THE EFFECT OF TOPIC ON WORD FORMATION AND THE FREQUENCY OF NEOLOGISM USE ON INTERNET FORUMS

Master's thesis
Juha Matti Mikael Kalima

University of Jyväskylä
Department of Languages
English
April 2013

JYVÄSKYLÄN YLIOPISTO

Tiedekunta – Faculty Humanistinen tiedekunta	Laitos – Department Kielten laitos
Tekijä – Author Juha Matti Mikael Kalima	
Työn nimi – Title THE EFFECT OF TOPIC ON WORD FORMATION AND THE FREQUENCY OF NEOLOGISM USE ON INTERNET FORUMS	
Oppiaine – Subject Englanti	Työn laji – Level Pro gradu-tutkielma
Aika – Month and year Huhtikuu 2013	Sivumäärä – Number of pages 84 sivua + 2 liitettä

Tiivistelmä – Abstract

Internetin englanninkieliset keskustelupalstat ovat merkittävä osa nykyaikaista kielenkäyttöä, ja niiden tutkimus tulisi nähdä osana ihmisten kielenkäytön tutkimuksen kokonaisuutta. Keskustelupalstoilla sekä luodaan, että käytetään myös paljon uudissanoja, joista moni siirtyy käytettäväksi myös keskustelupalstojen ulkopuolelle.

Tässä tutkielmassa selvitettiin, millä eri sananmuodostusprosesseilla keskustelupalstoilla käytetyt uudissanat on luotu ja vaikuttaako englanninkielisen keskustelupalstan aihe palstalla käytettäviin uudissanoihin. Tutkielman pohjana oli englanninkielen sananmuodostuksen tutkimus sekä tietokonevälitteisen kommunikaation tutkimus. Tutkielma toteutettiin vertailemalla saman keskustelupalstan kahta eri keskustelualuetta, joista toinen oli peliaiheinen ja toinen yleisen keskustelun alue, sekä niillä käytettyjä uudissanoja.

Peliaiheiselta keskustelualueelta löydettiin yli kolme kertaa enemmän uudissanoja kuin yleiseltä keskustelualueelta, siitäkin huolimatta, että yleisellä keskustelualueella käytetty sanavarasto oli laajempi. Uudissanojen käyttöaste oli myös lähes kolme kertaa korkeampi peliaiheisella keskustelualueella, sillä uudissanaa käytettiin yleisellä keskustelualueella keskimäärin 1,69 kertaa, kun taas peliaiheisella keskustelualueella uudissanaa käytettiin keskimäärin 4,02 kertaa. Sanojen erilainen lyhentäminen oli yleisin sananmuodostustapa molemmilla keskustelualueilla ja yli puolet uudissanoista molemmilla alueilla oli muodostettu lyhentämällä.

Tutkielman tulokset viittaavat siihen, että spesifit keskusteluaiheet kannustavat käyttäjiä luomaan ja käyttämään enemmän uudissanoja kuin yleisemmät keskusteluaiheet.

Asiasanat – Keywords: word formation, internet, CMC

Säilytyspaikka – Depository: Kielten Laitos

Muita tietoja – Additional information: -

1 INTRODUCTION 5

2 THEORETICAL BACKGROUND 7
2.1 Word formation 8
2.1.1 Defining the terminology 8
2.1.2 The formation of new words in general 12
2.1.3 Previous research on the internet and on the nature of the forums 20
2.2 Computer Mediated Communication (CMC) 21
2.3 Language of the internet 26
2.4 Linguistic features of the language of gamers 30

3 DATA ACQUISITION 32
3.1 Compiling the research materials 32
3.2 What is a neologism? 35

4 ANALYSIS OF THE DATA 37
4.1 Research questions for the study 37
4.2 Tools used with the data analysis 39
4.3 Methods of analyzing the data 40
4.4 Challenges during the analysis process 44

5 RESULTS 48
5.1 Findings of the analysis on the data gathered on the gaming subforum 48
5.1.1 The numbers in the gaming subforum 49
5.1.2 Findings on the gaming subforum 52
5.2 Findings on the data gathered on the general discussion subforum 54
5.2.1 The numbers in the general discussion subforum 54
5.2.2 Findings on the general discussion subforum 57
5.3 Differences between the two sub forums 58
5.4 Further points of interest in the forum language use 60
5.4.1 Nuances in meaning 60
5.4.2 Posting practices 66
5.4.3 The bits and pieces used to make a word 69

6 CONCLUSION 77

7 BIBLIOGRAPHY 83

1 INTRODUCTION

English has been and still remains the lingua franca of the internet and the multitude of people using the language online on a daily basis contributes to how the language changes through time (Crystal 1997, 105). As the number of internet users grows, so does computer mediated communication (CMC) increase in use. Instant messaging services are becoming more and more popular, internet message boards, aka. forums, are seeing more traffic, blogs are gathering increasing numbers of followers and social networking sites like Facebook and Twitter are becoming the number one pastime among young people.

