Imitation is the most sincere form of flattery

A corpus study of emoticons and abbreviations in English blogs written by Swedes

Oscar Svensson
Abstract

This study investigates the use of emoticons and abbreviations in a mode of Computer-Mediated Communication (CMC), blogs written by Swedish individuals in English. The aim of the study is to find out whether the emoticons and abbreviations used in these blogs are tailor-made by Swedish users or if they are the same as those found on the rest of the Internet. Another question asked by the study concerns if the use of emoticons and abbreviations differs depending on the gender and age group of the blogger. The method used is mainly concordance searches in a corpus, using the concordance program AntConc to find particular features of CMC. The data used for the study is a corpus of one hundred thousand words from blogs all over the Internet on a wide range of topics written by males and females of different age groups. The results of the investigation indicate that no local variations of abbreviations and emoticons exist within Swedish blogs written in English. While emoticons were used more frequently by males than their female counterparts, abbreviations were used mostly by female bloggers. It also appeared that bloggers aged thirty or younger use emoticons and abbreviations most frequently, suggesting that their use was influenced quite significantly by the age of the blogger. The conclusion drawn from this is that if any eventual local Swedish variations of emoticons and abbreviations, they exist in Swedish blogs and not English ones, as those discovered in this study do not deviate from what already exist in blogs written by native speakers of English.

KEYWORDS: abbreviations, computer-mediated communication, corpus, emoticons, gender, linguistics, Netspeak
# Table of Contents

1. Introduction  
   1.1 Aim, research questions and scope  

2. Material and Method  
   2.1 Material  
   2.2 Method  
   2.3 Problems and limitations  

3. Previous Research  
   3.1 Netspeak  
   3.2 Emoticons  
   3.3 Abbreviations  

4. Results & Discussion  
   4.1 Emoticon use  
   4.2 Abbreviation use  
   4.3 The Gender and age factors in emoticon and abbreviation use  

5. Conclusion  

References
1. Introduction

Computer-Mediated Communication, henceforth CMC, refers to the interaction between humans that is mediated by computers or mobile phones, mainly in text form (Herring, 2007:1). CMC can be both synchronous, where participants of conversations are required to be online simultaneously in order to conduct conversation, and asynchronous, where the sender's message is stored on a site and can be read there (Herring, 2007:13). While nearly all synchronous CMC is interactive, asynchronous CMC can be both interactive and non-interactive (Maricic 2005:15). Computer-Mediated Discourse, henceforth referred to as CMD, is the term used when classifying CMC for research purposes, focusing on “online language and language use” (Herring 2007:1).

The blog, a common occurrence on the Internet, is one example of asynchronous CMC. After the bloggers have published their posts, anyone can access the posts at any time, provided that the blog is open to the public. Blogs can be both interactive, where the bloggers interact with the readers of the blog, for instance by answering questions asked by readers, as well as non-interactive. While the blogs used for the research in this paper belong to both kinds, over 90% of the blogs are non-interactive. Blogs can be both formal and informal, although nearly all blogs deal with issues that authors face in their everyday lives or topics they are passionate about. The majority of the blogs used in this study are informal. Some blogs feature more pictures than words; others are more focused on writing. The topics of blogs can vary from post to post, making them one of the most flexible means of communication on the Internet.

Crystal (2001) introduced the term Netspeak as a an umbrella term for language features frequently encountered on the Internet, stating that it is a kind of language use that features aspects that can only be found on the Internet (2001:18). For the sake of clarity, this study uses Netspeak as its main term for the linguistic phenomena discussed. The study focuses on two features of Netspeak, namely emoticons and abbreviations. These terms and their use, explained further in Section 3, are something nearly every, if not all, Internet users have witnessed and make use of themselves.

According to Danet & Herring (2007), online communication tends to be of a nature that is less formal and that focuses more on interaction between individuals than other ways of communication (2007:27). The majority of blogs on the Internet make use of emoticons and
abbreviations. This study investigates, whether these two features of Netspeak occur within blogs in English authored by Swedish individuals. This topic is of particular interest because language is a constantly evolving phenomenon and an interesting question is whether Netspeak has evolved differently in Sweden compared to native speakers of English.

While a great deal of studies on Netspeak and online communication exist, such as Baron (2008), Crystal (2001) and Herring (2007), few of them look at it through a Swedish perspective and to the knowledge of the author, none of them look at how Netspeak is used by Swedes within the blogging medium. Therefore, this investigation is of particular interest as it focuses on Swedish blogger’s material in order to find out if a 'Swedish' variety of Netspeak exists or if Swedish bloggers use the same abbreviations and emoticons when writing in English as native speakers of the language.

1.1 Aim, Research Questions and Scope

In this section, the aim, research questions and scope of this study are presented and explained. The aim of this study is to investigate the English used in blogs by Swedish individuals written on the internet. To fulfill this aim, the following research questions have been posed:

1. What kind of abbreviations and emoticons do Swedes use when writing blogs in English?
2. What difference, if any, is there in the use of abbreviations and emoticons between Swedish bloggers who write in English compared to native English bloggers’ use of these features as evidenced in other studies?
3. What role does the age and gender of the blogger play in their use of abbreviations and emoticons?

The investigation achieves this by examining a large quantity of blog posts and finding the most commonly used abbreviations and emoticons in these sources. Whether Swedes innovate emoticons and abbreviations or prefer to use those that already exist is the focus of this investigation. It specifically examines blogs written in English authored by Swedish individuals and attempts to explore if the Netspeak used is a 'Swedish' variety or if it is used in blogs written by native speakers of English as well.
The scope of this study is limited to blogs that are partly or entirely written in English by Swedish authors. The definition of Swedish in this case is someone who was has Swedish as their first language and who use it as their primary language of communication.

