also in line with the findings of Zenner and Geeraerts (2015), who found that around 30% of the English switches in their data consisted of more than one word.

Some examples of the various types of switches are shown below. Because single-word switches and textism switches receive much more attention later on in other sections, they will be named only briefly here.

Single-word switches

Some examples are *nice, hey, cool, swag, account, shit, happy*.

Textism switches

Some examples are *btw* (by the way), *idk* (I don't know), *jk* (just kidding), *lol* (laughing out loud), *ofc* (of course), *omg* (oh my God), *thx/thnx/tnx* (thanks), *wtf* (what the fuck), *2* ('to' or 'too') *4* ('for'), *y* ('why').

Phrasal switches

Some examples: *by the way, common room, hi there, my pleasure, nice pic, the best, the fuck, who cares.*

Most of these phrases seem to be short, simple, and fixed. This is in line with results of Zenner and Geeraerts (2015), who found that more than half of the multiple-word switches in the data they studied had a certain level of fixedness. This also goes for many of the phrases in our data, such as *by the way, my pleasure,* and *who cares,* which are all present and named as fixed expressions in the online version of the Oxford English Dictionary (OED).

Some phrasal switches are not fixed expressions. An example is the phrasal switch '*You pretty girl'.* This shows that not all switches consisting of multiple-words are merely copied from the source language and pasted into the recipient language as one fixed whole.

Sentence switches

Although some of the sentence switches are short sentences such as *have fun* or *I know*, we also encountered longer sentences fully in English, as can be seen in the examples below:

(1)   *I've had better ideas*
(2)   *That's just how i roll*
(3)   *you unbright my day by ignoring me.. again*
(4)   *Me not really enjoys absolute silence*
(5)   *never change a winning team*

Examples 1 and 2 show grammatially correct English sentences. Example 3 also shows a regular English sentence, but its grammar is non-standard. 'Unbright' is not normally used as a verb. This shows that the speaker might not be fluent in English and just says what they assume is the correct English word, or they are purposefully creative with the English language. (4) is even more bluntly grammatically incorrect, but it is clear that was done intentionally here. The grammar mistakes were probably inserted for a comical effect.

Again, fixed expressions are found here as well, as in Zenner and Geeraerts (2015). Examples are (2) and (5).

Partial-word switches

As mentioned before, the category of partial word switches was added later during the coding phase, because some elements turned out not to fit the single-word switch category. De Decker and Vandekerckhove (2012) included them in the morphologically integrated switches as compounds and derivations, where most but not all of the partial-word switches I found could have been included, so I decided to create a separate category for them afterwards.

Some examples of nouns are *groepsapp* 'group app' and *downloadprogramma* 'download program'. Both are compounds of a Dutch and an English noun, one of them having the Dutch word first and the other with the English word first.

Examples of verbs are *doormailen* 'to forward an email' (literally 'to mail through') and *platgewhatsappt* 'bombed with WhatsApp messages' (literally 'flat-WhatsApped'). Both are English verbs with a Dutch prefix attached to it, creating a compound verb.

Some partial-word switch adjectives are *akwardheid* 'awkwardness' and *kkchill*, short for *kankerchill* 'fucking chill' (literally 'cancer chill').

The final two examples are no simple combinations of Dutch and English, but are a little more complicated.

(6)   *Dan ben ik 5 uur <u>ish</u> bij jou*
        'Then I'll be at your place around 5 ish'

In (6), only the morpheme *ish* is English; the rest of the sentence is entirely in Dutch.

(7)   *Consu(<u>fucking</u>)mentengedrag*
        'Con(fucking)sumer behavior'

In (7), the word *fucking* is inserted in a word, a process that occurs in English too, to give it more emphasis. Such intra-word insertion of expletives or profanity does not normally happen in Dutch, which means it is not only an English lexical switch, but also a loan of an English morphological process – expletive infixation. Even the position inside the word, before the syllable that has the main stress, is in accordance with the English custom, which means the speaker knows what they are doing.

These examples show the various ways in which English and Dutch words and morphemes are combined to create new words. Whether they are compounds, English words with Dutch derivation, or Dutch words with English derivation, nothing seems impossible.

This section has introduced the various lengths of switches that were found in the CMC corpus. It is clear that Dutch youths mostly use single English words when code-mixing with English and that switches consisting of multiple words are often (semi-)fixed expressions. This is not always the case, though, as creativity is shown in, for example, non-fixed expressions and partial-word switches.

Hypothesis confirmed: Yes

6.2 Research question 2: multiplicity

Hypothesis:
The large majority of the switches to English are the only switch in the CMC item they in which appear, and only few appear in a CMC item with one or more other switches.

RQ2: MULTIPLICITY OF SWITCH		
# of switches per item	Absolute frequency (#)	Relative frequency (%)
One switch per item	7079	82.1
Multiple switches per item	1540	17.9

Table 4. Absolute and relative frequencies of multiplicity.

As can be seen in Table 4, the majority of the switches appears in a sentence without any other English switches. This is in line with the claim that when code-mixing with English and

Dutch, Dutch youths use Dutch (their native language) as the matrix language and English (the foreign or second language) as the embedded language. However, there is a considerable part which does occur in an utterance with other switches. Let us take a closer look at them.

Some examples of items with more than one switch to English:

(8)   *Oke wil jij mij <u>please</u> de antwoorden van wis door<u>appen</u>?*
        'Okay can you please app me the math answers?'
(9)   *kunje gwoon <u>subtitles</u> <u>downloade</u> en dan via da programma erbij zette*
        'you can just download subtitles and then add them via that program
(10)   <u>*Seed*</u> *niemand die <u>torrent</u> ofzo :P*
        'Is noone seeding that torrent or what :P'
(11)   <u>*Backpack*</u> *weegt 20.06kg. Limiet is 20kg om hem in te <u>checken</u>. Gaan ze lopen <u>bitchen</u> om 60 gram*
        'Backpack weighs 20.06kg. Limit is 20kg to check it in. Then they bitch about 60 grams'
(12)   *Hee ik lees jullie <u>appjes</u> nu pas. Maar ik hoef zelf eigenlijk niet zo nodig mee te doen met de <u>bucketlisten</u>. Maar jullie kunnen hem wel voor elkaar maken.*
        'Hey I'm only reading your apps now. But to be honest, I don't really feel the need to participate with the bucketlists. But you guys can make them for each other.'

(8), (9) and (10) show items consisting of one sentence with multiple English items within that sentence. (11) and (12) show items made up of multiple sentenes, where each sentence contains one switch at most. Even though there is more than one switch in each of these examples, still the items are mostly in Dutch, with only a few English words mixed in. The English words are related to each other in some cases (such as *seed* and *torrent*), but generally they are not from the same semantic field.

De Decker and Vandekerckhove (2012) mentioned that they also found switches that appeared with at least one other switch in the same CMC item, but they did not calculate their relative frequency.

Hypothesis confirmed: Yes

6.3 Research question 3: word category

> Hypothesis:
> Over half of switches are lexical words (of the categories noun, verb, adjective and adverb) and most of these are nouns. Another category that occurs frequently are interjections, with a larger relative frequency than adjectives and adverbs. Very few switches are functional words. The most frequently used individual switches (apart from the total relative frequency of all adjectives and adverbs) are mainly adjectives and adverbs.

Because sentence switches, phrasal switches and textism switches do not have a word category overall and no categories were assigned to words within these switches, only single-word switches and partial-word switches are discussed in this section. The total number of tokens (individual words) of those two categories of switches is 6,263, of 1,162 different types (different words).

RQ3: FREQUENCY OF INDIVIDUAL WORD CATEGORIES				
Word category of switch	Tokens (abs. freq. (#))	Tokens (rel. freq. (%))	Types (abs. freq. (#))	Types (rel. req. (%))
Noun	2195	35.0	672	57.8
Verb	850	13.6	170	14.6
Adjective	1015	16.2	161	13.9
Adverb	621	9.9	52	4.5
Interjection	1531	24.4	87	7.5
Other	51	0.8	20	1.7
Total	6263	100	1162	100

Table 5. Absolute and relative frequencies of the tokens and types of the word categories.

Table 5 shows the frequencies of the word categories. First, consider the frequencies for tokens. As expected, nouns are the most common word category and interjections also score rather high. The category 'other', which includes function words, among other things, has only 0.8%. De Decker and Vandekerckhove (2012) also found the highest percentage for nouns, but they had a higher percentage for verbs than for interjections.

When looking at the frequencies for types, the distribution looks quite different. The percentage of interjections is much lower when looking at the variety of different interjections. This is in line with De Decker and Vandekerckhove (2012), who had a much lower percentage of types than of tokens for interjections. Nouns have by far the largest percentage of types, which was also the case in De Decker and Vandekerckhove (2012). The results for the types are in line with the borrowability hierarchy; it shows that nouns are the most borrowable and that verbs come in second. Although many switches are interjections, the variety is much smaller than that of some of the other word categories. This means that on average, the interjections that have been borrowed have a higher frequency per type. This makes sense, because interjections are easily borrowable for another reason than nouns and verbs: there is no lexical need to use interjections from another language, as can be the case with nouns and verbs, but their unique status in the sentence, makes them easy to be used in another language.

For every word category, I present a number of examples for illustration here.

Nouns:

(13) *Moet je alleen ff een 2e account aanmaken*
      'you just gotta create a 2nd account'
(14) *Ik zit gezellig te praten met mijn twin en vader hahahaha*
      'I'm having a nice chat with my twin and dad hahahaha'

The switches in these two CMC items are *account* and *twin.*

Verbs:

(15) *App wel als we er bijna zijn*
      '[I'll] app when we are almost there'
(16) *Kijk eens wie er als laatst getagged is*
      'Look who has been tagged latest'

The switches in these two CMC items are *app* ('app', first person present tense of 'to app') and *getagged* ('tagged', past participle of 'to tag').

Adjectives:

(17)　*Hij vindt het grappig dat iedereen zo <u>bitchy</u> is*
　　　'He thinks it's funny that everyone is so bitchy'
(18)　*Sowieso <u>sickste</u> festical [sic] ooit.. onbeschrijfbaar*
　　　'Definitely sickest festical ever.. indescribable'

The two switches in these CMC items are *bitchy* and *sickste* ('sickest').

Adverbs:

(19)　*alleen <u>facking</u> moe/duf/lui*
　　　'just fucking tired/dazy/lazy'
(20)　*Maar ik moet nog even <u>casual</u> aan de orde brengen dat ik een vriendin heb*
　　　'But I still have to casually bring up that I have a girlfriend'
(21)　*<u>Yep</u>, aan de andere kant van de binnenrand (ringbanen) van de stad.*
　　　'Yep, on the other side of the inner edge (ring roads) of the city'

The switches in these CMC items are *fucking* (stylized as *facking,* see section 6.4.1 for elaboration), *casual*, and *yep*.

Interjections:

(22)　*<u>Fack</u>. En weer die regen.*
　　　'Fuck. And again the rain.
(23)　*<u>Heey</u> heb je een goede foto van die klassenlijst??*
　　　'Hey do you have a good photo of that class list??'
(24)　*En de vraag van ernst bobbie en de rest wist ze wel... verassend.. <u>NOT</u>. #ihvh*
　　　'And she did know the question about *ernst bobbie en de rest...*[7] surprising.. NOT #ihvh'

In these sentences we find the interjection switches *fack* ('fuck'), *heey*, and *not*.

Other:

(25)　*Dan ben ik 5 uur <u>ish</u> bij jou*
　　　'Then I'll be at your place around 5 ish'
(26)　*Je kan geen dag<u>tickets</u> kopen though*
　　　'You can't buy day tickets though'
(27)　*en dan <u>after</u>:*
　　　'and then after:'

The switches of the miscellaneous category shown here are *ish*, a suffix, *though*, a conjunction, and *after*, a preposition.

Hypothesis confirmed: Yes

---

[7] *Ernst, Bobbie en de rest* is a Dutch children's TV show.

RQ3: LEMMA x WORD CATEGORY							
Lemma	Total Of ID	Noun	Verb	Adjective	Adverb	Interjection	Other
1 hey	661					661	
2 nice	225			221	4		
3 mail	202	202					
4 yup	180				180		
5 thanks	142					142	
6 yep	140				140		
7 cool	134			129	5		
8 mailen	124		124				
9 shit	111	53		3		55	
10 relaxed	107			98	9		
11 downloaden	94		94				
12 chill	93			89	4		
13 nope	85					85	
14 yes	85				56	27	2
15 checken	84		84				
16 hi	84					84	
17 fucking	80			21	58		1
18 site	77	77					
19 okay	75					72	3
20 online	73			45	28		
21 app	68	68					
22 spam	68	68					
23 dude	61	61					
24 laptop	60	60					
25 appen	46		46				
26 aw	46					46	
27 awesome	38			36	2		
28 yay	36					36	
29 fucken	36		36				
30 website	34	34					
31 fail	33	33					
32 chillen	33		33				
33 no	33				33		
34 aight	33					33	
35 date	32	32					
36 yeah	31					31	
37 wifi	29	29					
38 account	26	26					
39 random	25			21	4		
40 live	25			12	13		
41 damn	24					24	
42 show	22	22					
43 true	21			21			
44 cute	20			16	4		

Table 6. Frequencies of the most frequent lemmas.

Table 6 shows the most frequent single-word and partial-word switches. All other items, such as textisms, have been removed. This table includes all lemmas of single-word and partial-word switches that appeared in the corpus at least 20 times.

Of the top 10 most used lemmas:
- 5 have been used as an adverb [*nice, yup, yep, cool, relaxed*], with a total of 338 tokens
- 4 have been used as an adjective [*nice, cool, shit, relaxed, chill*], with a total of 451 tokens
- 3 have been used as an interjection [*hey, thanks, shit*], with a total of 858 tokens

- 2 have been used as a noun [*mail, shit*], with a total of 255 tokens
- 1 has been used as a verb [*mailen*], with a total of 124 tokens
- 0 have been used as another word category

As can be seen, adjectives and adverbs are indeed the two categories in which the most frequent words appear. Interjection also scores pretty high; while they have less variety in types appearing in the top 10, the number of tokens is by far the highest. Interjections are high in the top of word categories (see Table 5), because a few of them are used in code-mixing very frequently. This is also in line with the difference between the percentages of tokens and types that is displayed in Table 5. The majority of the most used lemmas are of categories that have a higher percentage of tokens than of types: adjectives, adverbs and interjections.