As CMC becomes increasingly embedded in our everyday communication, our lexicon will also begin to be influenced by phenomena taking place in CMC environments. This will also render older studies on a field such as word formation, if not altogether obsolete, at the very least outdated, as entirely new word formation types, such as corruption, begin producing more and more neologisms like *noob* and *l33t* into the English language. As such, new research is required to gain a better understanding of how and where present day word formation takes place. That being the case, this study will be looking at word formation types and neologisms used in an internet forum environment. The focus of this study will be on what word formation types contribute most words into the English used on the forum as well as how the topic of the discussion influences the frequency of neologisms in use and the formation types used in creating the neologisms.

Crystal (2001) has described the nature of forums, which he refers to as message or bulletin boards, as asynchronous. This means that the communication does not happen in real time, but is instead spaced over differing periods of time, sometimes with replies weeks or even months apart. Even though Crystal (2001) claims that asynchronous interaction is linguistically less creative than its

counterpart, synchronous interaction, forums nevertheless produce and demonstrate a number of neologisms, either through crossing over from synchronous means of communication or by simply producing new ones. In fact one of the chief motivations for this study has been the creativity exhibited by the users frequenting the forums and how many neologisms can be observed on the forums in the course of every day interaction. This observation is in fact so clearly visible that the users themselves have acknowledged the frequency of neologisms in their language use and responded accordingly. Having realized that the sheer volume of neologisms had begun to hamper the legibility of the forums for new users, the forum regulars had created lists of these neologisms and their meanings so that a new user would not have too much trouble understanding what is being discussed.

Due to the limitations of time and resources, the study will be a case-study on one forum alone with results that can not be fully generalized as representing general forum use. I do, however, hope to provide some observations that could be considered to be valid on a general level of internet forum language use as well as on the differences resulting from a change in topic.

First of all, however, I should introduce the basic outline of this paper. I will first lay a clear foundation for the study by going over previous research on the relevant fields in chapter 2. The chapter will cover some of the basics of word formation, computer mediated communication and language use on the internet. Also present in chapter 2 are some insights into online gaming culture and the unique linguistic items used within. Chapter 3 will discuss data acquisition, detailing what data was gathered, when and how it was gathered and the selection criteria for the data. After detailing data acquisition in chapter 3, chapter 4 will go over data analysis. Of particular interest here will be the research questions for the study and the hypotheses going into the study will be revealed in chapter 4 as well. Included in chapter 4 will also be a description of the methods for analyzing the data as well as a comprehensive outline of the

problems encountered in analyzing the data. Moving on from the analysis of the data, chapter 5 will display the results of this analysis. The results will be ordered according to the two research questions put forward in chapter 4. As such, the results will first explain how the neologisms found in the study were formed and which word formation processes were the most productive in each of the two subforums. After this, differences between the neologisms found from the two different subforums will be presented. Following the differences in the neologisms themselves, the differences in the use of these neologisms between the subforums will be reviewed. This part of the study will focus mostly on the frequency of neologism use. Once the results have been presented and the differences between the subforums have been established, the reasons for the differences will be discussed and finally a number of special features of online language use as observed from the data will be discussed. Chapter 6 will conclude the thesis, summarising the results of the study and discussing the implications of the results of the study. Additionally, the final chapter will offer suggestions for future research in this field. At the end of the study, will be the bibliography as well as the appendix, including a list of the neologisms found in the study.

2 THEORETICAL BACKGROUND

In this chapter, I will examine a number of topics related to the subject of this thesis. The first subchapter, 2.1, will discuss word formation. The three major areas covered about word formation will be: defining the terminology relevant to this study, general principles of word formation and finally a quick overview of the previous research on the field, specifically in the context of the internet accompanied by a brief explanation of what internet forums are. Subchapter 2.2 will give an overview of the academic field of computer mediated communication. The history of the field will be briefly explained, some of the current trends will be discussed and finally I will offer a brief look into the future of the field. In sub chapter 2.3, I will go over the concept of "the language

of the internet". I will present some characteristics of language use that are considered to be typical of the language used on the internet as well as offer some insights into the criticism of whether there even exists such a thing as an internet language. Concluding the chapter is sub chapter 2.4 which will go over the topic of gamers and the environment in which they use language.

2.1 Word formation

In this chapter I will provide information on the previous research on word formation, both in general as well as where it pertains to the internet. I will describe the different processes that take place in word formation, explain how word formation in the English language generally takes place and then consider how these processes may differ in the context of internet language use. The chapter will be in three parts, first one establishing and detailing the required terminology for the study. The second part will cover information on general word formation and a great deal of that part will deal with word formation processes that have remained the same for decades. It will, however, include new information on some word formation processes which are new and have only recently begun seeing use on the internet. The third part will deal with previous research on word formation in the context of the internet.

2.1.1 Defining the terminology

As the terms 'word' and 'neologism' are used constantly throughout this study, it is important to define them in some detail. Starting with 'word', there are different definitions available for the term depending on what is being studied. To start things off, as this study is conducted with data gathered through a written medium, phonological distinctions are both impractical and irrelevant and can be, for the most part, excluded from the definition entirely. The reason for this is that identifying the "correct" pronunciation would in many cases be

all but impossible, as there are no audio samples available for most of the words in the data. Due to these same reasons, the division between acronyms and initialisms, as further explained in 2.1.2, will also be impractical.