2. Material and Method

2.1 Material

The three following sections will discuss the material as well as the method used for this study. The material used as the primary source for this investigation consists of a collection of one hundred thousand words taken from fifteen personal blogs on various topics such as fashion, politics and gaming, found through searches conducted on the Internet. This was done in order to get as wide a range of different topics as possible. The material all comes from blogs written in English by Swedish authors during the last ten years. The bloggers’ online personas were scrutinized in order to make sure that they were indeed Swedish and their work was subsequently collected to be incorporated into the corpus. Five of the fifteen blogs were written by female authors and the remaining ten by males. Six of the bloggers were deduced to be above 30 years of age. This material was all collected by the author of this investigation and entered into a corpus named the BESC, short for Blogs in English written by Swedes Corpus, where corpus searches were performed in order to find trends and reoccurring instances of emoticons and abbreviations.

2.2 Method

The primary method used for material collection was searching the Internet, primarily blogging sites. One of the sites used for finding blogs fitting the criteria was Blogspot.se. However, the majority of the blogs on the site were written in Swedish. Therefore, additional searches were carried out to find more blogs in English on the site and other sites were found with the help of search engines, such as Google, in order to find bloggers who fit the criteria. The blogs that could be seen as fitting into the scope were examined more closely to ensure that the author indeed was Swedish one and thus fitting into the criteria given in the final paragraph of Section 1.1. Upon confirmation that this was the case, found by scrutinizing the author's public profile, blog posts were collected for the blog corpus. These profiles may not be an accurate representation of the actual identity of the blog’s author. However, in most of
the cases, other sources could be found which verified the identity of the blogger. This method was repeated until sufficient material had been found.

Since the majority of bloggers were male, the material has been normalized when looking at how the different genders use abbreviations and emoticons. Normalization is when you do not give results in absolute frequencies when comparing from sources of different sizes and instead rather make use of a common denominator which generally is a number close to the size of the smaller source (Lindquist 2010:42). For instance, when comparing two corpora of one respectively ten million words, results of the searches from the larger corpora are divided by ten to make the results more comparable to the smaller-sized corpus. Therefore, when comparing gender in Section 4.3, the results in this study have been normalized per ten thousand words.

The ages of the online personas of the bloggers were established through scrutinizing their online profiles. Some bloggers did not mention their exact age, but gave a decade-hint such as being in their ‘40s’ or ‘30s’, whereas others did not mention anything about how old they were. In these cases, should the online personas of the bloggers be present in encyclopedias when searched for on the Internet, the information about their age in these encyclopedias was used.

As was mentioned in Section 2.1, the material was compiled into the BESC, where abbreviations and emoticons were searched for using the concordance program AntConc. Then, the number of times they occurred was taken into consideration when researching just how frequently these phrases and words are used. Possible deviations have also been searched for with the help of the wildcard function * in AntConc, which finds everything in the corpus which uses the letters or symbols used in combination with the wildcard. This was done in order to find out whether Swedes emulate the already existing emoticons and abbreviations in English or if they create Swedish language-specific varieties of it.

Below follows Table 1, presenting the emoticons that were searched for in the corpus searches conducted in AntConc. In addition to these, the wildcard * has been used in combination with other components of emoticons such as ) and : in order to see if additional emoticons other than the ones given in the table existed. The emoticons have been examined in context in order to define exactly what kind of meaning they were supposed to convey.
Table 1. Emoticons searched for in the BESC

<table>
<thead>
<tr>
<th>Emoticon</th>
<th>Conventional interpretations</th>
</tr>
</thead>
<tbody>
<tr>
<td>:)</td>
<td>Happy, Smiling</td>
</tr>
<tr>
<td>:(</td>
<td>Angry, Sad</td>
</tr>
<tr>
<td>:)</td>
<td>Winking, Smirking</td>
</tr>
<tr>
<td>:D</td>
<td>Happy</td>
</tr>
<tr>
<td>T_T</td>
<td>Sad, Crying</td>
</tr>
<tr>
<td>:P</td>
<td>Playful</td>
</tr>
<tr>
<td>:/</td>
<td>Skeptical, Undecided</td>
</tr>
</tbody>
</table>

The emoticons in Table 1 were chosen due to their frequent occurrence on the Internet. The emoticons given in the table only consist of so-called ‘midget’ emoticons, lacking a nose, which is commonly visualized with a hyphen. However, when the wildcard searches found emoticons with a nose, they were subsequently incorporated into the results. Since AntConc only accepts .txt format, it was impossible to search for graphical or animated emoticons. Every graphical emoticon in the primary material was therefore converted to text-based emoticons for the purpose of this study.

Surprisingly, the most popular emoticon in the initial concordance search, with over one hundred tokens, was the winking emoticon. This seemed suspicious, as results of Huffaker & Calvert’s (2005 [www]) study had been different. Looking into the results in detail, it appeared that one blogger was a particularly avid user of said emoticon, being responsible for more than 90% of the hits. As a result, this particular blog was ignored in the search for the winking emoticon in order to make the results less skewed. Emoticons that were only present in one blog were also ignored to make sure that they were used by more than one individual.

Unfortunately, using the wildcard alongside letters to find new varieties of emoticons was not possible, as the results would show any word with the letters in it. For instance, searching for *D in order to find varieties of the :D emoticon would present every word with the letter in them. Therefore, it is possible that some varieties of the emoticons that incorporate letters have been missed during the searches.
Abbreviations are important for this investigation as well. The abbreviations that were searched for in the corpus are given in Table 2 below. Contrary to emoticons, abbreviations cannot be searched for with the wildcard * in AntConc, since the irrelevant results of those searches would far outnumber the relevant ones, as every word containing the letters searched for alongside the wildcard would have been shown, making such a search highly complicated as well as incredibly inefficient.