Hypothesis confirmed: Yes

6.4 Research question 4: integration

> Hypothesis:
> Most switches are not integrated into the Dutch language in any way. A minority of the switches are integrated in one or, even more rarely, two ways. The switches that *are* integrated, can be integrated as follows: morphologically, by compounding and grammatical inflection, and graphemically, by altering the spelling so it matches the Dutch pronunciation and orthographic rules more.

RQ4: INTEGRATION		
Integration of switch	**Absolute frequency (#)**	**Relative frequency (%)**
Integrated	1,706	19.8
Non-integrated	6,913	80.2

Table 7. Absolute and relative frequencies of integration.

Table 7 shows the grand total of integrated switches, taking graphemic and morpological integration together. As expected, most of the switches appear in Dutch sentences without being adapted at all. Although De Decker and Vandekerckhove (2012) also studied how English words were integrated into Flemish Dutch CMC, they did not calculate the absolute and relative frequencies of integrated and non-integrated switches in their corpus.

Hypothesis confirmed: Yes

6.4.1 Graphemic integration

RQ4: GRAPHEMIC INTEGRATION		
Integrated	**Absolute frequency (#)**	**Relative frequency (%)**
Yes	866	49.2
No	840	50.8

Table 8. Absolute and relative frequencies of graphemic integration, relative to the total of integrated switches.

49.2% of the integrated switches are graphemically integrated, which is 10.0% of the total of 8,619 switches. The graphemically integrated switches seem to mainly be adjectives, adverbs and interjections. Many of these words are used commonly and also commonly integrated in the same way. Some examples of these often graphemically integrated words are:

(28)   *fuck(ing)*   → *fack(ing)* or *fock(ing)*

(29)	*relaxed*	→ *relaxt*
(30)	*yep* or *yup*	→ *jep* or *jup*
(31)	*hey*	→ *ey* or *eey* or *heey*

Because the English pronunciation of the vocal 'u' in *fuck* and *fucking* is not part of the Dutch phonological inventory, it is generally phonetically integrated when used in speech, i.e. replaced by a Dutch vowel. This phonetic integration is graphemically displayed as an 'a' or 'o'. In some cases, the 'c' is completely left out or replaced by another 'k', resulting in forms like *fok(king)*, in which the omission of 'c' does not change the pronunciation but makes the orthography more transparent. (29), (30) and (31) also show words that are altered so they match the (Dutch) pronunciation better. In *relaxt*, the word has been integrated graphemically by changing the '-ed' suffix, pronounced as /t/, to a 't', which reflects not only the Dutch but also the English pronunciation more straightforwardly. Two words that are integrated very much are *yep* and *yup*. They are nearly always spelled with a 'j' instead of with a 'y', which raises the question if youths are even aware that these words came from English. Since they are spelled with a 'j' with such high consistency, one would think that it is considered normal to spell them in a 'Dutch' way – as opposed to the other switches, which are occasionally graphemically integrated. It could very well be possible that these words are already regarded by Dutch youths as Dutch words, despite the fact that they have not been added to the Van Dale dictionary (yet).

Not all graphemically integrated switches are integrated so naturally.

(32)	*have fun*	→ *hef fun*
(33)	*thanks*	→ *fenks*
(34)	*wifi*	→ *waifai*
(35)	*nice*	→ *nais* or *naise*
(36)	*playlist*	→ *pleelist*
(37)	*mail*	→ *meel*

Although the four examples above are integrated in such a way that they match their Dutch pronunciation, these words appear in their non-integrated form much more commonly. This may suggest that the graphemic integration was done here in a sort of joking manner. An interesting example is the form *naise*, in which the vowel and consonant have been changed for integration into Dutch, but the silent 'e' at the end has not been omitted.

De Decker and Vandekerckhove (2012) found the same kinds of graphemic integration, sometimes even the exact same words integrated in the same way, such as *fokking*, *pleelist* and *meel*. Apparently, these kinds of integration come natural to youths both from Flanders and the Netherlands in their Dutch CMC.

6.4.2 Morphological integration

RQ4: MORPHOLOGICAL INTEGRATION		
Integrated	Absolute frequency (#)	Relative frequency (%)
Yes	847	49.6
No	859	50.4

Table 9. Absolute and relative frequencies of morphologic integration, relative to the total of integrated switches.

45.5% of the integrated switches are morphologically integrated, which is 9.8% of the total of 8,619 switches. Phrasal switches and sentence switches were never morphologically

integrated, with one exception; there was one phrasal switch which contained morphological integration, which is discussed later on in this section. The large majority of the morphologically integrated switches are single-word switches, though there are a few partial-word switches as well. There is just one textism switch that has been morphologically integrated: *wtf'en* ('to wtf').

The morphological integration of switches occurs with verbs, nouns and adjectives. (38)-(40) show examples of integrated switches in these three categories:

(38) *to check* → *checken* (infinitive) or *gecheckt* (past participle)
(39) *app* → *appjes* (diminutive + plural)
(40) *cool* → *coole* (inflected)

Verbs

The inflection of verbs happens in a variety of ways, using a variety of spellings, some of which adhere to the Standard Dutch orthographic rules, but not all.

Past participles:

(41) *Ik heb dorus al <u>geappt</u>*
      'I have already apped dorus'

The switch in this sentence is the verb *geappt*, the past participle of the verb *to app*, in the meaning of sending someone a message via WhatsApp. It has been integrated into the Dutch language by using the Dutch past participle inflection, in this case with prefix 'ge-' and suffix '-t'.

(42) *Ach, het gaat vanzelf. Alle andere programma's installeren en configureren kost meer tijd. Windows was volledig <u>gecrasht</u> :S*
      'Ah, that's no trouble at all. Installing and configuring all the other programs takes more time. Windows had crashed completely :S'
(43) *<u>Gecrashed</u>*
      'Crashed'

The spelling of past participles of English verbs is not always done correctly, that is, correctly according to the Dutch spelling rules on inflecting English (borrowed) verbs. Both examples above are past participles of the English verb *crash*, but in (42) the Dutch suffix '-t' has been added, where in (43) the English suffix '-ed' has been added (besides the prefix 'ge'), creating a mixture beween Dutch and English inflection.

(44) *De moeder van Serena van Gossip Girl heeft hem <u>gedate</u>*
      'The mother of Serena of Gossip Girl dated him'

Another type of 'misspelling' can be seen in (44). Here, no suffix has been added at all, because in Dutch it would not be pronounced – still, it needs to be written in the Standard Dutch spelling.

(45) *Ik heb het boek Society 3.0 als gratis PDF gedownload! #society30*
      http://bit.ly/s30quest
      'I have downloaded the book Society 3.0 as free PDF! #society30
      http://bit.ly/s30quest'
(46) *ja, kan nu niet meer verder klikken, omdat eerst die anderen gedownloaded moeten*
      zijn :(
      'yeah, can't click any further now, because those others have to be downloaded first :('

In the second of these two examples, an extra '-ed', which is the English suffix for regular past participles, is added, for which there is no reason at all in Dutch. The word already ends with a 'd', so the extra '-ed' would not even be pronounced in Dutch.

(47) *jij tweet iets naar mij over domeinnamen en meteen wordt ik gefollowed door 2*
      domein registratie accounts.. ik return de favor!
      'you tweet something to me about domain names and immediately I am followed by 2
      domain registration accounts.. I return the favor!'
(48) *Ik heb ze gemailed.*
      'I have mailed them.'

Again, the English suffix '-ed' is used, in these cases where the Dutch suffix '-d' should have been used.

(49) *Heb je trouwens ook al je foto's geliket op dorus zn account☺*
      'By the way have you also liked all your photos on dorus's account☺
(50) *Oeh rik heeft ook geliked ^^*
      'Ooh rik has liked too ^^'

Using the English suffix '-ed' seems to be very popular with the Dutch youths, but the correct Standard Dutch spelling is the top one (*geliket*).

(51) *Ik heb de vorige ook niet ge-update*
      'I haven't updated the last one either'
(52) *heb jij je twitter app al geüpdate? Vind hem nu al beter werken*
      'have you updated your twitter app yet? Think it works better already'
(53) *khad virusscanner geupdate*
      'i updated virus scanner'

The same verb has been spelled in three different ways here, but none of them have been inflected correctly. A '-t' should be added, as well as a diaeresis (diacritic consisting of two dots) on the clashing vowels (eü) to mark disyllabicity.

Present tense (second/third person):

(54) *dadelijk unfollowt ze me dat wil ik niet hoor*
      'what if she unfollows me I don't want that'
(55) *Are you serious? En verkleed als een soldaat? Ga als Ghost dan, hij ownt.*
      'Are you serious? And dressed like a soldier? Then go as Ghost, he owns.'
(56) *maar je hebt t gelezen, dus je checked?*
      'but you have read it, so you check?'

(57)   *hmm.. dat <u>sucked</u> wel.. is super duur:D*
        'hmm.. that sucks yeah.. is super expensive:D'

(58)   *je moet wel uitkijken wat jij <u>shared</u>...anders halen ze bestanden van je pc af...*
        'you need to watch what you share though...or they'll take files from your pc…'

In the final three examples here, the English suffix '-ed' is used again where there should be a '-t', even though this is not even a suffix for the present tense in English! It has been morphologically integrated but is completely spelled like an English word, though not even an exsting English word.

It is clear that Dutch youths make use of switches to English creatively, using English verbs with Dutch inflection. However, it seems that not everyone knows how these inflections should be spelled according to the standard spelling rules.

Infinitives:

(59)   *Nu kan ik zn foto's <u>stalken</u>*
        'Now I can stalk his photos'
(60)   *K vertel zo alles, eerst <u>fking</u> Duits nog <u>killen</u> XD*
        'I'll tell everything in a minute, gonna kill fking German first XD'
(61)   *Mr ben al de hele middag aan t <u>gamen</u>*
        'But have already been gaming all afternoon'

In these examples, the verbs *to stalk*, *to kill,* and *to game* have been integrated, using the Dutch infinitive suffix '-en'. Because the stem of the word *game* already ends in an 'e', only an 'n' is added, which is in accordance with Dutch orthographic rules.

An interesting case of morphological integration is the following one:

(62)   *Ik zit nog steeds te <u>wtf'en</u>*
        'I am still wtf'ing'

The textism switch *wtf* is being used as a verb here. The infinitive suffix '-en' has been added, creating a verb.

A special case of phrasal/verbal integration is this one:

(63)   *Ik ga niet <u>thumbs uppen</u>*
        'I'm not going to thumbs up'

Here, the phrase *thumbs up* has been used as a verb, as can be seen by the sufix '-en'. It is used to mean 'giving something a thumbs up', which one can do to 'like' a YouTube video or other online content.

Nouns

 (64)   *wel <u>relaxed</u>:) heb je nu geluid op je tv of op de <u>boxen</u> van je pc?*
        'pretty relaxed:) do you have sound on your tv now or on the boxes of your pc?'

(65)   *ik doe wel ff een <u>screenshotje</u>*
        'i'll do a quick screenshot'
(66)   *oke.. ik heb geen <u>mailtje</u> gehad*
        'okay.. I did not get an email'

Because the plural suffix '-s' exists in both English and Dutch, many of the plural forms of nouns have received this suffix, which means we are unable to establish whether the '-s' is the English or Dutch suffix and have thus not coded them as integrated. There is a rare case of pluralisation with the other Dutch plural suffix, '-en', as seen in *boxen,* in example (654). Some other types of integration are also present. For example, the diminutive suffix '-(t)je' can be found a number of times, like in examples (65) en (66).

Adjectives

(67)   <u>*Coole pinguin;)*</u>
        'Cool penguin;)'
 (68)   *z ehebben iets super <u>chills</u> gedaan op de <u>server</u>*
        'the ydid something super chill on the server'
(69)   *Die is wel echt het aller <u>sexyste</u>*
        'That one is really the most sexy

There are a few ways in which adjectives can be inflected in Dutch. First, in most cases, the suffix '-e' is added when an adjective is placed before certain nouns, as in (67). The suffix '-s' is added when the phrase *iets (adverb)* ('something (adverb)') is used. In (68) this happened to the switch *chill*. Other forms of inflected adjectives are comparatives and superlatives. The comparative suffix is the same as in English ('-er'), which makes it impossible to determine if it is integrated or not. But the superlative suffix is a little different: in Dutch it is '-st(e)', which has been added to the switch *sexy* in (69).

(70)   *Jij vindt ooit een super <u>sexiert</u> met een goede muzieksmaak*
        'One day you will find a super sexy guy with good taste in music'

In (70), the adjective *sexy* has been changed to the noun *sexiert* here, which is supposed to mean 'someone sexy'. This is something that happens in Dutch with some words, such as *leuk* ('likable') and *leukerd* ('someone likeable'). However, '-erd' is not a suffix that is just added productively to any adjective. There are some fixed combinations of adjectives and corresponding nouns. Still, this contributor creatively added this morpheme to an English adjective, creating the word *sexiert*.

Partial-word switches

Even though partial-word switches form their own category *because* they are just partially English, they are counted as morphologically integrated switches nevertheless. More information on these switches has already been provided in section 6.1.

Other

One unique case of adjective/noun integration is the following:

(71)   <u>*shittepettit,*</u> *had ik net een goed cadeauidee, any names of zoiets?*
        'shit, I just had a good gift idea, any names or something?'

The word *shit* has been changed to the comical *shittepettit* here. It might be in analogy to the Dutch word *hittepetit* ('lively woman'), but it could also just be a way to lengthen the word, to add emphasis or a comical element.

Again, the same types of integration have also been found in the data analysed by De Decker and Vandekerckhove (2012).

As mentioned at the beginning of this section, there was one case of morphological integration with a phrasal switch.

(72)   *Er is trouwens iets <u>fucking awkwards</u> gebeurt*
        'Something fucking awkward happened by the way'

The phrasal switch here is *fucking awkwards*. It is clearly a phrasal switch, because *fucking* is an intensifier of *awkward*. According to Dutch grammar, an '-s' suffix is required when an adjective is in this position in a sentence. This is what happened.