Adams (2001) has presented some different definitions of words as follows: one definition would simply be a word form, an orthographical piece of text preceded and followed by spaces. When examining this definition more closely, however, it quickly becomes apparent that limiting the scope of this study to strings of letters separated by spaces would rule out some of the data. For example expressions such as *econ booming*, *turret farming* or *big gen* would have to be excluded from the data, were this definition to be adopted. Another way of looking at words given by Adams (2001) is to look at words as grammatical units. In this way, the singular *sheep* and the plural *sheep* are two separate words, even though they are orthographically identical. Similarly, the words *read* in the sentences "I used to read books when I was young" and "I read books when I was young" would count as two different words under this definition. The inclusive definition given by Adams is to look at words in the sense of a lexeme. Adams explains this definition as the "dictionary" sense, where any plural or grammatical forms of a core word are treated as a single word. As an example, *ride*, *rode*, and *ridden* would all count as just a single word. Of these definitions, I am prone to use the last one. The "dictionary" sense seems the most fitting as with the case of neologisms, these dictionary entries are exactly what I am trying to find. Whether a word *to buff* appears as *they buffed CSM* or *they thought about buffing CSM* is not relevant in this study, nor would it be pertinent to use the word forms *tanked*, *tanking* and *tank* all as unique words of their own. Doing so could potentially skew the data as all three forms mentioned above would count as instances of inflection (and inflectional derivation). This could lead to inflection being represented as up to three times larger than coinage for example due to the different forms that are more easily and naturally formed from verbs than for example nouns, where the only two forms available are singular and plural. In a similar vein, in the case of adjectives, *tanky*, *tankier* and *tankiest* would only count as a single word.

Likewise singular and plural forms would not constitute as two separate words in this study. It is worth nothing here that cases of legitimate conversion, such as when the word *buff* is used as both a verb ("Relic is never going to buff Terminators") and a noun ("They just got a buff in the last patch") will obviously be counted as two separate neologisms. Also, as typos and varying spellings are commonplace in the forums, relying on purely orthographical distinctness without paying attention to meaning would mean that forms like *ult* and *ulti* would count as two different words, where they could just as well be counted as alternative spellings for the clipped form of *ultimate*. No matter how exactly and definitively I attempt to define the terms, there are bound to be cases which will defy defining and will require further deliberation in the analysis chapter.

Now on to the second term to be defined: neologism. Neologism is, in essence, a new word. If one was to take an extreme stance here, one could look only for words that are truly new, by which I mean words that have never before been used. Such an approach would be too drastic and using this definition would complicate the study significantly as it would burden me with the task of finding out when and where a word was first used. It would also limit my study to nonce words i.e. words that are only found once in the data and in doing so reduce the amount of data dramatically. So clearly the definition of a neologism has to be expanded to include words that have seen some use already. When expanding the definition to include words that are not quite as unique as nonce expressions, one quickly finds oneself at the other end of the spectrum: when does a word stop being new? If a word appears in dictionaries, can it still be listed as a new word? Alternatively, if a word has seen decades of use in internet lingo, but has never appeared in any dictionaries, can it still be counted as a neologism in this study? Again, drawing the line will be difficult and it will have to be made on a case by case basis. Setting an arbitrary time period for how long a word actually remains "new" would be just that, arbitrary. However, for the purposes of this study, I set the 20-year mark as a guideline for when one could begin counting a word's age against the word

being counted as new. As the study is about word formation on the internet, it was decided that words that predate the internet (the Internet Protocol Suite, aka TCP/IP, one of the core technologies behind the internet, was standardized in 1982, which is what will be used for determining the age of the internet) should not be counted in this study.

A third aspect of defining neologisms is the frequency and use of a word. Should a neologism then have to be a word that is recent enough to still count as new, but established enough not to be just a nonce occurrence. On the other hand, is there any objection to nonce words? Does it matter whether a particular word has been used by one person or ten million? Does the number of uses somehow make the word a better neologism or a worse one? For this study, it was decided that frequency of use would not be a factor when determining what counts as a neologism and what does not. Nonce words would be right up there with *LOL* and *WTF* when collecting and analysing data. However, when dealing with different or alternative spellings for pre-existing words and separating them from typos, frequency of use would be one determining factor. Once again, such cases would have to be solved on a case by case basis, in order to accurately judge what is, and what is not, just a typo. If one version of a word appears time and again, it would have to count as an alternative spelling, or a corruption, despite appearing to be a typo. Problematic cases would be for example *turret* and *turrent*. *Turret* means an immobile defensive structure and the word is used in a multitude of games. However, *turrent* is a common misspelling of it. At which point does this common misspelling become an alternative spelling or a corruption? Is it enough for a single person to intentionally use the *turrent* spelling and if so, how does one determine intent?

To settle the matter of defining 'neologism', the following definition was chosen: 'neologism', in this study, is taken to mean a novel orthographical unit which dates back (where date of origin can be determined) at most no more than 30 years and is not a grammatical inflection of another neologism already

in the data. Where ambiguity exists, the decision will be made on a case by case basis (see section 3.2). This definition includes the fact that phonological distinctions will be difficult, excludes expressions that are outdated and prevents the overrepresentation of neologisms formed through grammatical inflection. It also allows for dealing with each individual neologism separately when the situation warrants it.