Table 2. Abbreviations searched for in the BESC

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Conventional interpretations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ttyl</td>
<td>Talk to you later</td>
</tr>
<tr>
<td>lol</td>
<td>Laughing out loud</td>
</tr>
<tr>
<td>omg</td>
<td>Oh my God</td>
</tr>
<tr>
<td>asap</td>
<td>As soon as possible</td>
</tr>
<tr>
<td>brb</td>
<td>Be right back</td>
</tr>
<tr>
<td>cya</td>
<td>See you around</td>
</tr>
</tbody>
</table>

The abbreviations here consist of both acronyms, where the abbreviation is pronounced as a single word, such as cya, and initialisms, where every letter in the abbreviation is pronounced separately like initials of a name, such as brb. However, some abbreviations, such as asap, can be used as both. These abbreviations were chosen mainly because they are informal abbreviations and not formal ones, such as NATO and the EU. Both upper-case and lower-case abbreviations have been searched for and taken into account regarding the results of the search.

Compared to emoticons, which are used relatively frequently in BESC, abbreviations were much more uncommon. Due to this lack of instances in AntConc, every hit in the concordance searches was recorded in order to have as much material as possible to analyze. However, the abbreviations searched for in the corpus are those which one would normally find within informal communication. Therefore, abbreviations of institutions, organizations and regions such as the EU, UN and ESEA are not recorded since they appear in formal contexts. Had they been incorporated into the searches, the number of abbreviations found would likely have increased significantly.
2.3 Problems & limitations

A limitation of this investigation is that the primary source material for this study is somewhat small, only amounting to one hundred thousand words, although it is relatively large for being collected by a single individual during a couple of weeks. However, the many different topics in the material may eventually result in unreliable results due to the overrepresentation of one topic compared to others. For instance, logs about politics rarely tend to make use of emoticons and informal abbreviations. Additionally, since the bloggers had to be scrutinized to make sure that they fit the criteria, all material collection had to be done by the author personally, making it very time-consuming, which was one of the reasons for the size of the corpus not being larger than it is.

The study is limited to bloggers who are Swedish and writing in English. Therefore, the material comes from a rather small source compared to all Swedish bloggers that write in English. The possible issue with this can be that the instances of emoticons and abbreviations found within this source material may eventually turn out not to be an adequate representation of what Swedes use. In addition, the gender-division, with two thirds of the bloggers in this study being male, may skew the results when comparing the two genders in the results.

At the beginning of the investigation, the major issue was finding appropriate material due to what seemed to be the decline of Swedish bloggers writing in English. However, this turned out to be an effect of looking in the wrong places, as additional searches on mediums other than the one where material was first searched for would show. There was plenty of material to be found all across the Internet, provided one looks hard enough.

One issue that arose was the ethical issue of whether or not the bloggers were made anonymous in the corpus and investigation. The aspects of anonymity that were considered stem from the guidelines suggested by the Association of Internet Researchers (2001) in Maricic (2005:72). All of the material of the BESC is publicly available online for free, including the names of the bloggers, and therefore the authors have not been made anonymous in the corpus itself. However, in the examples given in this study in Section 4, neither the name nor the gender of the blogger in question is given to allow the bloggers some degree of anonymity. In addition to this, the only individuals able to access the corpus are those authorized to use the learning platform of the course in which this study was carried out and therefore only students and professors at Linnaeus University are able to access the
corpus file directly. This provides further anonymity for the bloggers whose material has been used for this investigation.

3. Previous Research

3.1 Netspeak

This section presents previous research regarding Netspeak. As mentioned in Section 1, Netspeak is a term coined by Crystal in his book *Language and the Internet*, first published in 2001. The history of this type of language features that occur predominantly within electronic communication is quite a short one. In fact, it technically began with the introduction of the very first recorded conversation in email correspondence which is said to have taken place early in the 1960s (Thurlow et al, 2004:14). However, its rise in usage did not occur until computers were machines which could be owned by most people in the developed countries. Therefore, Netspeak has only flourished during the last few decades, which is not surprising considering the increase of these machines three decades after the first recorded emails. By then, many of the households in the western world owned a personal computer (ibid).

One factor that may have contributed to this explosive spread of computers might have been the introduction of the World Wide Web (www) by Tim Berners-Lee in the early 1990s, described by Baron (2008:13) as “a collection of software tools and protocols that make it relatively easy for computers to communicate across the Internet”. Naturally, this new addition to the Internet made it much easier for individuals to communicate with one another and therefore, since communication is required for a language to flourish and evolve, the implementation of the World Wide Web is certain to have affected the spread and development of Netspeak and its features.