6.4.3 Double integration: graphemic and morphological

There are only eight switches that have been integrated both graphemically and morphologically, 0.47% of all integrated words and 0.09% of all switches.

(73)   *Kapper heeft me kaulo haar <u>opgefokt</u>?!!!!? Hahaha ach boeie*
        'Hairdresser fucked up my fucking hair?!!!!? Hahaha oh well'

This is an interesting case, because *opgefokt* is also a Dutch word, meaning 'worked up'. Yet from this context it is clear that the English verb 'to fuck up' is what is meant. The word has been integrated to such an extent that it does not look like an English word anymore in the slightest.

Meeltjes: This noun, from the lemma *mail*, has been graphemically integrated by changing the spelling of the nucleus vowels *ai* to *ee*, matching the Dutch pronunciation more. Also, it has been morphologically integrated by adding the diminutive suffix '-the' and the plural suffix '-s'.

Meelen: This verb, from the lemma 'to mail', has been graphemically integrated in the same way as *meeltjes*, but has also been morphologically integrated by adding the suffix '-en'. The fact that it is spelled *meelen* with two e's is interesting, because according to the Dutch spelling rules, the plural of *meel* would be spelled *melen,* with just one 'e'. Perhaps the speaker felt that writing *melen* would hinder the intelligibility of the word.

Tjekken: This verb, from the lemma 'to check', has been doubly integrated graphemically by changing the 'ch' to 'tj' and the 'ck' to 'kk' and morphologically by adding the suffix '-en'.

Relaxter: This adjective, from the lemma 'relaxed', has been integrated graphemically by changing the '-ed' to '-t', as happens more often with the word *relaxed*. Also, it has been morphologically integrated by adding the comparative suffix '-er'.

Settelen: This verb, from the lemma 'to settle', has been integrated graphemically by changing the order of the 'e' and the 'l', to match the pronunciation of the word, and integrated morphologically by adding the Dutch suffix '-en'.

Two of the switches that have been integrated graphemically as well as morphologically are partial-word switches, namely *fakjou* (*fuck* → *fak*) and *universiteit-meel* (*mail* → *meel*).

We found one switch that has been integrated into Dutch, where the integration is neither graphemic nor morphological. Because this is a very rare occurrence, no separate category was created for this type of integration, but it could be called syntactic integration.

(74)   *why go you not*

Every single word in this sentence is English, but the word order is Dutch, with the finite verb occurring before the subject and the negative particle in that position in the question. When translating this sentence to Dutch word for word, one would get 'waarom ga je niet', which is a perfectly grammatical sentence, meaning 'why don't you go'. Other than syntactic integration, this could also be considered the opposite of a loan translation: rather than indiscriminately translating a foreign language (English) into Dutch (see section 6.9.4), the writer indiscriminately translated a Dutch utterance into incorrect English.

This section has made clear that even though most switches to English appear in their original form, Dutch youths integrate a considerable number of switches into the Dutch language, in various, sometimes creative, ways.

Hypothesis confirmed: Yes

6.5 Research question 5: semantic fields

> Hypothesis:
> The semantic fields of computer and technology and 'teenage talk' have a larger number of switches than other semantic fields.

Image 9. Word cloud of the most frequent switches. (Full Table included in Appendix A.)

Image 9 shows a word cloud of the 100 most frequent switches, made on an online word cloud generator on tagcrowd.com. These 100 words have been divided into semantic fields.

Semantic field: computer/technology
Account, app, appen, chat, downloaden, game, laptop, mail, mailen, online, posten, selfie, server, site, skypen, social media, spam, timeline, update, updaten, website, whatsappen, wifi

Semantic field: opinion/affective language
Awesome, awkward, chill, cool, cute, epic, fail, fun, gay, happy, nice, random, relaxed, sexy, sick

Semantic field: video games
Base

Semantic field: men and women
Bitch, boy, bro, chick, dude, girl

Semantic field: swearing
Bullshit, damn, fuck, fucken, fucking, holy shit, shit

Semantic field: conversational
Aight, aw, bye, hey, hi, indeed, night, nope, okay, please, thanks, true, woohoo, yay, yeah, yep, yup

Semantic field: leisure
Chillen, joint, live, meeten, music, party, shoppen, stoned, ticket

Semantic field: personal
Date, life, love, money, single, stalken

Semantic field: school/work
Checken, deadline, inchecken, workshop

Semantic field: textisms
Btw, idk, lol, np, ofc, omg, wtf, thnx, thx

Other
Dread

Three semantic fields that contain a large number of much-used switches are the fields of computer/technology, opinion/affective language and conversational words. Because conversational words do not carry much semantic weight, it is not very relevant to discuss in depth. The two fields that are more interesting are the first ones.

As expected, by far the largest category is that of computer and technology. Many of the words in this list do not have a Dutch equivalent, such as *mail*, *server*, and *online*. Some of the words do have a Dutch equivalent; *website*, for example, can be translated to *webpagina* or *internetpagina*. However, even in Dutch informal speech, these words seem to occur in English more often than in Dutch, making the English word the unmarked form and the Dutch word the marked form.

Many switches are from the semantic field of opinion/affective language. The words in this list all have a Dutch equivalent, with the exception of *sexy*.

Together with the semantic fields of 'men and women', 'swearing' and 'conversational', the 'opinion/affective language' semantic field show a person's way of speaking. Because there are many terms one can use to express an opinion, address someone or swear, one can choose terms to identify oneself and distinguish one from other speakers. In the case of this corpus, it shows what online teenage talk typically looks like. As an added factor, the fact that a considerable number of the words in the top 100 most freqent switches are textisms shows that these are definitely part of teenagers' computer-mediated communication.

Because only the 100 most frequent switches were divided into semantic fields rather than all switches that were found in the CMC corpus, it does not give a complete view on all semantic fields in which code-mixing happened. But since these 100 switches were by far the most frequent, this analysis does give an insight into the semantic fields in which code-mixing by Dutch youths in CMC is most frequent. Only few of the switches are part of semantic fields other than those that belong to 'teenage talk' and the computer/technology field. The fact that most of the most frequent switches fit into these categories, out of all possible semantic fields, shows that the English used in Dutch-English code-mixing are definitely not random of meaning.

De Decker and Vandekerckhove (2012) mention semantic fields more briefly, only discussing the fields of about 20 of the most frequent switches. Nonetheless, they also found many words in the semantic field of computer and technology.

Hypothesis confirmed: Yes

6.6 Research question 6: intentionality

> Hypothesis:
> The large majority of the switches are intentional. The intentionality of the switches interacts with a number of other factors. First, length: single-word switches and partial-word switches have a higher percentage of unintentional switches than textism, phrasal and sentence switches. Second, word category: nouns and verbs have a higher percentage of unintentional switches than adjectives, adverbs and interjections. Third, dictionary status: the majority of unintentional switches are included in the Dutch dictionary, whereas the majority of intentional switches are not.

RQ6: INTENTIONALITY		
Intentionality of switch	**Absolute frequency (#)**	**Relative frequency (%)**
Intentional	6883	79.9
Unintentional	1736	21.1

Table 10. Absolute and relative frequencies of intentionality.

As can be seen, the majority of the switches are intentional, namely roughly 4 out of 5. This is in line with our expectation and with the results of De Decker and Vandekerckhove (2012), who also found mainly intentional switches.

Hypothesis confirmed: Yes

6.6.1 Interaction with length

RQ6: INTENTIONALITY x LENGTH OF SWITCH												
Intentionality of switch	Total of ID		Partial-word switch		Phrasal switch		Sentence switch		Single-word switch		Textism switch	
	#	%	#	%	#	%	#	%	#	%	#	%
Intentional	6883	79.9	66	47.5	609	88.1	376	94.0	4603	75.2	1229	97.2
Unintentional	1736	21.1	73	52.5	82	11.9	24	6.0	1521	24.8	36	2.8

Table 11. Relation between intentionality and length.

Table 11 reveals the interaction between intentionality and length. There is a clear difference between the lengths we distinguished in switches. The relation between intentionality and length was found to be significant (χ^2 (4, 8619) = 488.927, p = .000). With sentence switches, phrasal switches and textism switches, the percentages of intentional switches are 94.7%, 88.3% and 97.2% respectively, while with single-word switches and partial-word switches the percentages of intentional switches are much lower, namely 75.2% and 47.5% respectively.

Single-word and partial-word switches

Some of the unintentional switches are regular English words, for which there is no Dutch equivalent. Some examples are *brownie, online, high tea* (single-word switches), *terugscrollen* ('to scroll back'), and *uitnodigingsmail* ('invitation email') (partial-word switches).

But unintentional switches also include puns. An example is the word *foutfit*, a combination of Dutch *fout* ('wrong', meaning 'unfashionable' in this context) and English *outfit*. It is not possible to translate such a pun into Dutch while preserving the joke.

Textism switches

It is not a surprise that most textism switches are intentional, since they are generally abbreviations for regular words. Some examples of intentional textism switches are *btw, omg, thx, thnx*. Just a tiny minority of the textism switches are counted as unintentional switches. Some examples are *FF* (Follow Friday) and *yolo* (you only live once). While these phrases can be translated into Dutch, such translations would lack the pragmatic value of the textisms. *FF* is used as a hashtag on Twitter, where *#FF* is tweeted along with the usernames of Twitter users the tweeter enjoys following. Because this is an internationally recognised hashtag, a Dutch translation would be icongruent with the global and public nature of Twitter. *Yolo* has become a type of idiom and is even being used as a regular word in sentences.

(75)   *Dat je gewoon <u>yolo</u> met <u>random</u> mensen mee gaat en kletsen enzo*
         'That you just yolo go along with random people and chat and stuff'

Yolo is used as an adverb here, which would receive the suffix '-ly' in an English sentence, but not in Dutch. Still, even without a suffix its meaning and function in the sentence is clear; it is all about hanging out and chatting with people in a lifestyle that reflects the meaning of *yolo,* you only live once. While the phrase *you only live once* can definitely be translated into Dutch, it is not possible to translate it in this specific context, where it is imbedded in the sentence as an adverb.

Phrasal switches

Many of the phrasal switches are short phrases that are easily translatable into Dutch; these are thus intentional. Some examples are *me too*, *as usual*, *no problem* and *fucking hell*. Some phrases do not have a Dutch equivalent, because they are fixed expressions, such as *guilty pleasure* and *like a boss,* the latter being a fixed expression in English that is mainly used online (a so-called internet meme, see section 6.9.5 for elaboration). Not all unintentional phrasal switches are fixed expressions: phrases that contain words that do not have a Dutch equivalent are also counted as unintentional, for example *nerd rage*.

Sentence switches

There are many short, general sentences which can be translated into Dutch, similar to the intentional phrasal switches. Some examples are *have fun, that's life,* and *what happened.* Again, sentences that are fixed expressions were counted as unintentional switches, along with puns and song lyrics.

Roemer has it is a pun on the English fixed expression 'rumor has it' with a Dutch politician, whose last name is 'Roemer'. The sentence switch in this case was a tweet by someone who was discussing a political issue involving the poitician.
Let's get one thing straight I'm not is another pun involving two meanings of the word 'straight' (clear/heterosexual).

There have been a few instances of English song lyrics.

(76)   *Now stop........ Hammer time*
(77)   *THE HILLS ARE ALIIIIIVE*
(78)   *WITH SOUND OF MUUUUSIC*

The three examples above are all parts of song lyrics. These are unintentional switches, because when translated, they are no longer recognisable as the original lyrics. It is clear that the song lyrics are supposed to imitate the singing of the song, as is reflected in the punctuation and spelling: repetition of periods and letters reflects rhythm and emphasis, while capitalisation reflects volume.

My results are very different from those of De Decker and Vandekerckhove (2012); they found many more unintentional multiple-word switches. This can be explained by a methodological difference: they have included proper names (e.g. titles of films), which would be unintentional, and I have excluded these from my counts.

As mentioned in section 6.1, these results are more in line with the findings of Zenner and Geeraerts (2015), who also found that switches containing more than one word are often fixed expressions, song lyrics among them. This supports the hypothesis that multiple-word switches are often copied from the source language and pasted into the recipient language as a whole.

Hypothesis confirmed: Yes

6.6.2 Interaction with word category

RQ6: INTENTIONALITY x WORD CATEGORY														
Intentionality of switch	**Total of ID**		**Noun**		**Verb**		**Adjective**		**Adverb**		**Interjection**		**Other**	
	#	**%**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**
Intentional	4669	74.5	1252	57.0	378	44.5	905	89.2	570	91.8	1523	99.5	41	80.4
Unintentional	1594	25.5	943	43.0	472	55.5	110	10.8	51	8.2	8	0.5	10	19.6

Table 12. Relation between intentionality and word category.

Table 12 shows the interaction between intentionality and word category. The relation between intentionality and word category was found to be significant (χ^2 (5, 6263) = 1473.876, p = .000). Nouns and verbs have relatively more unintentional switches than the other word categories. The former categories often have a specific, lexical meaning, referring to an object, concept or action in the world. It has been discussed in section 3.2 that these types of words are very borrowable, because of their specific meaning, for which there might not be a term in the recipient language. This is why it makes sense that relatively many of these words are unintentional switches.

Some common examples of intentional single-word switches are: *awesome, awkward, checken, chill, cool, dude, fucking, gay, have fun, hey, I know, nice, relaxed, site, thanks, yep, yup*. Many of these words are typical of 'teenage talk'. Especially when used in a non-English sentence, it is considered cool to use such short English words and phrases. Even though they can be easily translated into Dutch, using English words exudes a certain cool vibe. Even a word such as 'gay' is used not only to refer to a homosexual invididual, but also to mean that something is silly.

(79)   *Das kk gay*
         'That's fucking gay'

Some common examples of unintentional single-word switches are *download, downloaden, game, gamen, laptop, mail, mailen, offline, online, selfie, show, stalken, wifi*. At a first glance it is obvious that the nature of these words, almost all nouns and verbs, are different. They have specific concrete meaning and are not necessarily part of teenage talk.