2.1.2 The formation of new words in general

There are several different word formation processes in which new words are introduced into the English language and the first to be described here is **borrowing.** Borrowing is not strictly speaking word formation per se, as it is simply the process of taking a word from another language and using it in one's own language. Therefore, no new word is formed, even though the vocabulary of the target language is expanded. With borrowing, the word retains the meaning that it had in the original language. An example of borrowing would be the word *sauna* in English, borrowed from Finnish. Reciprocally, a borrowing from English into Finnish would be the word *internet*.

Coinage bears some resemblance to borrowing in that a new word can be added into a language without a root of any kind in the target language. However, whereas borrowing simply copies the word and meaning from one language to another, coinage is the process of adding a word to a language from thin air, by naming an object by giving it a new name. Yule (2006) gives *nylon*, *kleenex* and *teflon* as examples of coinage. This name can be based on the name of the company that makes the product or a particular person and in that case, the word is called an **eponym**. There can be a team of language experts deciding on the best and most appealing possible name for a product or it could be that someone comes up with a descriptive term for something and it just gets picked up by the general population in every day use. One example of an

eponym is the word *hoover* in British English, which is used both as a verb to mean 'to use a vacuum cleaner' and as a noun for 'vacuum cleaner'. The word received its meaning from the Hoover Company, founded by William Henry Hoover, which manufactured and sold vacuum cleaners.

Conversion and **semantic shift**, as their names suggest, are two processes where existing words acquire new meaning. **Conversion** (also referred to as functional shift) is the process of using a word from one word class to create another for a different word class, such as *a shovel -> to shovel snow, up -> to up the price* or *a head -> to head the initiative*. In this case the word retains a very close relation in meaning to the original. As conversion takes words and bends their meaning into something slightly different, **semantic shifts** take existing, although often obsolete or antiquated, words and "recycle" them in order to create new meaning for them in a language. This process can take place over decades and even centuries. An example of semantic shift the word *gay* which has changed its meaning from 'happy' to 'homosexual' during the twentieth century, providing a great deal of amusement to adolescents reading older texts where the word is used in its older meaning and then read with the new meaning in mind.

Compounding is a word formation process that also takes existing words and uses them to create words with a new meaning. The difference to the previous two processes is that with compounding, two separate words are put together in order to create a new word. Examples of this process are words like *eyeglasses*, *highlight* and *broadband*. In many cases the meaning of the new word can easily be deduced from the component words, such as the case of *eyeglasses* but in some cases the meaning is not as straightforward: for example, the meaning of 'high-speed internet' is not as instantly apparent from the word *broadband*.

Reduplication is a word formation process which is related to compounding, and even categorized as a sub category of it by some (Bauer, 1983), in which a word or a close-sounding word is repeated to give an emphasized meaning. Bauer (1983, 212-213) describes two different categories: rhyming- and ablaut-motivated compounds and these are what Katamba (2005) refers to as the two main types of reduplication. In the first category a rhyming word element is added either to the end or the beginning of a word to achieve the rhyming effect, for example *super-duper* or *teeny-weeny*. With ablaut reduplication the added word element includes a vowel change in such a manner that it does not rhyme with the original word. Examples of ablaut reduplication are words such as *riff-raff* and *tip-top*. Exact reduplication is a third type of reduplication, however it is the least productive of the three and mostly used in child-speak (*no-no, pee-pee*). There is some obvious overlap between compounding and reduplication and Bauer (1983) does include reduplication as a sub-category of compounding. Katamba (2005) more explicitly defines reduplication as "the repetition of the base of a word in part or in its entirety". It is worth noting that the word-element added in ablaut and rhyming reduplication can be either meaningless (think of *duper* in *super-duper*) or meaningful (either *brain* or *drain* in the rhyming reduplicative *brain-drain*). It is also possible that reduplicatives are formed from two word elements, neither of which could function independently. For example it would not be possible to divide the word *riff-raff* either into two independent components *riff* and *raff*, nor to indentify a clear base for the reduplicative, as *riff-raff* is neither a *raffy riff*, nor a *riffy raff*, so to speak.

Now we get to what Bauer (1983) calls "unpredictable formations". These formation types are clippings, blends, acronyms and word manufacture. Of these, the fourth formation type, word manufacture, is in fact the same as coinage, discussed earlier in this chapter. As such word manufacture will not be discussed further here.