According to Greiffenstern (2010), Crystal gives a presentation and analysis of these new features found in online communication in his aforementioned book (Greiffenstern, 2010:22). Therefore, it is a good starting point when looking at different features of Netspeak. However, it is important to consider the effects of time on Crystal’s (2001) reflections. For instance, he mentions how individuals used computer terms in order to make it seem as if what they were saying had “a cool cutting-edge” (2001:19). Among the examples listed slightly after this observation is the verb “multitasking” (ibid). Said today, this would most likely not make conversational partners bat an eyelash as it has been integrated into our everyday language.
Semantic change, which is the change of meaning of a word, affects Netspeak just like any other language. There are also different kinds of semantic change. For instance, *Multitasking* was changed from only referring to tasks on computers to other tasks as well, effectively broadening its meaning. Perhaps one of these two meanings will eventually oust the other, as is mentioned by Atchison in her “cuckoo” metaphor (2013:131), where a new meaning (*multitasking* referring to non-computer tasks in this case) usually joins the other meaning and the two co-exist. Of course, it is not necessary for the new meaning to overtake the old one, as several meanings are more than capable of co-existing, as is the case with words like *hand*. Several other instances of semantic change can be found, e.g. *flaming*, which now not only refers to something that is on fire, but also deliberately posting “derogatory messages and *ad hominem* attacks targeting individuals or groups” (Maricic, 2005:60).

In his study, Crystal also mentions how Netspeak allows Internet users to vary their writing in order to mimic emphasis in talk; for instance, a common feature is using “all capitals for ‘shouting’” (2001:35), such as when a participant in a conversation online writes ‘LISTEN HERE’ in order to get their point across. Repeating letters, punctuation marks, exclamation marks and question marks is also a feature of Netspeak, which serves to emphasize some form of a phrase. The same goes for alternating capitalization in a word. For instance, ‘WhAAAAAT??????’ emphasizes the final syllables of the word as well as notifies that the writer is surprised by whatever they are commenting on.

Other examples of linguistic features in order to emphasize something that are given by Crystal are spacing between letters to appear as speaking clearly (r i g h t n o w) or using asterisks in order to signal that a certain word is emphasized more than others, such as “the *real* answer” (2001:35). However, these are far from the only ways a writer can vary emphasis in their statements online, and the variation seems to only depend on the creativity of the writer. For instance, capitalizing a word in a sentence can be used to signal the importance of that word instead of using asterisks.

### 3.2 Emoticons

This section presents previous research on emoticons, a feature of Netspeak. Like CMC at large, Netspeak features such as emoticons and abbreviations are relatively new phenomena. Emoticons are symbols that can be used in order to mimic facial expressions. The use of an emoticon a.k.a. a smiley was proposed by Scott Fahlman in 1982 when he suggested that :-)
should be used as a joke marker while `:-(' would mark that something was not a joke (Microsoft [www]). Mike Jones, the researcher behind the effort to locate the origin of emoticons, also claims that the original emoticons “have been in widespread use since the early ‘80s” (ibid), suggesting that they filled functions which online communication were lacking, such as the ability to mark whether something was to be taken as a joke or not and compensating for the lack of body language in online communication.

Derks et al. (2007) suggest that this function which is filled by emoticons in Netspeak is the same as that of non-verbal signals in face-to-face communication, which are as follows: “providing information, regulating interaction; and expressing intimacy” (2007:843). Furthermore, they also state that because of this, emoticons might “add a paralinguistic component” to messages, providing something more than what is already written (ibid). In other words, emoticons are a feature of Netspeak that fulfill many different functions. In their study, Derks et al. also discovered that their subjects utilized emoticons more often in contexts that were “socio-emotional” compared to contexts that were more oriented towards accomplishing tasks (2007:846), which points to emoticons being a tool used not to accomplish tasks, but to connect with one's conversational partner and convey feelings. They also found that subjects made use of emoticons regardless of positive or negative context, suggesting that emoticons are not bound to a certain conversational mood. Another conclusion drawn by Derks et al. was that the emoticons used often mimicked the mood of the social contexts. According to the authors, it appeared as if the subjects of the study used emoticons denoting positivity and negativity chiefly in their corresponding contexts (ibid). However, it is also interesting to note that the study found that subjects would use the most emoticons in “negative socio-emotional contexts” (ibid), hinting that emoticons are a powerful tool for requesting sympathy.

Huffaker & Calvert’s (2005 [www]) study on the language used in blogs written by teenagers concluded that although previous research had shown that females made use of emoticons more often than males did in IM (Instant Messaging) programs such as IRC, no such evidence could be found regarding blogs (ibid). In fact, the results of their study suggested that the use of emoticons was more prevalent among males than females. In addition, the study observed that text-based and graphical emoticons were quite evenly divided, making up 51 and 49% of the total number of emoticons respectively. In the cases of the more common emoticons, which indicated happiness, playfulness, flirtatiousness or winking, bloggers tended to create
them using ASCII characters. On the other hand, emoticons for indicating sleepiness or anger, which were lesser known emoticons, were nearly all graphical. The study suggested that the reason for this was that the replication of those emoticons in text form would be a difficult task (ibid).

Vasic et al. (2010 [www]) carried out a study amongst university students, which showed that although more senior students than freshmen considered emoticons not to have a place in formal communication, there was no difference regarding usage of emoticons in informal conversations. The conclusion drawn from this was that the use of emoticons had become something that gave a human touch to informal online communication (ibid), similar to the observations of Derks et al. (2007).

Dresner & Herring (2010:251) argue that emoticons are nearly universally seen as a way of indicating emotions without saying anything. However, they also claim that emoticons are a conveyor of the speech act that an individual performs by producing an utterance, confirming the intention of what an individual has written (2010:256). Therefore, emoticons are able to work as indicators of illocutionary force as well, which can be seen when a smiley is accompanying a complaint, effectively downtoning the face-threatening force of the complaint. One’s face in pragmatics is one’s public image of self (Yule, 2010:135). Dresner & Herring argue that emoticons are not only ways of expressing emotion. Instead, they are indicators of the way in which one should interpret a string of words (2010:258). The authors conclude that there are three aspects of emoticons. Firstly, they are used to express emotions, which directly correspond to facial expressions. Secondly, they also have non-emotional meanings, which conventionally correspond to a facial expression such as the winking emoticon for jokes (since joking is not an emotion). Thirdly and finally, they can be ‘illocutionary force indicators’ which do not have a conventional facial expression to correspond to, such as the aforementioned use of the smiling emoticon when softening complaints (Dresner & Herring, 2010:263).