Hypothesis confirmed: Yes

6.6.3 Interaction with dictionary status

RQ6: INTENTIONALITY x DICTIONARY STATUS						
Intentionality of switch	**Total Of ID**		**Not in dictionary**		**In dictionary**	
	Abs. freq. (#)	**Rel. freq. (%)**	**Abs. freq. (#)**	**Rel. freq. (%)**	**Abs. freq. (#)**	**Rel. freq. (%)**
Intentional	6883	100	4302	62.5	2581	37.5
Unintentional	1736	100	236	13.6	1500	86.4

Table 13. Relation between intentionality and dictionary status.

Table 13 shows the relation between intentionality of switch and dictionary status. This relation was found to be significant (χ^2 (1, 8619) = 1330.153, p = .000). Most words that do not have a Dutch equivalent have been added to the dictionary (86.4%). Some examples of these are *account, chatten, downloaden, skypen, stalken,* and *voicemail*.

The number of words that are not in the dictionary, but still do not have a (semantically and pragmatically) adequate Dutch equivalent and were hence borrowed unintentionally is very small, only 13.4%. Some examples of these words are *autocorrect, pre-order, though,* and *wrap.*

The percentage of words that are in the dictionary that are intentional is somewhat higher (37.5%). The reason why these words would be added to the dictionary despite the fact that an adequate Dutch equivalent for them already exists, would be that they are commonly used in Dutch, so they are in the process of being adopted into the Dutch language as a loanword. Some examples of these words are *app* (in the meaning of 'text message'), *bitch, checken, chill, cool, money, party,* and *random.*

By far the largest category of single-word switches are the ones that are intentional and have not been added to the Dutch dictionary (62.5%). Some examples are *awesome, barfen, btw, hurray, nice, thanks,* and *yeah.*

Hypothesis confirmed: Yes

We can conclude that the switches to English in Dutch youths' CMC are largely intentional. Intentionality clearly interacts with some of the other factors relevant to code-mixing: length of switch, word category and dictionary status. This shows that the intentionality of switches is not random: intentional switches are of a different nature than unintentional ones.

6.7 Research question 7: frequency of switches

> Hypothesis:
> The most frequent intentional switches mainly fall in the category of 'teenage talk' and the most frequent unintentional switches are mostly part of the semantic field of computer and technology. There is also a correlation between the frequency of switches and their dictionary status. Many of the most frequent switches are included in the Dutch dictionary.

Before going into the interactions of the frequency of switches with intentionality and semantic fields and with dictionary status, the most frequent lemmas (total) and the most frequent textism switches are displayed in Tables 14 and 15 below.

RQ7: MOST FREQUENT LEMMAS	
Lemma	**Absolute frequency (#)**
hey	661
nice	225
mail	202
lol	184
yup	180
omg	165
thanks	142
yep	140
cool	134
mailen	124
wtf	111
shit	111
relaxed	107
downloaden	94

chill	93
nope	85
yes	85
hi	84
checken	84
fucking	80

Table 14. Most frequent lemmas.

Table 14 shows the top 20 of most frequent switches, excluding the many *lol*'s that were uttered by an outlier dubbed the LOL-girl. This female contributor used the textism *lol* exceedingly more often than the other contributors. She made 373 textism switches with *lol*, making her an extreme outlier. The undisputable 'winner' is *hey*, which may very well have to do with the fact that the Dutch word *hé* is pronounced and the same carries the same meaning as the English *hey*. Notable is that *nice* came in second. It seems that this currently is the fashionable word to use to express that something is cool.

The most frequent lemmas found by De Decker and Vandekerckhove (2012) are very similar to the ones found here. Their most frequent word is *nice*, and *downloaden*, *nope*, *checken*, *lol*, *omg* and *wtf* are also named as switches with high frequencies.

RQ7: MOST FREQUENT TEXTISM SWITCHES		
Lemma	**Full version**	**Absolute frequency (#)**
lol	Laughing out loud	184
omg	Oh my God	165
wtf	What the fuck	111
btw	By the way	72
idk	I don't know	57
thnx	Thanks	42
thx	Thanks	30
k	Okay	21
np	Now playing/no problem	13
ofc	Of course	12
g	Gangster	9
u	You	8
fml	Fuck my life	8
ly	Love you	8
pls	Please	8
tnx	Thanks	7
yolo	You only live once	7
ff	Follow Friday	7
jk	Just kidding	6
ftw	For the win	6
brb	Be right back	6
b-day	Birthday	6
2	To/too	6
nvm	Never mind	6
ty	Thank you	6

Table 15. Most frequent textism switches.

Table 15 shows the top 25 of most frequent English textisms, again excluding the LOL-girl. Even with her excluded, *lol* is still the most frequently used textism switch. While we found quite a large variety of English textisms in our Dutch CMC data, there are just a few that are used very frequently. Most of these have been used fewer than 10 times. The following textisms also appeared in De Decker and Vandekerckhove's (2012) top 10 English textisms:

lol, btw, wtf, brb, omg, ofc. They also counted *aight* as a textism, whereas in this study it was regarded as an abbreviation.

6.7.1 Interaction with intentionality and semantic fields

RQ7: MOST FREQUENT INTENTIONAL LEMMAS	
Lemma	**Intentional, absolute frequency (#)**
hey	661
nice	225
lol	184
yup	180
omg	165
thanks	142
yep	140
cool	134
shit	111
wtf	111
relaxed	107
chill	93
yes	85
nope	85
checken	84
hi	84
fucking	79
site	77
okay	75
btw	72

Table 16. Most frequent intentional lemmas.

The above table shows the top 20 of most frequent *intentional* switches, again excluding the 373 *lol*'s by the LOL-girl. With the exception of *site* and *checken*, the words seem rather general without much specific meaning or semantic weight. Many of them fit into the picture of 'teenage talk': textisms such as *lol* and *omg*, swear words such as *shit* and *fucking* and words describing opinion and affective language, such as *nice* and *relaxed*.

In their top 10 of most frequent intentional switches De Decker and Vandekerckhove (2012) also found *nice, nope,* and *checken*, which shows that these are definitely popular words with Dutch speaking youths both in the Netherlands and Flanders.

Hypothesis confirmed: Yes

RQ7: MOST FREQUENT UNINTENTIONAL LEMMAS	
Lemma	**Unintentional, absolute frequency (#)**
mail	202
mailen	124
downloaden	94
online	73
spam	68
laptop	60
appen	46
fucken	33
wifi	29
account	26
live	25
show	22

inchecken	17
sexy	16
server	15
workshop	15
update	15
game	14
updaten	13
stalken	13

Table 17. Most frequent unintentional lemmas.

The above table shows the top 20 of most frequent *unintentional* switches. The nature of these lemmas is clearly very different from the intentional lemmas. With 10 nouns and 7 verbs, it shows that these words do have specific semantic meaning, many of them being part of the semantic field of computer/technology.

In their top 10 of most frequent unintentional switches De Decker and Vandekerckhove (2012) also found *mail, online, downloaden,* and *account*, all words in the semantic field of computer/technology, specifically internet-related words.[8]

Hypothesis confirmed: Yes

6.7.2 Interaction with dictionary status

RQ7: LEMMA x DICTIONARY STATUS			
Lemma	Total Of ID	No	Yes
hey	661	661	
lol	557		557
nice	225	225	
mail	202		202
yup	180	180	
omg	165		165
thanks	142	142	
yep	140	140	
cool	134		134
mailen	124		124
wtf	111		111
shit	111		111
relaxed	107		107
downloaden	94		94
chill	93		93
nope	85	85	
yes	85		85
checken	84		84
hi	84	84	
fucking	80		80
site	77		77
okay	75	75	
online	73		73
btw	72	72	
app	68		68
spam	68		68

---

[8] It is important to note that my most frequent words cannot be compared exactly to theirs, because De Decker and Vandekerckhove's (2012) criteria for switches differed somewhat from mine.

dude	61	61	
laptop	60		60
idk	57	57	
appen	46		46
aw	46	46	
thnx	42	42	
awesome	38	38	
k	38	38	
I know	36	36	
yay	36	36	
fucken	35	2	33
website	34		34
fail	33		33
chillen	33		33
no	33	33	
aight	33	33	
date	32		32
yeah	31	31	
thx	30	30	
wifi	29		29
account	26		26
random	25		25
live	25		25
damn	24	24	
show	22		22
true	21	21	
cute	20		20

Table 18. Most frequent lemmas and their dictionary statuses.

Table 18 lists every English lemma that was used in code-mixing in our CMC data at least 20 times and shows whether they occur in to the Dutch Van Dale dictionary or not. Of the top 20 lemmas, 13 have been added to the dictionary and 7 have not. Of the entire table (54 lemmas), 31 have been added to the dictionary and 23 have not. So, the majority occurs in the dictionary. Also, some of these English words are very similar to Dutch counterparts that are, of cource, in the dictionary (such as *hey*, *hi, okay* and *aw* which resemble the Dutch *hé*, *hai, oké* and *ah* respectively, in both form and meaning). Some of the lemmas that do not occur in the dictionary are textism switches, which are less likely to be added to a Standard dictionary anyway, because they are generally limited in use to informal online written language. A few words are used very frequently in the CMC corpus, but are not included in the Dutch dictionary: *nice, yup, thanks, yep* and *nope*. Because these words seem to be so frequent in everyday Dutch (youth) language, adding them to the Van Dale dictionary should be taken into consideration, especially if these words end up staying around for a while.

Hypothesis confirmed: Yes

From data on the frequency of switches, we can conclude that they are largely intentional single-word switches, without much of a specific concrete semantic meaning. There are also a few unintentional single-word switches among the most frequent switches, mainly of the semantic field of computer and technology. Only a handful of textism switches seem to be used very frequently by the Dutch youths in their CMC; only eleven of them were used more than 10 times. A clear relation between the frequency of switches and their dictionary status is also present: most of the most frequent switches are already part of the Dutch Van Dale dictionary. It may only be a matter of time until the others are added too.

6.8 Research question 8: language-external factors

6.8.1 CMC mode

> Hypothesis:
> WhatsApp and MSN chat have relatively the most switches to English. There will also be an interaction between CMC mode and word category: Twitter will have a lower percentage of English interjections than MSN chat, SMS and WhatsApp.

RQ8: CMC MODE			
CMC mode	Absolute frequency (#)	# of words (total)	Relative frequency (%)
MSN	1,213	49,107	2.47
SMS	347	24,799	1.40
Twitter	2,398	122,264	1.96
WhatsApp	4,661	195,999	2.38

Table 19. Absolute and relative frequencies of switches per CMC mode.

Table 19 shows the interaction between frequency of switches and CMC mode. The relation between CMC mode and code-mixing was found to be significant (χ^2 (3, 392169) = 151.953, p = .000). MSN chats have the highest relative frequency of switches to English, followed by WhatsApp, Twitter and SMS. It is notable that MSN and WhatsApp, which are very comparable in features of interactivity and synchronity as they are both forms of instant messaging, are very close in the frequency of English switches as well.

Hypothesis confirmed: Yes

6.8.1.1 Interaction with word category

RQ8: WORD CATEGORY x CMC MODE								
Word category of switch	MSN chat		SMS		Twitter		WhatsApp	
	Abs. freq. (#)	Rel. freq. (%)	Abs. freq. (#)	Rel. freq. (%)	Abs. freq. (#)	Rel. freq. (%)	Abs. freq. (#)	Rel. freq. (%)
Noun	277	25.9	87	30.9	813	49.7	1018	31.1
Verb	165	15.4	64	22.7	250	15.3	371	11.3
Adjective	181	16.9	29	10.3	215	13.1	590	18.0
Adverb	156	14.6	13	4.6	105	6.4	347	10.6
Interjection	288	26.9	88	31.2	239	14.6	916	28.0
Other	3	0.3	1	0.4	13	0.8	34	1.0
Total	1070	100	282	100	1635	100	3276	100

Table 20. Interaction between CMC mode and word category.

Table 20 shows the interaction between word category and CMC mode. This relation was found to be significant (χ^2 (15, 6263) = 34.173, p = .000). As can be seen in the table above, the percentage of English interjections is much lower on Twitter compared to the other CMC modes, and the percentage of English adverbs is lower on Twitter and SMS than on MSN and WhatsApp.

This can likely be explained by the conversational nature of Whatsapp and MSN. WhatsApp and MSN chats often start with a greeting:

(80)   *Heey willemijn. Alles goed?*
         'Hey willemijn. Everything all right?'

(81)   *Hey juuntjeeee*
       'Hey juuntjeeee'
(82)   *Hi hoe is ie*
       'Hi how are you'

In synchronous communication, many other types of conversational words are used as well:

(83)   *haha indeed:D*
       'haha indeed:D'
(84)   *Nope zeker nie, maar dat had ik ook niet verwacht*
       'Nope definitely not, but then again I hadn't expected it'
(85)   *Okay dan zien we wel*
       'Okay then we'll see'
(86)   *Thanks. Tot morgen*
       'Thanks. See you tomorrow'
(87)   *jup, heb m doorgestuurd naar thalia :P zij weet iets meer over de GO-pass dan ik :P*
       'yup, forwarded it to thalia :P she knows a bit more about the GO-pass than I :P'

The examples presented here are those of English elements inserted into Dutch utterances, but it also happens very often that only the English element is said, words like *hey, hi,* and *yup* can be stand-alone switches.

SMS text messages often start with a greeting as well, but the content of the posts make clear that the writer does not intend to start a full conversation here. Rather, the SMS is a completed whole, often even with an explicit closure (e.g. kiss, greeting):

(88)   *Hey lieve jij :) je mag ook wel eerder komen als je wilt ;) kusjee*
       'Hey dear you :) you can come earlier if you want ;) kisss'
(89)   *hey ik fiets langs claudia's straat!!!1!*
       'hey I'm biking along claudia's street!!!1!'
(90)   *Hi! Gaat goed hier :) m'n pakje uit China is er, die eerst lekker uitgepakt, nu aan het samenvatten. Mooizo, dat is fijn, sporten is vast ook lekker :) Uiteraard vind ik dat goed, heb er zin in! Lijkt me een broodnodige compensator ja (K) sportze!*
       'Hi! Doing well here :) my package from China is here, unpacked that first, now making a summary. Good, that's great, exercising must also be nice :) Of course that's fine by me, looking forward to it! Seems a much-needed compensator indeed (K) enjoy your exercise!'