These word formation types are so categorized because from the point of view of generative grammar, they are "very awkward", as Bauer (1983, 232) puts it, in that there is some difficulty involved in creating air-tight rules for these processes. **Clippings** then represent a word formation process that creates new words by reducing already existing words in length, examples being *fax* for 'facsimile' and *the Met* for 'the Metropolitan Museum'. Plag (2003) refers to clippings as a sub category of truncation, which he describes as a word formation process which functions identically to the definition of clippings, that is reducing already existing words in length. Truncation includes the *–y* diminutive (*Andrew -> Andy*) as well as the practice of reducing names in length (*Albert -> Al*). I would argue that clippings are overtaking both the *–y* diminutive and the shortening of names in number and productivity. As such, I would propose that clippings should in fact be considered the main category and the *–y* diminutive and name shortenings should be preferably considered to be a subcategory of clippings as the number of names available for shortening is relatively fixed. As such, *Ron* could be seen as a clipped form of *Aaron* and *Randy* could be seen as a clipped form of *Randolph* using the *–y* diminutive. Plag (2003) also notes that this kind of terminology is in fact what is used in other publications and it is the one I will be using in this study.

Blends are words that have come into existence through combining parts from other words into a single entity. Blends could be confused with compounds which, too, are formed by combining two different words. The distinction is that blends, as their name suggests, blend two words into one, unique entity, whereas in compounding one could easily determine the different components by simply separating the two (Jackson & Zé Amvela 2000). Another definition, which further separates compounding from blending, is offered by Yule (2006, 238). According to this definition, blending is "the process of combining the beginning of one word and the end of another word to form a new word". This definition is also expanded with the addition of "in a few blends, we combine the beginnings of both words". Using these definitions, examples of blends are words such as *modem* for 'modulator/demodulator' and *motel* for 'motor hotel'.

One notably productive source of blends is the yellow press, which creates many of the "fad" words using this formation process. Particularly the practice of talking about celebrity couples by using a blend created by combining the first names of the couple is commonly used. Popular examples of this are the likes of *Brangelina* (referring to the couple Brad Pitt and Angelina Jolie) and *Bennifer* (referring to Ben Affleck and Jennifer Lopez). This practice shows well the few different patterns that are most often used in blending; combining the beginning of one word to the end of another and combining the beginnings of two words to create a new one. In order to show the fluidity and productivity of this word formation type, one can examine the latter of the above examples, *Bennifer*, and its constituent components in more detail. The word was coined when Ben Affleck was coupled with Jennifer Lopez. When Ben Affleck broke up with Jennifer Lopez and began courting Jennifer Garner, *Bennifer* ceased to be used and the media had to invent a new blend. As *Bennifer* was already used and as such, simply combining the first names was not an option, a new formula for creating a blend was also required. Through several different attempts at a new blend, from *BenJen* to more cumbersome ones such as *Bennifer Redux*, it appears that *Garfleck* was finally adopted as the one to use. With this particular practice, words created by its use describe current matters and as such are bound to be rather short lived. This is due to the fact that once the relationship is over or the fad has passed, the need to refer to it also ceases and the word quickly falls into disuse, both by the media that created it, as well as people who used it.

While normally the word *abbreviation* includes all abbreviated forms such as *abbr.*, which would be considered a clipping in this study, **abbreviation** is used in this study as a category of words that can be divided into two sub groups: initialisations and acronyms. These two word formation processes are closely tied together with the only differentiating factor being how the resulting word is pronounced. **Initialisations** are words that are pronounced letter by letter, such as *FBI*, *USA* and *SMS*, whereas **acronyms** are pronounced as regular words, examples being *NATO* and *SETI*. This division into initialisations and

acronyms is a matter of phonology as in order to make the distinction between the two the word must first be pronounced, used in spoken language. It is worth mentioning here that the division into the two sub types will be problematic for the purposes of this study as there will not be audio samples of the neologisms available, complicating the division. In the *Oxford guide to English usage* (Weiner 1983), it is noted about abbreviations that "It is usual to indicate an abbreviation by placing a point (full stop) after it". Examples given for the use of full stop include words such as *Jan.* (meaning 'January') and *Sun.* (meaning 'Sunday'). There are five exceptions given to this general rule. Firstly there are "sequences of capital letters alone" (*MA* for 'Master of Arts', *NA* for 'North America') and acronyms (though, what are acronyms if not sequences of capital letters). Secondly there are numerical abbreviations (*8th*, *10th*) and thirdly symbols for temperature, chemicals, measurements etc. A fourth exception is titles (*Mr*, *Mrs* and *Dr*) and finally something referred to as "words that are colloquial abbreviations". It is interesting to note that these "colloquialisms", so to speak, are viewed as an exception to the rule, something out of the norm. Due to the age of the publication, it might be the case that use of abbreviations has since increased in number so much that these colloquialisms might begin to be the norm and that the more traditional abbreviations are now becoming the exception, at least in online communication. Also, when one takes into account the primary motivation behind using abbreviations in the fast paced, text based communication of the internet; the speed of typing and thus communication, it is obvious that the practice of adding extra characters such as full stops only to indicate that a given expression is an abbreviation is counterproductive and as such the practice has fallen into disuse. Aronoff (1976) also notes that acronyms are a fairly recent addition to the linguistic landscape of the world, having only been present in the English language for over a century. Aronoff (1976) also points out that acronyms are not a natural formation process of language as they require an alphabet in order to take place. As a feature of written language, it is fitting that they find use and thrive in a purely text-based language environment such as the internet.