3.3 Abbreviations

Previous research on another prominent feature of Netspeak, namely abbreviations, is discussed in this section. Abbreviations are shortened forms of words or phrases and Netspeak, especially synchronous chatting, is heavily laden with abbreviations. This is partly due to their time-saving nature and ability to compress a long phrase quickly which helps with
keeping the conversational floor, making them occur very frequently within synchronous online communication. Initialisms are a subgroup of abbreviations where every letter in the abbreviations is pronounced separately, such as *brb* (pronounced BEE-ARR-BEE). Acronyms are another subgroup, where the abbreviation is pronounced as a single word, such as *lol*. Abbreviations, contrary to emoticons, occurred before the introduction of computers. In fact, they were used as early as the Middle Ages (Baron 2002:97). Fragmentation occurs quite often within some of these abbreviations. For instance, the noun phrase 'I will' (commonly shortened to *I'll*) often seems to fall victim to ellipsis, or intentional omission. Consequently, the abbreviation used for the phrase 'I'll be right there' is simply *'BRT'* , where the personal pronoun and auxiliary verb are not mentioned at all.

In their study of university students, Vasic et al. (2010) reported that the use of abbreviations was increasing when their subjects were communicating informally, as those who were studying on their 1st semester using fewer abbreviations than those on their 7th semester. Additionally, the investigation also indicated that acronym use is reduced in formal communication when comparing the 1st semester students to the 7th semester ones. This suggests that abbreviation use is mainly relegated to informal communication with the exception of the abbreviation of names of “public institutions” such as the EU, NATO and other organizations which are used in formal communication as well (ibid).

Greiffenstern (2010) regards acronyms as being one of “the most distinctive and productive classes of abbreviations” (2010:46). She also noted that Netspeak is where abbreviations are used most frequently compared to other ways of communicating. This suggests that they are used for more reasons than to be time-saving and make conversation flow faster through rapid message-producing. Greiffenstern claims that when they make use of abbreviations, individuals appear to use them as a way of expressing a particular attitude, or identifying with some form of group (ibid). She argues that the sheer number of acronyms of different forms might make it hard for non-frequent users of the Internet to comprehend some Netspeak users, creating a division between casual and more experienced netizens when browsing the web (ibid).

Danet & Herring (2007) state that respectable evidence exists that features of Netspeak, among those abbreviations and emoticons, are used in languages other than English (2007:27). They also suggest that the features that are shared possibly have one and the same origin, *viz.* the United States. This would mean that these features had spread from the U.S.
across the globe. However, they also acknowledge that there are some local deviations that are likely to have arisen due to “technical constraints of the medium” (ibid), which refers to the telephone keypad found on phones in that particular context. Therefore, although the majority of abbreviations or other features of Netspeak appear to be the same regardless of language, the existence of exceptions is an ever-present possibility. However, it should be noted that in the past few years, the introduction of smart phones with entire touch-keyboards in them may have caused the reduction of these regional variations of abbreviations and emoticons, although differences naturally exist between users of the Roman alphabet and others, such as Hangul, the Korean alphabet.

Merchant (2001) carried out a study in which he commented that using abbreviations was quite a common occurrence in informal e-mails, in text messages on telephones as well as in synchronous chatting. He also noted that every individual he interviewed for his study displayed some form of knowledge about them (2001:302). In addition, he categorized abbreviations into four types. The first category listed was emoticons. However, other researchers, such as Crystal, treat them as a category of their own. This is because emoticons are not often used as a way of abbreviating a message. Instead, they are used to convey what was explained by Dresner & Herring (2010) in Section 3.2. Secondly, there was the case of “simple abbreviations” where the letters in the beginning of every word in a phrase are used to form the abbreviation, such as the ‘asap’ abbreviation. The third category was the combining of numbers and letters to form a “phonetic rendering of the message”, for instance writing some1 or NE1 instead of the more traditional forms of someone and anyone. Crystal (2001) comments that these seem almost like rebuses, since the sound of the numbers when pronounced “acts as a syllable of a word” (2001:86). Phonetic spelling, such as writing gonna instead of going to, made up the fourth and final category in Merchant's investigation. However, other researchers, such as Greiffenstern, do not view phonetic spelling as a way of abbreviating but instead a way of making the tone of messages emulate that of speech (2010:49). Merchant also noticed that sometimes individuals were creative and were able to mix the categories with one another, suggesting that the boundaries between these four categories are not very strict (2001:302).

To summarize Section 3, Netspeak is a broad and general definition of features of language frequently occurring online. Netspeak is affected by semantic change, changing the meaning of certain words. Netspeak can also be used to frame one's words differently in order to
emphasize something. Emoticons are used in order to express some form of intimacy and to make messages more personal. Emoticons fulfill, among others, the function which non-linguistic cues such as mimicry are responsible for in face-to-face communication. Emoticons have different functions and can express emotions as well as non-emotional meanings while also carrying illocutionary force. Abbreviations seem to be quite frequent in synchronous and less frequent in asynchronous CMC due to their time-saving nature, which is needed more in synchronous communication as the writers in asynchronous communication tend to have plenty of time to edit what they write before they post it. While different categories of abbreviations exist, their boundaries are blurred and they can therefore easily be mixed with one another. In the following sections, the results of the different searches for emoticons and abbreviations in the corpus are presented and discussed in relation to previous studies.