On Twitter, the percentage of nouns is much higher than on the other CMC modes. This shows that the language used on Twitter definitely differs from the language used on the other CMC modes. People have fewer conversations on Twitter and they generally do not start by greeting each other, but by a reply to someone else's puretweet (indicated by '@(username)'), after which an exchange of tweets arises. Still, these 'chats' are generally not synchronous, as opposed to those on WhatsApp and MSN.

Here are some examples:

(91)   *Okay! (a)*
       'okay! (a)'

(92)   *thanks voor de tip girl! ;)*
        'thanks for the tip girl! ;)'
(93)   *yep.. dat is nieuw rooster :-(*
        'yep.. that is new time schedule :-('

Some examples follow of puretweets on Twitter, which are not sent to one specific Twitter user, but to all followers of the person sending the tweet. The different nature of these tweets causes different types of switches.

(94)   *En eindelijk weer verder met het updaten van de blog! Check 'em http://bit.ly/fhvCQ8*
        'And finally back to updating the blog again! Check it http://bit.ly/fhvCQ8'
(95)   *Voor de insiders: the Woolshed, Cairns*
        'For the insiders: the Woolshed, Cairns'
(96)   *Gordon is de grootste loser allertijden.*
        'Gordon is the biggest loser ever'
(97)   *Mensen die denken dat de hele #starwars serie maar tien jaar oud is, vind ik grappig. Zoek de release van deel IV, V en VI anders even op!*
        'I find people who think that the entire #starwars series is only ten years old, funny. Look up the release of parts IV, V and VI!'
(98)   *Hee.. hele timeline weer vol met #TVOH onzin. Ik moet echt een andere twitter-app hebben waar ik mee kan filteren.*
        'Hey.. entire timeline full of #TVOH crap again. I really need to get another twitter app I can filter with.'
(99)   *Woohoo! http://t.co/ZPqg6Ras*
        'Woohoo! http://t.co/ZPqg6Ras'

These messages are not part of a conversation, but rather stand-alone utterances. The words switched here have more lexical substance than the ones we encountered before in the WhatsApp and MSN data.

Hypothesis confirmed: Yes

In sum, the way Dutch youths use English depends on the CMC mode they are using for communication. The results point towards the youths using more English overall and more conversational English in particular on CMC modes that are closer to informal face-to-face conversations (i.e. WhatsApp and MSN chat). This prodives support for the theory that English is used as a part of youth language, which youths know how to keep separate from standard language.

6.8.2 Gender

> Hypothesis:
> Male contributers switch relatively more to English than female contributors. Also, there is an interaction between gender and the semantic fields of the switches: the male youths use more terms inside the semantic field of video games than female youths.

Because gender was not known for the contributors of the MSN chat, SMS and Twitter data, only the WhatsApp data were used for this subcategory.

RQ8: GENDER			
Gender	**Absolute frequency (#)**	**# of words (total)**	**Relative frequency (%)**
Female	2918	133928	2.18
Male (excluding gay)	1442	52072	2.77
Male (gay)	301	9999	3.01

Table 21. Absolute and relative frequencies of switches per gender.

Table 21 shows the interaction between frequency of switches and gender. This relation was found to be significant (χ^2 (2, 195999) = 74.450, p = .000). The difference in frequency of switching between the two genders is small, but still significant. One of the male contributors, who happened to be overtly gay, as was clearly evident from the contents of his WhatsApp chat, used a markedly different vocabulary, which is why he has been set apart. The quantity of his switches is higher than that of his male peers.

Hypothesis confirmed: Yes

6.8.2.1 Interaction with semantic fields
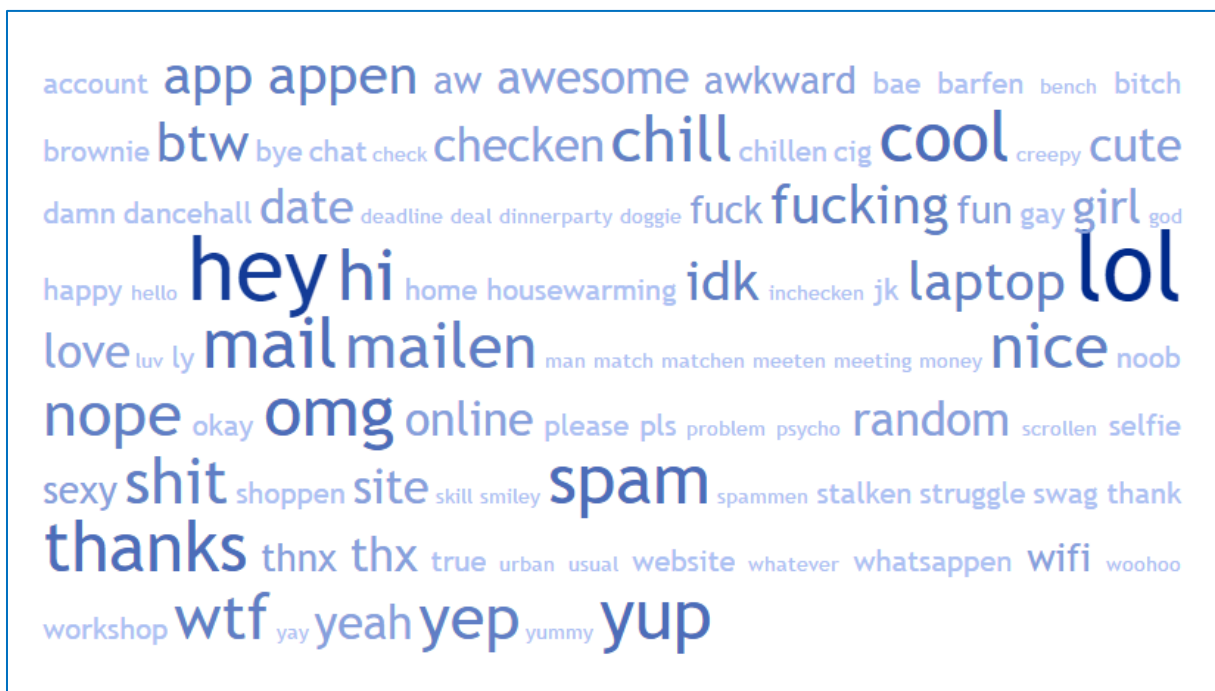
Image 10. Most frequent English lemmas by female contributors. (Full Table included in Appendix B.)

Image 11. Most frequent English lemmas by male contributors (excluding gay). (Full Table included in Appendix C.)

The two images displayed above show word clouds of the 100 most frequent English elements by females and males (excluding the gay male), respectively. The larger a word is displayed, the more it was used. A difference between the genders is clearly visible. Some words that are of large size in the male version, do not appear in the female version at all, for example: *admin, aight, base, downloaden, fucken, looten, maybe, metal, ofc, posten, relaxed, server, sick, sniper, stoned, suv*. Many of these words (*admin, base, downloaden, looten, metal, posten, server, sniper, suv*) are words used in relation to computers, video games and technology. The female contributors have fewer words unique to them; some examples are *awkward, cute, love, random, sexy, spam,* and notably*, wifi*. Most of these words seem to express some type of emotion, with the exceptions of *spam* and *wifi*. The high position of *spam* can be attributed to one female contributor, who kept sending the word 'spam' in a WhatsApp chat to literally spam her conversational partner. Out of the 63 occurences of 'spam' by female contributors, 59 were by this girl.

Image 12. Most frequent English lemmas by male contributors (including gay). (Full Table included in Appendix D.)

The above word cloud shows the most frequent English elements by males, including the gay male. Examples such *awkward, cutie, love,* and *random* seem to fit better into the word cloud of female contributors, which features similar or the same words on a large font size.
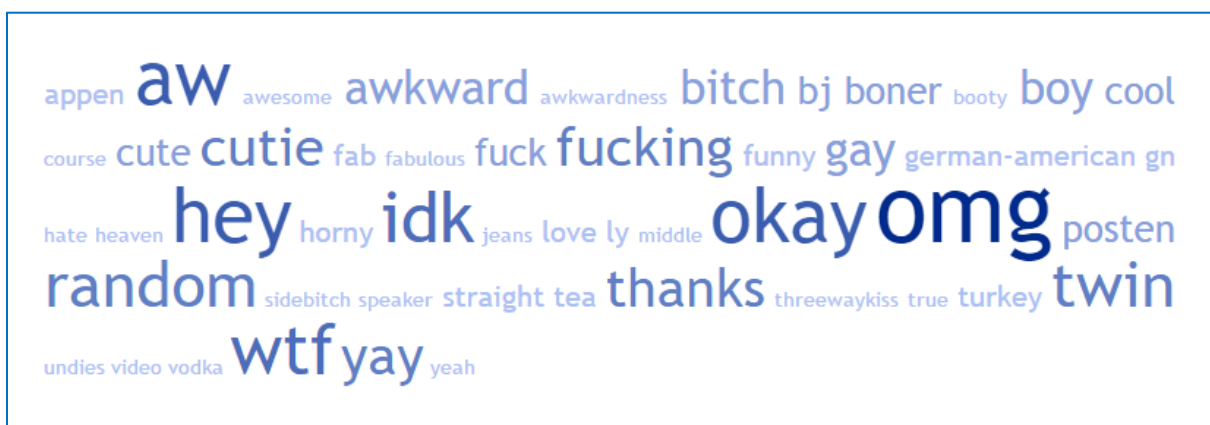
Image 13. Most frequent English lemmas by male (gay). (Full Table included in Appendix E.)

Last but not least, the final word cloud shows the 50 most frequent English elements by the abovementioned gay male. When comparing this to image 11, the most frequently switched words by the other males, the differences are striking. With the exception of *appen* ('to send a WhatsApp message') and *video*, none of the switches related to the topics of video games, computer and technology are present in his most frequent switches. While in no way this study intends to imply that this contributor differs from the others *because* he is gay, or to play into the stereotype that young homosexual males speak differently and have other interests than their heterosexual peers, the difference between this contributor and the others was obvious enough to make note of it. It is even likely that he was not the only gay contributor in the corpus, but he was the only one for which both contents and language made him stand out.

In sum, there clearly is a difference between the male and female contributors in their choice of English switches. However, this difference does not necessarily stem from code-mixing in itself. Based on the most frequent switches by the male contributors, which contained many terms related to video games, computers and technology, it is more likely that the topics that are discussed differ between the genders, also resulting in different switches.

6.8.3 Age group

Hypothesis:

Age group interacts with a number of other factors. The younger age group is less conforming to the standard language in their written CMC than the older age group. Accordingly, differences may crop up between the two age groups in a number of factors. First of all, an interaction with length: the younger age group uses more textism switches than the older age group, as textisms represent non-standard orthography typical of CMC. Second, word category: the younger age group uses relatively more English interjections and fewer English nouns than the older age group, thus using relatively more 'teenage talk' in English than the older age group. Third, graphemic integration: the younger age group integrates relatively more English elements than the older age group, thus diverging more from the English spelling. Fourth, intentionality: the younger age group uses more intentional switches than the older age group, consciously deviating from Standard Dutch. And last, dictionary status: the younger age group uses fewer English words that have been added to the Dutch dictionary and are thus part of Standard Dutch.

RQ8: AGE GROUP			
Age Group	Absolute frequency (#)	# of words (total)	Relative frequency (%)
12-17	3209	124,893	2.60
18-23	5410	267,276	2.02

Table 22. Absolute and relative frequencies of English switches per age group.

Table 22 shows the interaction between frequency of switches and age group. This interaction was found to be significant (χ^2 (1, 392169) = 117.740, p = .000). The frequency of switches is a little higher for the younger age group. This might be caused by the adolescents' tendency to use more non-conforming language, so they might use English words instead of Dutch to do set themselves apart from the Standard Dutch language.

6.8.3.1 Interaction with length

RQ8: AGE GROUP x TEXTISM SWITCHES			
Age Group	Absolute frequency (#)	# of switches (total)	Relative frequency (%)
12-17	340	3209	10.6
18-23	925	5410	17.1

Table 23. Interaction between age group and textism switches.

Table 23 shows the interaction between age group and textism switches. Contrary to what was hypothesised, the younger age group does not use relatively more textisms than the older age group. Actually, the older age group seems to use them much more, but this is due to the LOL-girl in the 18-23 age group. When her 373 tokens of the textism *lol* are excluded, the absolute frequency of textism switches by 18-23 year olds is 552 and the absolute frequency of switches (total) is 5037, which results in 11.0%. This brings the relative frequencies of the two age groups close together, but the adolescents still did not use more textisms than the

older age group.young adults: after leaving out this outlier, the relation between age group and textism switches was found to be non-significant (χ^2 (1, 8246) = .269, p = .315).

Hypothesis confirmed: No

This suggests that youths do not use English textisms as a way to rebel any more than young adults do.

RQ8: AGE GROUP x LENGTH OF SWITCH											
Age group	Total Of ID	Partial-word switches		Phrasal switches		Sentence switches		Single-word switches		Textism switches	
		#	%	#	%	#	%	#	%	#	%
12-17	3209	39	1.2	174	5.4	112	3.5	2544	79.3	340	10.6
18-23	5410	100	1.8	517	9.6	228	5.3	3580	66.2	925	17.1

Table 24. Interaction between age group and length of switch.

The interaction between age group and length of switch in total is displayed in Table 24. After leaving out the *lol*'s by the outlier again, this relation was found to be significant (χ^2 (1, 8246) = 99.782, p = .000), as opposed to the interaction between age group and textism switches only. The largest difference lies in the relative frequencies of single-word switches; it is higher for the younger age group (after excluding the LOL-girl, the relative frequency for the older age group is 71.1% - still lower than the 79.3% of the younger age group). Switches of all other lengths are used more by the older age group, in particular phrasal switches.

6.8.3.2 Interaction with word category

Before calculating the percentages of the word categories, I took a look at the most frequent switches per age group to see if any clear differences crop up.

RQ8: MOST FREQUENT SWITCHES PER AGE GROUP			
12-17		18-23	
Lemma	Total Of ID	Lemma	Total Of ID
hey	495	lol	482
yup	114	mail	167
relaxed	93	hey	166
nice	90	nice	135
lol	75	omg	133
hi	71	thanks	99
yep	70	cool	94
downloaden	66	mailen	93
spam	60	wtf	74
shit	58	yep	70
nope	57	yup	66
chill	51	okay	66
dude	50	checken	62
thanks	43	shit	53
cool	40	yes	52
wtf	37	site	50
mail	35	laptop	48
fucking	34	app	48
yes	33	btw	47
omg	32	fucking	46

Table 28. Most frequent switches per age group.