Hatch and Brown (1995, 210) report that these two word formation processes, acronyms and initialisations, "are not generally powerful processes for forming new words in English. Still, there are many acronyms in particular fields...". The publication could be showing its age here as such fields where the two are not powerful processes are becoming more and more rare, especially with information technology becoming more integrated with every day life. Acronyms and clippings are, based on the results of Kalima (2007), highly productive in an internet forum environment. This is likely a result of both of these types shortening the words so that typing them takes less key presses and thus reducing the amount of time it takes to finish writing a post.

Abbreviations, as Bauer (2003, 238) points out, are based on orthography in that without a writing system, they would almost certainly not exist. This connection to orthography is further highlighted in the internet language use as language use takes place almost entirely in written form (with the exception of Voice Over IP, a term for transferring speech over the internet with services like Skype). So in addition to the word formation types listed above, it should also be noted that are also some typographical word formation processes being used solely on the internet and I will detail them here. The first of these I will call **corruption**, and words in this category are formed by changing letters in pre-existing words, sometimes with little to no change in meanings, and at other times with accompanying changes in meaning as well. The most extreme example of this is the *leet speak* or *1337 5P34K*, which is the practice of substituting letters in words with numerals and other characters in such a way that the characters used resemble the intended letter. Basic examples of such substitution include replacing the letter *e* with the numeral *3* (the numeral 3 resembles a vertically mirrored capital E), replacing the letter *l* with numeral *1* (lower case L resembles the numeral 1) and replacing the letter *t* with the numeral *7* (the numeral again bears a resemblance to a capital T, with the vertical line simply being tilted and meeting the horizontal line at a different

point). Using these substitutions you can deduce that *1337* in *1337 5P34K* means *leet* (or, *l-eet*), which itself will be relevant when discussing the next orthographical formation type, homophonic literation.

Homophonic literation is what I call the practice of spelling words differently, taking advantage of the sound properties of letters when pronounced out loud. To use the example from the previous chapter, *leet*, the word is formed from the base word *elite*. Before we can understand how *elite* becomes *leet*, we must first break *leet* down into its core components and examine their audio properties. When examining the components, we can effectively define two separate parts, *l* and *eet*. The first of these, if pronounced as the letter *l*, is pronounced in a very similar way as the first two letters of the word *elite*. The second part, *eet* is then pronounced in a similar way as the remaining three letters, *-ite*, of the word *elite*. Now combining the two, *l* and *eet*, we have something that is pronounced in a way similar to *elite*. The easiest example of homophonic literation is of course using the letter *u* when talking about the word *you*. Other commonly used examples are words such as *cya* or *cu* for *see you* and *l8er* for *later*. An often seen feature of words produced in this way is that they are shorter than their original versions (*you* -> *u*, *later* -> *l8er*). The shortness of these words can be attributed to the ease of using them when typing as less characters used in a word translates directly to less key presses on the keyboard when typing the word. As such, it would be counterintuitive to create words if using them meant having to use more characters than using the original word would. However when looking at the words formed using this process, such as *kewl* for *cool* and *bewbies* for *boobies*, one notices that not all of the words are shorter than the original words. The fact is that even though these example words are not longer than their original counterparts, they are not shorter either. Without taking a stance on whether *kewl* is somehow easier to type on a keyboard than *cool*, it appears that character length can not be the sole motivation for the creation of such words and there clearly has to be some other motivation for this formation process to be used. I would hypothesize that this motivation is nothing else than being playful with language and using it in creative and different ways.

2.1.3 Previous research on the internet and on the nature of the forums

As the concept of an internet forum is rather central to my study, I will now describe the phenomenon, as well as offer some insight into previous research on it.

The nature of internet forums is such that they are an asynchronous way of communication, which means that the discussion does not take place in real time (Crystal 2001, 22). Crystal refers to forums as "bulletin boards", which is a rather accurate description of how these forums work, i.e. all users read a topic of their choice and then comment on it if they so wish. I am however more accustomed to using the word 'forum' when discussing them and also feel that it is a more suitable term to describe them, due to their active users. Because some users are highly active on these forums, the exchange of ideas can become practically instantaneous, effectively taking more the characteristics of a forum with each individual speaker taking their time to formulate a response, rather than a bulletin board where people occasionally leave notes for others to read. As such I will be using the term 'forum' when discussing them in this paper.

One of the linguistic studies done on this field is an undergraduate work by Driscoll from 2002. The focus of Driscoll's study is very similar to mine in that it tries to determine how a particular group of internet gamers uses language and how new words are formed in that particular group and medium. The study, while done on Internet Relay Chat, or IRC, is related to mine as both focus on internet gamers and as such could be expected to share a fair amount of lexical features, even though forums and IRC are two fairly different types of media for discussion, IRC being a synchronous, real-time medium and forums being an asynchronous one. Another difference between the two is that Driscoll's

study was done on a First Person Shooter (FPS) game called Quake, whereas this study's game-specific section deals with Real Time Strategy (RTS) games. The nature of RTS games is further explained in section 2.4, but regarding the differences between the two, it could be said that in FPS games there is very little down-time when player action is not required (as an opponent could at any time appear from behind a corner or fire at the player from behind, requiring constant readiness), whereas in RTS games there is more time between meaningful player actions (such as when waiting for selected units to be built or waiting for units to complete their orders). This down-time can then be allotted to for example typing messages to your allies for discussing strategy or taunting your opponents. Driscoll (2002) discovered that out of the 72 words found in the data, 29 words were unique coinages, 24 words were clippings, 10 words were acronyms, five blends and four compounds. This should provide me with a good comparison point to see if the shift in the mode of communication (from IRC to a forum) will have an impact on the frequency of word formation processes in use.