4. Results & Discussion

The concordance hits of emoticons and abbreviations found in the searches carried out using AntConc have been put into different tables and analyzed in this section. The results were also compared with previous studies on emoticons and abbreviations and how they relate to or differ from one another. The first subsection discusses emoticons and the second abbreviations. Gender and age is discussed in the third and final subsection, where the possible effects these factors may have on a blogger’s use of emoticons and abbreviations are presented.

4.1 Emoticon Use

All of the emoticons presented in Table 1 were present in the blogs in the primary material for the present study with the exception of :/. The results following the search for emoticons can be seen in Table 3 below. These results are interesting because compared to Huffaker & Calvert's study (2005 [www]), where emoticons denoting happiness made up 53% of the emoticons observed, the percentage of these emoticons in the present study is much higher, calculated to be 79.24%. Of course, as a product of this, sadness is not nearly as prevalent in these blogs as it was in the blogs analyzed by Huffaker & Calvert (30%), since the emoticons suggesting this only make up 8.49% of all emoticons in the corpus. This suggests that blogs written by Swedes do not make use of sad emoticons as often as those authored by native speakers of English.
Table 3. Emoticons found in the BESC

<table>
<thead>
<tr>
<th>Emoticon</th>
<th>Meaning in context</th>
<th>Concordance Hits</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>:)/=)</td>
<td>Happy, Smiling</td>
<td>42</td>
<td>39.62</td>
</tr>
<tr>
<td>:D</td>
<td>Happy, Laughing</td>
<td>29</td>
<td>27.36</td>
</tr>
<tr>
<td>^^/^_^</td>
<td>Happy</td>
<td>13</td>
<td>12.26</td>
</tr>
<tr>
<td>:)</td>
<td>Joking, Sarcastic, Hinting</td>
<td>11</td>
<td>10.38</td>
</tr>
<tr>
<td>:(</td>
<td>Sad, Angry</td>
<td>7</td>
<td>6.60</td>
</tr>
<tr>
<td>:P</td>
<td>Silly</td>
<td>2</td>
<td>1.89</td>
</tr>
<tr>
<td>T_T</td>
<td>Sad</td>
<td>2</td>
<td>1.89</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>106</td>
<td>100</td>
</tr>
</tbody>
</table>

The number of winking emoticons differs from Huffaker & Calvert's study, where only 5% of the emoticons used were of that kind. Even when removing 90% of the instances produced by one blogger, they managed to more than double the percentage that was presented by Huffaker & Calvert (2005 [www]), making up about 11.2% of the total number of emoticons. This would suggest that the emoticon has been subject to semantic change during the last nine years, changing its meaning from flirtatious to more of an indicator of a joke, which would explain its increase in frequency, as joking is far more common than flirting. In fact, the majority of the hits in the present study are jokes, such as when one blogger describes something as:

(1) nothing out of the ordinary! ;)  

Here, the blogger is referring to living in a team house focused on gaming, something quite unusual. Another way of using this emoticon was to indicate sarcasm, such as when another blogger writes about teams from another part of the world:

(2) they were so much superior ;)  

This was written after the blogger’s team had thoroughly beaten said teams in a tournament. This clearly indicates that they were being sarcastic when writing this, perhaps wanting to mock individuals who were claiming that such was the case.
However, while not identical in the percentages for happy and sad emoticons, the general results of the investigations remain relatively similar to those in Huffaker & Calvert (2005): Emoticons are mainly used to describe happiness and happy emoticons are primarily used in a positive context, agreeing with what Derks et al. (2007) suggested. One can but wonder if Scott Fahlman knew just how widespread his suggestion for using the smiley would become when he forwarded it in 1982. The reason for such a large percentage is unclear. However, the general mood for the majority of the blog posts was a positive one, which may have contributed to more emoticons denoting happiness being used.

When conducting searches using the wildcard * in AntConc, no other variety of emoticon appeared. Instead, the only tokens that were deemed relevant and were added to Table 3 were hits where an emoticon already selected for analysis had not been found by the initial search due to additional space between the semicolon and whatever component that was used to form the emoticon. Therefore, the results of this search appear to comply with the suggestion of Danet & Herring (2007) that features of Netspeak are shared regardless of language.

Dresner & Herring (2010) suggested three different functions of emoticons, which were explained in further detail in Section 3.2. In the BESC, all three of the functions were found. The direct relation to facial expression seemed to be the most common one, which is not too surprising, since blogs often relate to issues bloggers are facing and the emotions they are feeling when confronting them. Nevertheless, the non-emotional meaning was also there, especially regarding marking jokes with the winking emoticon. The most interesting finding, however, was that the illocutionary force mentioned by the authors was quite apparent in the results as well. Sometimes this force even affected other emoticons such as when one blogger was told by his friend:

(3) you receive the leftovers :(<:D

The laughing emoticon is used to soften the would-be complaint of the sad emoticon referring to the direct facial expression in this particular instance. It is used since the post refers to a game the blogger and his friends were playing and was therefore not a serious issue. Another instance of where the emoticon has an illocutionary function is where a blogger, after losing a game, commented:

(4) you cant [sic] always win :)
Naturally, losing a game is nothing which would make anyone smile, but the smiling emoticon works like the laughing emoticon in the previous example, softening the statement, making it seem as if the blogger is recuperating from the loss quite well in this case. This sentence would have been interpreted quite differently had the author chosen other emoticons such as T_T, which would suggest that the loss saddened the blogger.