Unique to the top 20 of most frequent switches of the younger age group are the following: *relaxed, hi, downloaden, spam, nope, chill,* and *dude.* Unique to the top 20 of the older age group are *mailen, okay, checken, site, laptop, app,* and *btw.*

Though there are a few differences between the two top 20s, they are overall similar. When comparing the word categories per age group, the difference becomes more obvious:

RQ8: AGE GROUP x WORD CATEGORY													
Age group	Total Of ID	Noun		Verb		Adjective		Adverb		Interjection		Other	
		#	%	#	%	#	%	#	%	#	%	#	%
12-17	2583	669	25.9	312	12.1	406	15.7	304	11.8	875	33.9	17	0.7
18-23	3680	1526	41.5	538	14.6	609	16.5	317	8.6	656	17.8	34	0.9

Table 29. Interaction between age group and word category.

The differences in the choice of words between the two age groups are clearer to see in Table 29. The relation between age group and word category was found to be significant (χ^2 (5, 6263) = 289.285, p = .000). The largest difference is the percentage of interjections. 33.9%, over a third, of the switches of the adolescents is an interjection, almost twice as many as the 17.8% of the young adults. Another clear difference is between the nouns, which has a much higher percentage for the older age group, 41.5%, as opposed to 25.9% for the younger age group. There are slight differences in the other word categories as well. The large difference between the two age groups shows that young adults definitely switch differently from adolescents. Fewer interjections and more nouns may mean that they use more switches with more semantic weight, whereas the younger age group uses semantically emptier words such as interjections because they regard it as cool.

Hypothesis confirmed: Yes

6.8.3.3 Interaction with graphemic integration

RQ8: AGE GROUP x GRAPHEMIC INTEGRATION			
Age Group	Absolute frequency (#)	# of switches (total)	Relative frequency (%)
12-17	574	3209	17.9
18-23	292	5410	5.4

Table 25. Interaction between age group and graphemic integration.

The table above shows the interaction between age group and graphemic integration. The younger age group uses much more graphemic integration than the older age group, which is another way of writing in a non-conforming way. However, there is one specific lemma that is constantly integrated, is the lemma *hey*. This very frequently used lemma is more often integrated than not. When the 299 integrated tokens of *hey* are left out, this results in an absolute frequency of integrated switches of 275 and an absolute frequency of switches (total) of 2,910. This results in a percentage of 9.5% integrated switches by the adolescents, which is still much higher than that of the young adults. After leaving out this outlier, the relation between age group and graphemic integration was found to be significant (χ^2 (1, 8320) = 48.940, p = .000).

Some examples of graphemic integration by the younger age group are displayed below:

(100)   *school is vandaag zo <u>facking</u> saai*
        'school is so fucking boring today'
(101)   *<u>Hellooww</u>*🐱🐱
        'Helloo🐱🐱'
(102)   *ok, lijkt me wel <u>relaxt</u> als dat door gaat*
        'ok, I think it would be relaxed if that's on'
(103)   *<u>Heey</u> hoe is get*
        'Hey how is it going'

(104)   *<u>fenks</u>*
        'thanks'
(105)   *<u>jup</u> :) ik ben er ook, dus je kan binnen... ben wel met opdracht bezig'*
        'yup :) I'm here too, so you can [come] inside... am busy with assignment though'

These are some examples of graphemic integration by the older age group:

(106)   *<u>Faaack</u>! Storing Arnhem-Nijmegen -_-'*
        'Fuuuck! Interruption Arnhem-Nijmegen -_-'
(107)   *<u>Nais gurl</u>*
        'Nice girl'
(108)   *<u>Joe ken doe it!</u>*
        'You can do it!'

Hypothesis confirmed: Yes

6.8.3.4 Interaction with intentionality

RQ8: AGE GROUP x INTENTIONALITY					
Age group	Total Of ID	Intentional		Unintentional	
		Abs. freq. (#)	Rel. freq. (%)	Abs. freq. (#)	Rel. freq. (%)
12-17	3209	2662	83.0	547	17.0
18-23	5410	4221	78.0	1189	22.0

Table 26. Interaction between age group and intentionality.

This Table shows the interaction between age group and intentionality of switches. This relation was found to be significant (χ^2 (1, 8619) = 30.461, p = .000). The adolescents use relatively more intentional switches than the young adults. As stated in the previous section, this suggests that the younger age group uses English words where Dutch words could have been used, showing a non-conforming attitude towards the Dutch standard language. Some examples of these intentional switches by adolescents are *awkward, btw, chill, I know, idk,* and *skippen* ('to skip').

Hypothesis confirmed: Yes

6.8.3.5 Interaction with dictionary status

RQ8: AGE GROUP x DICTIONARY STATUS					
Age group	Total Of ID	No		Yes	
		Abs. freq. (#)	Rel. freq. (%)	Abs. freq. (#)	Rel. freq. (%)
12-17	3209	1889	58.9	1320	41.1
18-23	5410	2649	49.0	2761	51.0

Table 27. Interaction between age group and dictionary status.

Table 27 shows the interaction between age group and dictionary status. This relation was found to be significant (χ^2 (1, 8619) = 79.202, p = .000). The percentage of switches that are not added to the Van Dale dictionary is 58.9% for the younger age group and 49.0% for the older age group. This is a considerable difference. Perhaps adolescents use more novel English elements because of their status (covert prestige); their use of many words that have not been added to the Dutch dictionary might reveal a genuine effort to use words that do not conform to the standard Dutch language.

Some examples of such words that are not in the Dutch dictionary, but the adolescents use nevertheless:

(109) *Ait ik okk*
      'Aight me tooo'
(110) *Jaaa awesome*
      'Yeees awesome'
(111) *Nice, ging het goed?*
      'Nice, did it go well?'
(112) *nawttt*
      'not'

These are all rather basic words that can unproblematically be translated into Dutch (to 'oké', 'geweldig', 'fijn', and 'niet' respectively).

Hypothesis confirmed: Yes

In conclusion, there are a number of clear differences between the two age groups. These differences are in line with the theory that adolescents use more non-conforming language, known as youth language. Although it is likely that young adults still make use of youth language, the results indicate that they do this to a lesser extent than the adolescents.

6.9 Other findings

I found some interesting occurrences of code-mixing in the corpus which could not be classified into one of the eight research questions. These are discussed in this miscellaneous section.

6.9.1 Repetition of English element

There were cases in which contributors repeated the English element, as in (113) and (114).

(113) *Nog een paar dagen:) en dan heb ik em! #iPhone4s #happyhappyhappy*
      'A few more days:) and then I've got it! #iPhone4s #happyhappyhappy

(114) *party party party :p*
       'party party party :p'

The repetition of words shows sheer youthful excitement. The fact that these words conveying excitement are in English shows that English elements are used to express emotions and affective language. It is in line with the hypothesis that youths code-mix English in their Dutch language to boost their expressivity and to intensify their excitement.

These words were coded as one single-word switch, including a note 'repetition of switch'.

6.9.2 Insertion of Dutch into English sentence

While the data have revealed that the way Dutch youths switch to English is mostly by inserting English words and phrases into Dutch sentences, there were two cases where the exact opposite happened: the entire sentence was written in English, but a Dutch element was inserted into it. Both cases have been coded as one English sentence switch, including a note that Dutch was inserted into the switch.

(115) *so yes I am sogging*
       'so yes I am procrastinating my studies'

In this sentence, the only non-English word is the Dutch verb *soggen*, inflected to fit into the English sentence. *Soggen* is derived from the abbreviation *sog*, which is short for *studie-ontwijkend gedrag* ('study avoiding behavior') and means doing anything to avoid having to study, or procrastinating one's studies. An English equivalent to this word does not exist. The term *soggen* is so common among Dutch university students, that any translation would not carry the same pragmatic weight.

(116) *Stoefpears run the world*
       'stewed pears run the world'

In example (116), only the word *stoefpears* is not truly English. Yet, *stoefpears* is not a Dutch word either. It exemplifies intentional Dunglish, in Dutch 'steenkolenengels', English filled with errors because Dutch constructions are translated literally. The correct Dutch word is *stoofperen* ('stewed pears'), where the second part of the compound *peren* has been correctly translated to the English 'pears', but the first part *stoof*, has not been translated at all, only the spelling of the vowel has been changed to how *stoof* would be pronounced in English. The letters 'oo' would be pronounced as an /u:/ in English, but as an /o:/ in Dutch, which in its turn writes /u/ as 'oe'. This is why *stoof* has been changed to *stoef*: to make it look more 'English'. In all probability this has been done intentionally, as a joke. Still, the speaker may have jokingly used this Dunglish switch because they did not know the proper English translation of *stoofperen*.

For both sentences in this section, the question remains why these youths used English sentences to begin with when they could not translate them entirely into English. Though we can only speculate about this, it can be assumed that it is done just for fun, or to generate a comical effect.

6.9.3 Dutch abbreviations of English words

There were a few cases where English words were abbreviated in a way they are not by native speakers of English.

(117)   *wrong turn benk nu owk aant <u>downe</u>*
        'i am downing wrong turn too now'

The English verb *to download* has been abbreviated, only leaving the first half of the original verb, *down*, and a Dutch clipped suffix. Though it would not be impossible for native speakers of English to abbreviate *to download* in the same way, it is not a common practice.[9] This reveals creative, inventive code-mixing.

(118)   *Stuur es <u>screen</u> van zn profiel?*
        'send a screen of his profile?'

Here, the English word *screenshot* has been abbreviated to *screen*. Again, this is not a common abbreviation in English, as attested by the Urban Dictionary, but invented by the Dutch youth him/herself.

Such abbreviations of English words were also found by De Decker and Vandekerckhove (2012), who coded them as semantic integration (which is a form of integration that was not included in the present study, see section 3.5 for elaboration).

6.9.4 Calques

In some cases, when using English elements in their sentences, the Dutch youths used calques, i.e. indiscriminate translations of foreign expressions.

(119)   *Vanaf daar wordt dit <u>shit</u> <u>real</u>*
        'From that point on this shit gets real'

In (119), the English phrase *this shit gets real* is partially translated into Dutch and partially kept English. The problem here lies with the Dutch demonstrative pronoun *dit* ('this'). In Dutch, there are two forms of this word, namely *dit* and *deze*. In this case, because of the gender of *shit*, the grammatically correct form would be *deze*, but *dit* is used instead. Although it cannot be proven beyond a doubt, it is likely that it was done because *dit* is phonologically more similar to the English *this*.

(120)   *Ja ben ik dat <u>casual</u>*
        'Yeah am I that casual'

A similar thing happens in (120). The word *that* in the phrase *that casual* would be translated to Dutch as *zo*, not as *dat*, the determiner which is the Dutch equivalent for *that* in most cases, but not in this one! The fact that *dat* is a translation for *that* in some cases, plus their phonological similarity, may be the reasons why the contributor chose to write *dat* here.

---

[9] The Urban Dictionary, an Internet-based dictionary of slang words and phrases which contains much (online) youth language and is kept up-to-date by its users, does not name it as one of the meanings for 'down' or 'to down'.

6.9.5 Memes

In section 6.6.1, the fixed expression *like a boss* was named as an example of a textual internet meme, also called meme for short. The online version of the Oxford English Dictionary defines the term 'meme' as follows:

> "An image, video, piece of text, etc., typically humorous in nature, that is copied and spread rapidly by Internet users, often with slight variations."

Like a boss refers to the song 'Like a boss' by American comedy trio The Lonely Island.[10] It is not the only meme that appeared in the corpus. Here are some more examples:

(121)   *Can't tell if troll or just very very stupid*
(122)   *http://t.co/gawSAep Double rainbow all the way :p*
(123)   *Xbox meenemen naar tentamen #winning*
        'Bringing xbox to exam #winning'

While (121) and (122) also fall into the category of fixed expressions, not all memes are fixed expressions. As can be seen, memes can take the form of phrases and sentences, but also single words, like (123).

The above three memes are all references to popular culture. (121) is a reference to a screencap of the TV cartoon *Futurama* with the text 'not sure if trolling or just stupid' (see image 14 below), (122) is a reference to an utterance in a popular YouTube video,[11] and (123) refers to the American actor Charlie Sheen, who often repeated the term 'winning' in interviews back in 2011.[12] All of them became popular online and were quoted often.