There are also a number of dictionaries on netspeak, internet dialect, internet slang, computer jargon, hackish and Net-lingo which all refer to the same linguistic phenomenon. The novelty and the sheer amount of new words have also generated some scientific interest and studies on the topic. It is interesting to note that in their book from only little over a decade ago, Jackson and Zé Amvela (2001, 129) categorise this whole field as an occupational jargon. Clearly that is not the case anymore, as it is used by people completely unrelated to occupational computing and between people who in no way share any occupational features.

2.2 Computer Mediated Communication (CMC)

In this section I will give a brief overview of the history of the field of research called computer mediated communication. I will introduce some of the research

done in this area and I will also present some ideas where the field is believed to be heading in the years to come.

There have been a number of studies in the recent decades, focusing on whichever technical innovation has surfaced and spiked in popularity. Be it email, IRC or more recently Facebook or Twitter, there are usually researchers ready to study the new piece of technology that everybody is buzzing about. This has led to a new field of study called Computer Mediated Communication (CMC for short). In this section I will discuss the development of this field of research as well as introduce some of the main concepts in the field. One very productive researcher in this field is Susan Herring, who has written a number of studies on CMC, ranging from introductory works to more in-depth studies on the pragmatics of CMC, the use of non-verbal communication in CMC as well as gender issues dealing with CMC. This study will refer in great part to her works when discussing CMC and how it is relevant to this study.

To introduce the brief history of CMC, I will paraphrase Herring (2010). The first cases of actual human communication via the use of computers took place in 1972 when the first emails were sent. Text-based CMC became more widespread as internet Service Providers allowed access to the internet from home computers in the later 1980s and early 1990s and finally bloomed in mid and late 1990s. With the internet being accessible to more people, the services available on it, such as electronic mailing lists, MUDs (Multi-User Dungeon / Domain, an online virtual game or educational world that was text-based and had several people interacting with the same environment) and IRC (Internet Relay Chat), became popular. As a result of this popularity, research into these areas of language use began. While new technologies have offered new modes of CMC (Voice over IP, such as Skype or video conferencing), text-based communication has remained the primary method of communication in CMC as is clear when one thinks of the killer applications of today (Facebook, Twitter). Initially, it was debated whether CMC could even be consider to be

conversation or not. While at the early stages of asynchronous CMC, the point might have been arguable, by the time synchronous CMC became popular, the case could no longer be made that CMC could not be counted as conversation simply because "it was not produced orally or received auditorily like speech, and conversation" (Herring 2010). Today, all types of CMC, be they VoIP or text-based, are considered as conversations between two or more people.

Herring (2008) has divided CMC into five different major areas of research: (1) classification, (2) structural features, (3) discourse patterns, (4) lens through which to study human behaviour and finally (5) language and language ecologies. Only the first two will be explained in greater detail here as the other three have little relevance to the study at hand. The first of these assigns and categories to different types of CMC to make it more manageable and to facilitate further research. This has lead to labels such as 'language of email' or 'language of chat groups' with each having distinct features affiliated with them. These features have been the focus of structural feature studies, which has been and continues to be the most popular of the five. Research on structural features has then focused on typography, orthography and neologisms. The conventions associated with the language used on the internet, with its wealth of acronyms, emoticons and varying and creative typography has attracted a great deal of interest and attention. Studying the structural features has lead to an idea of what Crystal (2001) calls 'netspeak', in effect a fairly unified language variety with an established set of shared features. Crystal (2001, 18) defines it as a type of language "displaying features that are unique to the internet, … arising out of its character as a medium which is electronic, global, and interactive." Androutsopoulos (2006) has criticized Crystal's approach, which he terms the first wave of linguistic CMC studies, by observing that as for example the uses and contexts for email use are so varied, so do the practices of language use vary, therefore making an umbrella term such as 'language of emails' inaccurate and possibly misleading. Androutsopoulos (2006) notes that while the variety of group practices is noted, it is not accounted for in any systematic way. His critique also promotes a move

away from medium-related and towards user-related studies, no longer searching for the typical features of a given type of CMC, but instead for how the different features contribute to the language use of the participants. This supports the current study's aim to look at not only the neologisms found and their formation processes, but also at the frequency of their use within different contexts.

Returning to Herring's (2008) categorization of the five major areas of CMC study, Herring also notes that studies on these five areas have appeared roughly in the order given. This is logical, considering that in order to study structural features, one must first classify those features. Likewise, study of discourse patterns is quite difficult if the structural features that make up the discourse patterns are not mapped out first. The focus by researchers has been a descriptive one, where the different aspects of CMC are mapped out and described on how they are seen in actual use (Herring 2008). While Herring (2010) suggests that prime areas for further research are non-English, non-text-based areas of CMC, I find that study of the language used is equally important as the way it is used. Herring (2011) could be seen to promote a move "beyond" the study of the structural features of CMC, such as typography and orthography, while at the same time lauding the contributions to just such fields. As such I believe that there is still a great deal of viable research available in studying the language itself.