Illocutionary force often works to tone down potential face-threatening acts of one’s statements. For instance, in Example (4), the complaint that the initial emoticon embodies is a face-threatening act towards the author’s friend, that his gift was insufficient. The author realizes this and therefore decides to make the second emoticon work as a way of making sure it is not taken as such by suggesting that they are only joking, downtoning the face-threatening act.

4.2 Abbreviation Use

The second of the features of Netspeak examined in this study are abbreviations. Within this subsection, results from the concordance searches are given. Table 4 below features two subcategories of abbreviations, namely acronyms and initialisms, which were explained more in detail in Section 3.3. The results are analyzed more closely in the paragraphs following it.

Table 4. Abbreviations found in the BESC

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning in context</th>
<th>Concordance Hits</th>
<th>Type</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>lol</td>
<td>Laughing out loud</td>
<td>9</td>
<td>Both</td>
<td>75</td>
</tr>
<tr>
<td>asap</td>
<td>As soon as possible</td>
<td>1</td>
<td>Both</td>
<td>8.33</td>
</tr>
<tr>
<td>cya</td>
<td>See you</td>
<td>1</td>
<td>Acronym</td>
<td>8.33</td>
</tr>
<tr>
<td>ftw</td>
<td>For the win</td>
<td>1</td>
<td>Initialism</td>
<td>8.33</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>12</td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

The hits are far from as numerous as those in the search for emoticons. All of these tokens include both the capitalized and non-capitalized spelling of the abbreviations. The lack of abbreviations may be a result of blogs being asynchronous, and therefore the time saved when writing would not be of as much importance as when one is engaging in a synchronous conversation with another individual. For instance, the unabbreviated forms See you and as
soon as possible had more instances than their abbreviated counterpart. This provides further evidence that the time-saving aspect of abbreviations is not a major influence for bloggers using them. A likely reason for this is that bloggers have sufficient time to edit and are therefore not as pressured for time as individuals participating in synchronous online communication.

One of the most common acronyms online is lol, the abbreviation of the phrase 'laughing out loud', recognized as a marker of the hilarity of something. This was one of the abbreviations that did not even have an unabbreviated form present in the corpus. Perhaps this suggests that the abbreviation has been conventionalized can only be found in its abbreviated form. In fact, lol has become so conventionalized that it can be even be found in informal spoken interaction. However, it is important to note that this abbreviation mainly is not used as a substitute for the actual phrase. One could argue that through abbreviating the phrase, lol has broadened its meaning from a phrase to a speech act, since it is able to refer to more than simply laughing, such as finding something funny. This is another way in which semantic change might affect Netspeak features, i.e. a change of syntactic function.

Despite being in a minority in the BESC, abbreviations seem to share one thing with emoticons, which is what they are used to convey. If Huffaker & Calvert (2005) had included abbreviations, it would have been easy to compare the results without looking for similarities between emoticons and abbreviations. However, they did not, and therefore it is difficult to compare abbreviations directly, although when comparing to emoticons, both features seem to be used for the same effect, which is mainly for indicating happiness or enjoyment. As was mentioned earlier, lol was the most common abbreviation and although laughing does not necessarily have to indicate happiness, most of the contexts were with something that the blogger found enjoyable, such as when one blogger is:

(5) feeling a tad sexier lol.

The blogger in Example (5) seems to like feeling attractive and is therefore happy and enjoying it. In fact, while neutral abbreviations were found in the corpus, no abbreviations with a negative meaning such as wtf were present.

Comparing these results to Merchant’s (2001) study, three of the four categories that he lists abbreviations in are present. These are: emoticons, abbreviations where the initial letters are
used as shorthand, and phonetic spelling. However, phonetic spelling is not present in the tables due to being considered as not being abbreviating but more a way of making a message emulate speech. No instances of the combination of letters and numbers were found. Therefore, one could assume that this final category tends to occur mainly in synchronous communication, which blogs are not. Nevertheless, emoticons, using initial letters and phonetic spelling can be seen in both asynchronous and synchronous CMC. Perhaps this means that abbreviations which combine letters and numbers are more suited for synchronous modes of CMC when compared to the other categories. However, the reason for this is unknown and warrants additional research.

Vasic et al. (2010) suggested that in more formal contexts, abbreviations tend to be used only when referring to organizations and public institutions and not for shortening phrases or the like. This held true for the results of this study as well. In the blogs dealing with more formal issues, abbreviations were scarce and the most common ones were those who referred to official organizations such as NATO and the EU. Therefore, it seems logical to deduce that the level of formality in a blog plays an important role for just what kind of abbreviations are used in it.

Due to the incredibly small turnout of the concordance searches, deducing whether Swedes use abbreviations that differ from the rest of the world is a difficult task. However, judging from results given so far, it would appear as if that is not the case, much like the case of emoticons. What seems to be the most likely situation is that if Swedish individuals make use of any possible local variations, they exist within the Swedish language and not in English. However, since all of the primary source material for this study is in English, it is impossible to find that out in this study. Additional research needs to be conducted about blogs in Swedish in order to find an answer for this particular hypothesis.

4.3 The Gender and age factors in emoticon and abbreviation use

As mentioned in Section 3.2, previous studies have observed that while females utilize emoticons more in chat modes like IM programs, blogs written by males seem to contain a higher density of emoticons (Huffaker & Calvert 2005[www]). With previous research in mind, the most expected result would be that the blogs by male authors would be the ones that make use of the most emoticons. The results of this study comply with this. Even when normalized to per ten thousand words, males significantly outnumbered females in regards to
emoticon usage. One interesting thing to note about females regarding emoticons is that with the exception of one individual, not one female used the same emoticons as other bloggers. The females who made use of emoticons in their blog posts tended to use different emoticons than the others and therefore, these emoticons are not counted as valid for the study due to only one blogger making use of them. For instance, one female blogger made use of the emoticon xx, which denotes kisses, as a way of ending some of her posts. However, this may be a product of insufficient material written by females and if more blogs by females had been sampled, the emoticon may have appeared in more than one blog. As was mentioned in Section 2.1, the reason for the small number of female blogs in this study is due to the fact that the majority of the primary source material used for this investigation was written by males.