---

[10] The music video for the song 'Like a boss' by The Lonely Island:
https://www.youtube.com/watch?v=NisCkxU544c
[11] The 'Double Rainbow' YouTube video that went viral: https://www.youtube.com/watch?v=OQSNhk5ICTI
[12] A compilation of Charlie Sheen using the word 'winning' back in 2011:
https://www.youtube.com/watch?v=pipTwjwrQYQ
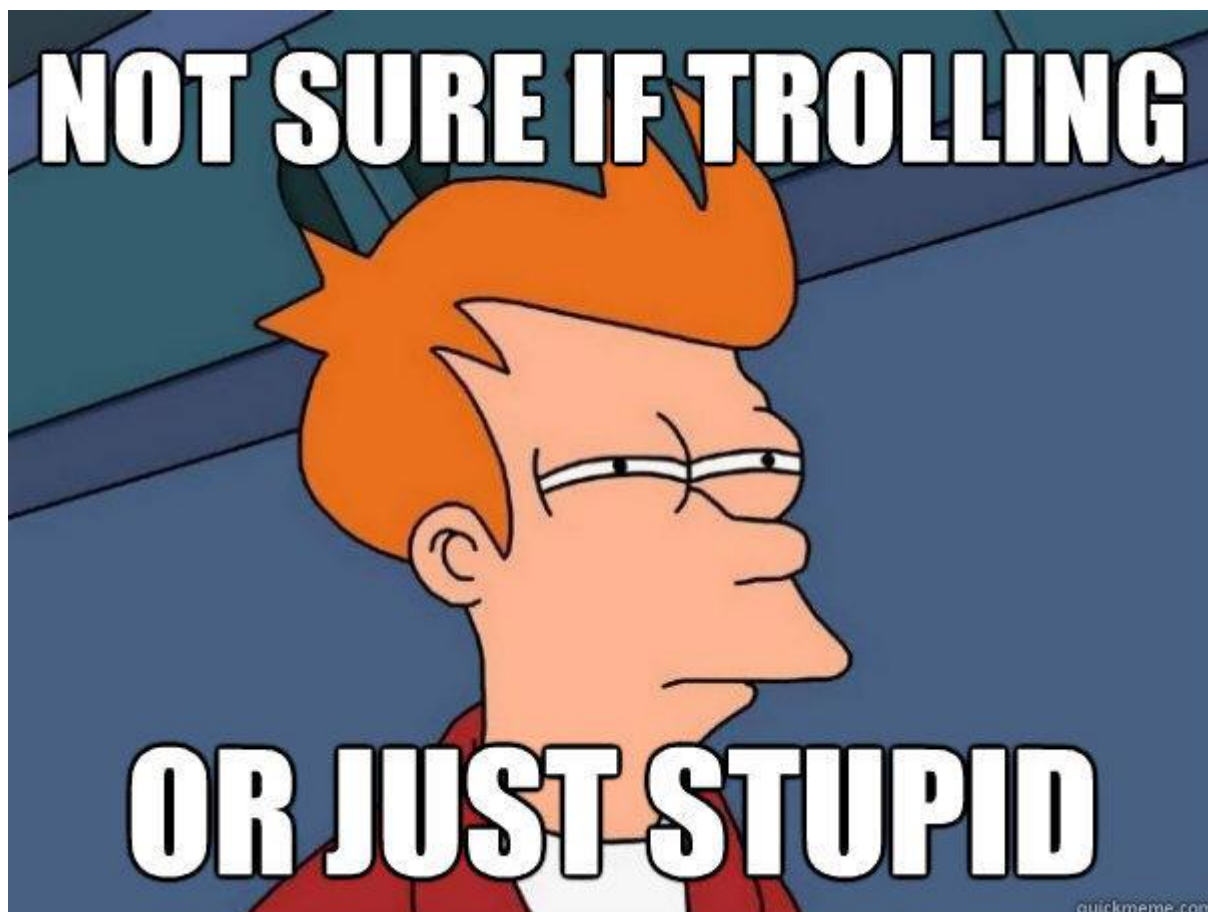
Image 14. The *Futurama* meme.

What makes memes interesting in this data is that they are not merely fixed expressions or single words that have a pragmatic meaning, but the fact that they have become popular and are mainly used in computer-mediated communication. While memes may also be referred to in face-to-face conversations, this does not happen until they have earned a popular status online, and they are definitely more popular on the internet than *irl* (in real life).

Another property of memes is that they are generally short-lived. As the Oxford English Dictionary mentions, they are spread rapidly; and they often lose popularity as fast as they gain it. The meme *winning* of (123) is a great example of this: it rose to popularity in 2011, when a number of the interviews in which Charlie Sheen said the word were broadcasted. The example from the corpus came from the Twitter data, which was also collected in 2011 – right when the meme was popular. Nowadays, this meme is already considered old and not cool to use anymore among youths. If a corpus of tweets by Dutch youths was collected at the time of writing, in 2016, the memes found would be very different from the ones found in the current study. This just shows fleeting much youth language trends are.

6.9.6 Conclusion

This final section of results discussed particularly creative and interesting uses of English by Dutch youths. It shows that Dutch youths, besides inserting English words into their Dutch language and integrating them in accordance with Dutch grammar rules, inventively use English to boost their expressivity, play with language, be pragmatic in code-mixing, and perhaps because they inadvertly got the languages mixed up sometimes.

7 <u>Conclusion</u>

This master's thesis has described the way Dutch youths make use of English elements in CMC. The general research question I have aimed to answer was as follows: How and how much do Dutch youths code-mix and insert elements from the English language in Dutch written computer-mediated communication? A variety of factors contributing to code-mixing, youth language and computer-mediated communication have been taken into account to give a detailed analysis of how the English elements are inserted in the Dutch language exactly. By coding a large corpus of Dutch youths' CMC for all these factors, calculating and comparing frequencies and discussing many examples relating to the results, I have provided a quantitative and qualitative analysis of the Dutch youths' code-mixing.

Based on the De Decker and Vandekerckhove's (2012) study on English in Flemish youths' CMC, eight sub-questions and hypotheses were formulated, considering the following factors: length of switch, multiplicity, word category, integration, semantic fields, intentionality, frequency of switches and the language-external factors CMC mode, gender and age group. Because of the large number of hypotheses, the results are summarised in Table 30, showing which hypotheses are confirmed and which are refuted.

CONCLUSION: RESULTS SUMMARISED											
	Length	Multiplicity	Word category	Integration	Semantic fields	Dictionary status	Intentionality	Frequency of switches	CMC mode	Gender	Age
Length	✓										
Multiplicity		✓									
Word category			✓								
Integration				✓							
Semantic fields					✓						
Dictionary status											
Intentionality	✓		✓			✓	✓				
Frequency of switches					✓	✓	✓				
CMC mode			✓						✓		
Gender					✓					✓	
Age	✗		✓	✓		✓	✓				

Table 30. Outcome of the hypotheses (✓ = hypothesis confirmed, ✗ = hypothesis refuted).

Table 30 shows the analysed factors and whether the hypotheses about them were confirmed or not, including hypotheses about interactions between factors. The empty cells in this table present variables/interactions about which no hypotheses were formulated. As can be seen, all hypotheses, except for one, are confirmed. Many of the findings were in line with those of De Decker and Vandekerckhove (2012); this shows that the use of English does not seem to differ greatly between Dutch and Flemish youths. The present thesis thus serves as an addition to and elaboration on their study.

There are a number of reasons why Dutch youths code-mix between Dutch and English: because of lexical need, but also to help create their own variant of the language, 'teenage talk'. They show a certain amount of creativity, e.g. through integration, abbreviations, and puns. Yet their use of English mainly consists of single English words inserted into Dutch sentences, a perfect example of insertional code-mixing. Most of these single-word switches are nouns or interjections; this is in line with the theories about these two word categories being very borrowable (Muysken, 1999; Matras, 2007). Even when using switches longer than one word, they are often (semi-)fixed expressions, in line with the study on multiple-word switches by Zenner and Geeraerts (2015). It is clear that Dutch youngsters are not native speakers of English and do not use it with the same proficiency level as Dutch.

Still, switches that fall in the category of teenage talk are the most frequent. The frequency of switches show us which words are currently fashionable to use among Duth speaking youths. Words such as *nice, thanks* and *relaxed* and textisms such as *lol* and *omg* are very frequent in Dutch youths' CMC. Another semantic field that occurs notably often is that of computer and technology, because many of these terms do not have a Dutch equivalent.

An influence of language-external factors has also been found. In particular, the age of the contributor plays a role in how they use English in CMC. The adolescents display a much more non-conforming behavior to the standard language when it comes to their lexical choices than young adults; this shows in their choice of relatively more intentional switches, more English elements that are not part of the Standard Dutch dictionary (yet) and more interjections than the young adults. This is in line with the findings of Verheijen (2016), who also found that adolescents are less conforming to the standard language. Depending on the CMC mode, youths adjust the amount of English they use in their Dutch sentences. When a CMC mode is more synchronous and has one-on-one interactivity, the use of English switches, in particular conversational words such as interjections, is more frequent, in line with research on the difference between CMC modes by Verheijen (2016). Lastly, it seems that boys use switches from different semantic fields than girls. However, it is likely that this has more to do with their conversational topics than with their use of English.

To answer the general research question, it is clear that Dutch youths use a considerable amount of English in their CMC: about one in forty words in the analysed data was an English element. This seems like a rather large amount at first sight, but the results of the semantic fields, the CMC modes and the word categories show that the code-mixing is not random and all over the place, even though the youths will not be conscious of that (Myers-Scotton, 2002). Most of the switches to English are conversational words, such as greetings and swear words. Teenage talk is also frequently in English. English has also been shown to be used for untranslatable jokes, such as memes or puns. These are all indicative of informal language used by youths among each other. This suggests that their use of English is only temporary and may possibly decrease with age – since they adjust their use of English depending on the CMC mode used, and generally do not speak youth language to adults (Schoonen & Appel, 2005), it is questionable whether these words will remain part of their vocabulary in the long run.

Because of the huge amount of data and the given time limit of this thesis, it was unfortunately not possible to manually search through the entire corpus, which means that some switches to English might be missing, in particular, non-standard spelling variants. Despite this, I have attempted to be as complete as possible. It was also beyond the scope of this thesis to go deeper into the influence of the individual contributors; here lies an opportunity for future studies.

The fact that over 8,000 switches have been coded for many factors makes it a promising data set for future research. There are definitely more aspects of code-mixing that could be studied with this data. Also, the data could be compared a corpus of Dutch youths' use of English in face-to-face conversations. Another fruitful area of research could be to study the use of English in CMC by older people in comparison, to find out if English also features prominently in their Dutch, or if it is really only typical of cool youth language.

All in all, code-mixing by Dutch youths is definitely not something that happens because English is completely replacing their Dutch. Although English has an undeniable influence on

the Dutch language in some fields, their CMC is mostly in Dutch. Still, the use of English gives the youths' online language colour. Whether it is done through textism switches and internet memes or by simply using an English word where the Dutch equivalent could have sufficed, it distinguishes them from other (older) speakers of Dutch – something youths can do like no other.

References

Bi, N. Z. (2011). An investigation into English mixing in Chinese Internet language. *World Journal of English Language*, 1(2), 60-67.

De Bond tegen leenwoorden. (2011). Retrieved from: http://bondtegenleenwoorden.nl/

Braak, J. van den (2002). Het verschijnsel straattaal: een verkenning. *Respons: Mededelingen van het Meertens Intituut*, 13-19.

Cenoz, J., & Jessner, U. (Eds.). (2000). *English in Europe: The acquisition of a third language*. Clevedon: Multilingual Matters.

Cornips, L. (2004). Straattaal: Sociale betekenis en morfo-syntactische verschijnselen. In: De Caluwe, J., De Schutter, G., Devos, M., & Van Keymeulen, J. (Ed.). *Taeldeman, man van de taal, schatbewaarder van de taal [Liber Amicorum Johan Taeldeman]*. Gent: Academia Press.

Danet, B., & Herring, S. C. (2007). *The multilingual Internet: Language, culture, and communication online*. Oxford: Oxford University Press.

De Decker, B. & Vandekerckhove, R. (2012). English in Flemish adolescents' computer-mediated discourse: A corpus-based study. *English World-Wide*, 33(3), 321-351. DOI: 10.1075/eww.33.3.04ded

De Decker, B., & Vandekerckhove, R. (2013). De integratie van Engels in Vlaamse jongerentaal kwantitatief en kwalitatief bekeken: das wel nice! :p. *Nederlandse Taalkunde*, *18*(1), 2-34.

Desai, R. A., Krishnan-Sarin, S., Cavallo, D., & Potenza, M. N. (2010). Video-gaming among high school students: health correlates, gender differences, and problematic gaming. *Pediatrics*, 126(6), 1414-1424.

Di Sciullo, A., Muysken, P., & Singh, R. (1986). Government and code-mixing. *Journal of Linguistics*, 22(1), 1-24.

Dor, D. (2004). From Englishization to imposed multilingualism: Globalization, the Internet, and the political economy of the linguistic code. *Public Culture*, 16(1), 97-118.

Fromkin, V.A. (Ed.). (2000). *Linguistics: an introduction to linguistic theory.* Oxford: Blackwell.

Georgakopoulou, A. (2011). Computer-mediated communication. In: Östman, J., & Verschueren, J. (Eds.). *Pragmatics in practice* (pp. 93-110). Amsterdam, Philadelphia: John Benjamins Publishing.

Gerritsen, M., Korzilius, H., Meurs, F. van, & Gijsbers, I. (2000). English in Dutch commercials: Not understood and not appreciated. *Journal of Advertising Research*, 40(4), 17-31.

Görlach, M. (2002). *English in Europe*. Oxford: Oxford University Press.

Griffiths, M. D., Davies, M. N., & Chappell, D. (2004). Online computer gaming: a comparison of adolescent and adult gamers. *Journal of Adolescence*, 27(1), 87-96.

Hargittai, E., & Hinnant, A. (2008). Digital inequality differences in young adults' use of the Internet. *Communication Research*, 35(5), 602-621.

Hassan, N., & Hashim, A. (2009). Electronic English in Malaysia: Features and language in use. *English Today*, 25(04), 39-46.

Herring, S. C. (2010). Computer-mediated conversation: Introduction and overview. *Language@ internet*, 7(2), 1-12.

Hornikx, J., Meurs, F. van, & de Boer, A. (2010). English or a local language in advertising? The appreciation of easy and difficult English slogans in the Netherlands. *Journal of Business Communication*, 47(2), 169-188.

Joshi, A. 1985. Processing of sentences with intrasentential code switching. In: Dowty, D.R., Karttunen, L. & Zwicky, A. M. (Eds.) *Natural language parsing: Psychological,*

computational and theoretical perspectives (pp. 190-205). Cambridge: Cambridge University Press.

Kytölä, S. (2013). Multilingual language use and metapragmatic reflexivity in Finnish internet football forums: A study in the sociolinguistics of globalization. *Jyväskylä studies in humanities,* 200. University of Jyväskylä.

Matras, Y. (2007). The borrowability of structural categories. In: Matras, Y. & Sakel, J. *Empirical Approaches to Language Typology* (pp. 31-74).

Meurs, F. van, Korzilius, H. P. L. M., & Hermans, J. J. (2004). The influence of the use of English in Dutch job advertisements: An experimental study into the effects on text evaluation, on attitudes towards the organisation and the job, and on comprehension. *ESP Across Cultures,* 1(1), 93-110.

Meurs, F. van, Korzilius, H. P. L. M., & Hollander, A. D. (2006). The use of English in job advertisements on the Dutch job site Monsterboard.nl and factors on which it depends. *ESP Across Cultures*, 3, 103-123.

Meyerhoff, M. (2006). *Introduction to sociolinguistics.* London, New York: Routledge.

Millar, R. M. (2007). *Trask's historical linguistics*. London, New York: Routledge.

Muysken, P. C. (1995). Code-switching and grammatical theory. In: Milroy, L., & Muysken, P. (Eds). *One speaker, two languages: Cross-disciplinary perspectives on code-switching* (pp. 177-198). Cambridge: Cambridge University Press.

Muysken, P. (1999). Three processes of borrowing: borrowability revisited. In: Extra, G., & Verhoeven, L. (Eds.) *Bilingualism and migration* (pp. 229-246). Berlin, New York: Mouton de Gruyter.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in code-switching.* Oxford: Clarendon.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes.* Oxford: Oxford University Press.

Nortier, J. (2016). Characterizing Urban Youth Speech Styles in Utrecht and on the Internet. *Journal of Language Contact*, 9(1), 163-185.

Onna, B. van, & Jansen, C. (2006). How multilingual are the Dutch really? On proficiency in Dutch, English, French, and German in Dutch organizations. *BELL-Belgian journal of English language and literatures*, 169-180.

Oostdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2013). The construction of a 500-million-word reference corpus of contemporary written Dutch. In: P. Spyns & J. Odijk (Eds.) *Essential Speech and Language Technology for Dutch: Results by the STEVIN Programme* (pp. 219-247). Heidelberg: Springer.

Oxford English Dictionary. "By the way." [cited 9 August 2016.] Retrieved from: http://www.oed.com/view/Entry/25523?rskey=rSAD2l&result=1#eid11216604

Oxford English Dictionary. "Meme." [cited 9 August 2016.] Retrieved from: http://www.oed.com/view/Entry/239909?redirectedFrom=meme#eid

Oxford English Dictionary. "My pleasure." [cited 9 August 2016.] Retrieved from: http://www.oed.com/view/Entry/145578?redirectedFrom=my+pleasure#eid29606326

Oxford English Dictionary. "Who cares." [cited 9 August 2016.] Retrieved from: http://www.oed.com/view/Entry/27902?redirectedFrom=who+cares#eid10191517

Pfaff, C.W. (1979). Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language*, 55(2), 291–318.

Poplack, S. (1980). "Sometimes I'll start a sentence in Spanish y termino en espaniol": Toward a typology of code-switching. *Linguistics,* 18(7-8), 581-618. DOI: 10.1515/ling.1980.18.7-8.581

Schoonen, R., & Appel, R. (2005). Street language: A multilingual youth register in the Netherlands. *Journal of multilingual and multicultural development*, 26(2), 85-117.

Sebba, M. (2012). Multilingualism in written discourse: An approach to the analysis of multilingual texts. *International Journal of Bilingualism*, 1367006912438301. DOI: 10.1177/1367006912438301

Sijs, N. van der (2009). Loanwords in Dutch. In: Haspelmath, M., & Tadmor, U. *Loanwords in the world's languages: A comparative handbook* (pp 338-359). Berlin: De Gruyter Mouton.

Steinbock, D. (2016). *Tagcrowd: create your own word cloud from any text to visualize word frequency.* Retrieved from: http://tagcrowd.com/

Stichting LOUT. (2016). Retrieved from: https://onzetaal.nl/over/stichting-lout/

Stichting Nederlands. (2016). Retrieved from: http://www.stichtingnederlands.nl/

Thurlow, C., Lengel, L., & Tomic, A. (2004). *Computer mediated communication.* SAGE, London, Thousand Oaks, California, New Delhi. Retrieved from: https://books.google.nl/books?id=ED5Htazhr38C&hl=nl

Treurniet, M., De Clercq, O., van den Heuvel, H., & Oostdijk, N. (2012). Collecting a corpus of Dutch SMS. *Proceedings LREC 2012*, 2268-2273.

Treurniet, M., & Sanders, E. (2012). Chats, tweets and SMS in the SoNaR corpus: Social media collection. In: D. Newman (Ed.) *Proceedings of the 1st Annual International Conference Language, Literature & Linguistics* (pp. 268–271). Singapore: Global Science and Technology Forum.

Urban Dictionary. "Down." [cited 9 August 2016]. Available from http://www.urbandictionary.com/define.php?term=down.

Van Dale's Great Dictionary of the Dutch Language. (2016). Retrieved from: vandale.nl/zoeken/zoeken.do

Vandekerckhove, R., Cuvelier, P., & Decker, B. (2015). *Language Matters*, 46(3), 344-363.

Verheijen, L. (2015). Out-of-the-ordinary orthography: The use of textisms in Dutch youngsters' written computer-mediated communication. *York Papers in Linguistics, special issue, PARLAY Proceedings*, 2, 127-142.

Verheijen, L. (2016). De macht van nieuwe media: hoe Nederlandse jongeren communiceren in sms'jes, chats en tweets. In: Van de Mieroop, D., Buysse, L., Coesemans, R., & Gillaerts, P. (Eds.). *De macht van de taal: Taalbeheersingsonderzoek in Nederland en Vlaanderen* (pp. 275-293). Leuven / Den Haag: Acco.

Warschauer, M., Said, G. R. E., & Zohry, A. G. (2002). Language choice online: Globalization and identity in Egypt. *Journal of Computer-Mediated Communication*, 7(4). DOI: 10.1111/j.1083-6101.2002.tb00157.x

Zenner, E., & Geeraerts, D. (2015). I'm Queen of the World! (Semi-)fixed English expressions and constructions in Dutch. *Taal & Tongval*, 67(2), 247-274.

Zenner, E., Speelman, D., & Geeraerts, D. (2015). A sociolinguistic analysis of borrowing in weak contact situations: English loanwords and phrases in expressive utterances in a Dutch reality TV show. *International Journal of Bilingualism*, 19(3), 333-346. DOI: 10.1177/1367006914521699

Appendices

Appendix A: Frequency of the 100 most frequent switches.

	Lemma	Frequency
1	hey	661
2	lol	557
3	nice	225
4	mail	202
5	yup	180
6	omg	165
7	thanks	142
8	yep	140
9	cool	134
10	mailen	124
11	wtf	111
12	shit	111
13	relaxed	107
14	downloaden	94
15	chill	93
16	nope	85
17	yes	85
18	checken	84
19	hi	84
20	fucking	80
21	site	77
22	okay	75
23	online	73
24	btw	72
25	app	68
26	spam	68
27	dude	61
28	laptop	60
29	idk	57
30	appen	46
31	aw	46
32	thnx	42
33	awesome	38
34	k	38
35	fucken	36
36	yay	36
37	i know	36
38	website	34
39	fail	33
40	chillen	33
41	no	33
42	aight	33
43	date	32
44	yeah	31
45	thx	30
46	wifi	29
47	account	26
48	random	25
49	live	25
50	damn	24
51	show	22
52	true	21

53	cute	20
54	bitch	19
55	awkward	19
56	fuck	19
57	gay	19
58	posten	18
59	have fun	18
60	shoppen	18
61	base	18
62	party	17
63	deadline	17
64	inchecken	17
65	boy	17
66	what	16
67	server	16
68	woohoo	16
69	sexy	16
70	bye	16
71	workshop	15
72	update	15
73	meeten	14
74	right	14
75	please	14
76	game	14
77	np	13
78	stalken	13
79	timeline	13
80	updaten	13
81	sick	12
82	whatsappen	12
83	bullshit	12
84	social media	12
85	dread	12
86	ofc	12
87	on	11
88	single	11
89	stoned	11
90	skypen	11
91	go	11
92	chat	11
93	joint	11
94	epic	11
95	twin	10
96	yas	10
97	selfie	10
98	portfolio	10
99	indeed	10
100	check	10

Appendix B: Frequency of the 100 most frequent switches by female contributors.

	Lemma	Frequency
1	lol	413
2	hey	314
3	omg	82
4	mail	79
5	yup	74
6	hi	69
7	cool	67
8	spam	63
9	thanks	59
10	yep	46
11	yes	45
12	mailen	41
13	nice	40
14	shit	39
15	nope	37
16	chill	37
17	wtf	37
18	appen	35
19	idk	29
20	btw	27
21	laptop	26
22	i know	26
23	app	24
24	fucking	20
25	no	17
26	online	17
27	cute	16
28	thx	15
29	yeah	15
30	random	14
31	checken	14
32	site	14
33	date	14
34	awesome	14
35	wifi	13
36	thnx	12
37	awkward	12
38	have fun	11
39	aw	9
40	right	9
41	stalken	8
42	chat	8
43	whatsappen	7
44	selfie	7
45	pls	7
46	sexy	7
47	account	6
48	please	6
49	ly	6
50	okay	6
51	true	6
52	workshop	6
53	cig	6
54	brownie	6
55	jk	6

56	thanks girl	6
57	housewarming	5
58	thank you	5
59	struggle	5
60	ever	5
61	dancehall	5
62	chillen	5
63	shoppen	5
64	website	5
65	bitch	5
66	noob	5
67	barfen	5
68	gay	5
69	whatever	4
70	swag	4
71	inchecken	4
72	yummy	4
73	take him	4
74	woohoo	4
75	damn	4
76	check	4
77	deadline	4
78	as usual	4
79	money	4
80	scrollen	4
81	dinnerparty	4
82	me too	4
83	meeten	4
84	spammen	4
85	hello	4
86	sister	3
87	outfit	3
88	o no	3
89	np	3
90	not	3
91	psycho	3
92	supernice	3
93	kids	3
94	nice girl	3
95	single	3
96	liken	3
97	look	3
98	love you	3
99	skippen	3
100	matchen	3

Appendix C: Frequency of the 100 most frequent switches by male contributors (excluding gay).

	Lemma	Frequency
1	nice	118
2	hey	75
3	okay	49
4	yep	43
5	shit	35
6	lol	32
7	chill	26
8	aight	24
9	fucken	23
10	k	22
11	base	18
12	downloaden	17
13	nope	17
14	thanks	15
15	cool	14
16	fucking	14
17	thx	12
18	wtf	12
19	ofc	12
20	btw	12
21	omg	11
22	site	11
23	yes	11
24	server	11
25	on	10
26	no	10
27	online	10
28	app	10
29	thnx	10
30	yay	9
31	yas	9
32	awesome	9
33	date	9
34	idk	9
35	g	9
36	show	8
37	admin	8
38	maybe	8
39	aw	8
40	appen	8
41	yup	7
42	mailen	7
43	sick	7
44	chillen	7
45	posten	6
46	ticket	6
47	account	6
48	damn	6
49	relaxed	6
50	stoned	6
51	looten	6
52	mail	6
53	money	6

54	gay	6
55	suv	5
56	fuck	5
57	boy	5
58	checken	5
59	live	5
60	true	5
61	laptop	5
62	crashen	4
63	fk	4
64	meeten	4
65	metal floor	4
66	taggen	4
67	chainsaw	4
68	dude	4
69	bullshit	4
70	why	4
71	I know	4
72	gamen	4
73	bitch	4
74	dmr	4
75	update	4
76	safe	4
77	the fuck	4
78	yolo	4
79	joint	4
80	right	3
81	jesus	3
82	like	3
83	torrent	3
84	easy	3
85	epic	3
86	event	3
87	selfie	3
88	skippen	3
89	spectaten	3
90	sniper nest	3
91	uploaden	3
92	body	3
93	sale	3
94	though	3
95	tho	3
96	gamer	3
97	u	3
98	ay	3
99	moshpit	3
100	website	3

Appendix D: Frequency of the 100 most frequent switches by male contributors (including gay).

	Lemma	Frequency
1	nice	118
2	hey	94
3	okay	66
4	omg	66
5	yep	43
6	shit	35
7	lol	32
8	wtf	27
9	chill	26
10	idk	25
11	aight	24
12	aw	24
13	fucken	23
14	k	22
15	thanks	21
16	fucking	19
17	base	18
18	downloaden	17
19	Nope	17
20	cool	17
21	yay	15
22	ofc	12
23	thx	12
24	btw	12
25	gay	11
26	server	11
27	site	11
28	online	11
29	yes	11
30	awesome	10
31	appen	10
32	app	10
33	random	10
34	no	10
35	thnx	10
36	on	10
37	date	9
38	boy	9
39	g	9
40	twin	9
41	yas	9
42	posten	9
43	admin	8
44	maybe	8
45	show	8
46	bitch	8
47	chillen	8
48	sick	7
49	i know	7
50	yup	7
51	mailen	7
52	ticket	6
53	true	6

54	stoned	6
55	money	6
56	account	6
57	damn	6
58	live	6
59	looten	6
60	mail	6
61	relaxed	6
62	yolo	5
63	fuck	5
64	awkward	5
65	checken	5
66	by the way	5
67	laptop	5
68	suv	5
69	metal floor	4
70	meeten	4
71	gamen	4
72	safe	4
73	the fuck	4
74	joint	4
75	like	4
76	taggen	4
77	update	4
78	u	4
79	dude	4
80	crashen	4
81	dmr	4
82	why	4
83	fk	4
84	chainsaw	4
85	bullshit	4
86	body	3
87	bj	3
88	uploaden	3
89	skippen	3
90	moshpit	3
91	cute	3
92	cutie	3
93	wifi	3
94	liken	3
95	website	3
96	selfie	3
97	sniper nest	3
98	ay	3
99	event	3
100	fast lane	3

Appendix E: Frequency of the 50 most frequent switches by male (gay).

	Lemma	Frequency
1	omg	55
2	hey	19
3	okay	17
4	idk	16
5	aw	16
6	wtf	15
7	twin	8
8	random	8
9	yay	6
10	thanks	6
11	gay	5
12	by the way	5
13	fucking	5
14	bitch	4
15	boy	4
16	awkward	4
17	i know	3
18	posten	3
19	cutie	3
20	cute	3
21	cool	3
22	bj	3
23	fab	2
24	high tea	2
25	ly	2
26	appen	2
27	boner	2
28	hey cutie	2
29	love you too	2
30	horny	2
31	turkey	2
32	funny	2
33	german-american	2
34	fuck off	2
35	gbf	1
36	gfy	1
37	gn	1
38	heaven	1
39	as fuck	1
40	amazing	1
41	gn cutie	1
42	aw cutie	1
43	daten	1
44	awesome	1
45	chillen	1
46	fucking awkward	1
47	awkwardness	1
48	first like	1
49	fabulous	1
50	deadline	1