In contrast, when it comes to abbreviations, the gender-division is the other way around. Even without any form of normalizing, females contribute for over half of the instances of abbreviations found in the corpus. A possible reason for this is provided by Greiffenstern's (2010) study as was discussed in Section 3.3. Females may make use of abbreviations more often than males in order to identify with some group. It is possible that females consider themselves to be non-typical users of abbreviations and use them as a way of claiming that they in fact make use of abbreviations often. Had Huffaker & Calvert (2005) looked at abbreviations more closely in their study, it would have been very interesting to compare the results. Further research of this would be useful for drawing a more certain conclusion. In addition, further collection of material, in particular blogs written by females, would be beneficial since the amount of primary material for this study is not evenly distributed gender-wise, as was mentioned in the previous paragraph.

In terms of age, there seems to be a group where emoticons and abbreviations are used most frequently. In the corpus, emoticons are used relatively often up until around thirty years of age, where their use decreased significantly. Only 7.5% of the emoticons in the corpus are used by bloggers who are thirty years or older. However, this is mostly true for males. The only female who was over thirty years old was in fact the source for the majority of emoticons and abbreviations for the female blogs in the primary source material. One likely reason for this is the strategy of identification suggested by Greiffenstern (2010). The female blogger is quite fond of computer games. However, one could argue that women over thirty years of age are not the typical age group that comes to mind when thinking of computer gamers.
Therefore, perhaps in order to identify with this kind of group, the blogger makes use of emoticons and abbreviations. The same can be said for the sole male blogger over thirty who employs abbreviations. This might be to identify with this target group, in this case people who are interested in finding things out about the world through mediums on the Internet, in this case Google Earth. With the exception of these two blogs, the only users of emoticons and abbreviations in the corpus are below thirty years of age.

5. Conclusion

The aim of this thesis was to investigate what emoticons and abbreviations Swedes use when writing blogs in English. Emoticons were used the most frequently, although some abbreviations were also present. None of the instances of emoticons and abbreviations found were exclusive to Swedish bloggers, as they could be found in blogs written by native speakers of English as well. Age and gender played quite a significant role for the bloggers’ use of Netspeak, as no males and only one female over thirty years of age made use of both emoticons and abbreviations. Male bloggers tended to use more emoticons whereas female bloggers utilized abbreviations more frequently. With these conclusions drawn, the aim of this essay has been fulfilled.

Just like the results of Huffaker & Calvert’s (2005 [www]) study, emoticons were used mainly to convey happiness. However, the present study had a significantly higher percentage of emoticons denoting happiness present and subsequently a lower percentage of sad emoticons. The same held true for abbreviations, with lol being the most frequently used abbreviation in the BESC.

Bloggers did not use any emoticons and abbreviations when writing in English that were specific to Swedish individuals and it seems more likely that eventual deviations would be present within the native language of the bloggers. It also seems logical that young individuals were more likely to use emoticons and abbreviations. However, one result which was surprising was the use of abbreviations. Not only were they only used twelve times in one hundred thousand words, females were also the ones using abbreviations most frequently. However, females making use of abbreviations more often than males, conforms to the suspicions of Greiffenstern (2010).
The method as well as the materials used both have their strengths and weaknesses. In terms of the method, the personal collection of blog posts ensured that all of the material comes from bloggers fitting the criteria. However, the collection was a very arduous and time-inefficient process. While it may have been accelerated through using some form of a bot for collecting material, that collection method may have resulted in bloggers who are not considered Swedish by the criteria being present in the material. One of the weaknesses of the concordance program used when searching for results is that the searches conducted had to be spaced specifically. Otherwise, some instances where a blogger had used an extra space may have been left out in the results. However, its strengths, which mainly refers to the time saved and number of results it was able to process more than make up for this.

Taking additional time to collect material is one of the ways in which one could have proceeded with this investigation differently. However, considering the time-constraints for the study, which corresponds to one university semester, the material collected was of a reasonable size. Another way in which the study could alternatively have been carried out is related to the concordance program. As was mentioned in Section 2.2, using the wildcard * to search for abbreviations would be a very inefficient method due to the high number of irrelevant results. However, looking at some other aspects, such as capitalizing all the letters in a word to emphasize it, as suggested by Crystal (2001), might have been interesting as the concordance program is able to match search hits after case. Eventual issues with this method would be that one would have to specifically search for the word, and considering the sheer number of words in the corpus, that would have been quite a monotonous and inefficient endeavor.

This study can be of use to individuals who are interested in emoticon and abbreviation use by Swedes in blogs and linguists whose main interest is CMC as well as those who want a basic example of how to make use of a corpus to analyze linguistic data. It may also serve as a basis for comparing results with a similar study carried out with blogs written by Swedes in their native language, which is one of the possible areas where further research can be carried out related to this study. Other areas of additional research may be examining the use of abbreviations and emoticons in blogs written in English by individuals from different countries such as the other Nordic countries and comparing the results found there with those in this study.
References


Greiffenstern, Sandra. 2010. *The influence of computers, the Internet and computer-mediated communication on everyday English*. Berlin: Logos Verlag Berlin GmbH.