Moving on to possible areas for future research in the field of CMC, Herring (2008) suggests turning attention towards theorizing about the effects of CMC on language. These theories should then be empirically tested on large-scale samples, systematically compared between different modes, contexts and languages as well. She also calls for the preservation of as much data as possible, particularly in synchronous modes of communication where automatic logging does not take place. Herring (2008) also predicts increased attention to modes of CMC using spoken language and multimedia. There have been continuous predictions of voice and multimedia modes booming in

popularity due to increased technical resources such as increased bandwidth and increased number of webcameras thanks to the popularity of laptops and the integrated web cameras and microphones within. Despite the technologies allowing for the possibility to move beyond simple text-based communication, that shift has yet to come. Personally I would hypothesize that the main reason for this is the ease of text-based CMC. You are not required to set up your microphone or webcam, nor are you required to worry about your appearance or having a cough or the flu when using text-based CMC. Another extra complication with audiovisual communication is the intensity of the conversation, as you have to pay attention to pauses and other dimensions of face-to-face and live communication. With text-based CMC you can type your response and focus your attention elsewhere whilst waiting for a reply. There is also no need for an immediate reply like there is in face-to-face communication. If someone asks you a question in real life, you are expected to respond within seconds, whereas in text-based chat, it is automatically implied that the reply might not be instantaneous.

Herring (2004), in 2004, proposed that in the coming five years, CMC would become, through technological integration and other factors, a simpler, safer and less fascinating communication environment. Herring (2004) also commented on how more and more of the more traditional ways of CMC are being "united under a simpler browser-accessible format". This is certainly true as the creation and popularity of services such as Facebook, where users communicate under their real names and mostly with people they know in real life, are shifting the average mode of CMC from anonymous to distinct. In older modes of CMC, such as IRC, Windows Messenger and forums, communication has been largely conducted with the use of aliases. I would hypothesize that as the identity of the people communicating was restricted to what the participants decided to reveal about themselves, they were free to use different kinds of language that they might not have used in the company of people they knew in real life. For example *l33t 5p34k*, aka. *leet speak*, the use of substituting letters with numerals and other non-alphanumeric characters available on the

keyboard, would not only have been frowned upon when communicating with someone not familiar with it, it would have been nigh indecipherable. Even if CMC is taking steps toward regular, face to face communication, the anonymous modes of communication are still there and I would wager that they are not going anywhere anytime soon, even if their popularity might diminish with the ease of use and social aspects of the other CMC methods.

2.3 Language of the internet

With the introduction of the internet, there has also been a type of language variant among the users of the internet which has evolved through different stages and modes of communication. Despite the criticism to such a broad term presented in 2.2, it is important to explain certain aspects about the language use on the internet. In this chapter I will discuss this language variant, which will be called, for the lack of a better term, the language of the internet. For the most part, this chapter will offer examples of the features of such language.

Cheater (2006, 20) has listed some features of what she calls 'Hackish', i.e. the English used by hackers and others tech-savvy English speakers on the internet. Among these features are inter-intelligibility among L2 speakers and a preference towards multi-functional forms over single-function nouns. She also suggests that English would be heading towards a pure-positional grammar. As "the single most critical driver" she posits the fact that the digital world is based on binary opposites. "Every concept has its antithesis, every state its negation, every action its reversal. Yet English verbs are generally asymmetrical, the main sets of antonyms being paired prepositions." She goes on to suggest that this binary way of thinking is also creeping into the language, producing neologisms in order to provide such binary counterparts to already existing words. In the data for this study there were some examples of such neologisms, such as the neologism pair *underpowered* and *overpowered* as well as the neologism *non-repair*.

In the following, I will go through the different ways in which e-English varies from regular English, as brought up by Cheater (2006). When examining what is happening with different kinds of words, she notes that adjectives are formed more and more with the use of the affixes such as *–y*, *-less* and *–free*. This would be in line with the notion that Hackish follows clear patterns and logic and thus, when given a slew of examples where such affixes are used successfully to create adjectives, a logical L2 speaker uses these same affixes to create new words, even in situations where a native speaker might be inclined to use other ways of expression. Some of the examples given by Cheater are *deadlockfree*, *lossy* and *mismatchfree*. Related to the creation of new adjectives, Cheater also gives examples of the before mentioned binary opposites and discusses whether the negation of *deadlocked* should be *deadlockfree* or *livelocked*. When aiming for inter-intelligibility, the term *deadlockfree* seems more suitable, as it does not require the knowledge of the antonym pair *dead – alive*. The most logical word, however, would be the form *non-deadlocked*, created with the help of what Cheater calls "the generic reversor", the prefix *non-*. Cheater also mentions "collapses" (here in the case of adjectives, "collapses" refers to blends) as a typical feature of Hackish. She gives adjectival speed and efficiency as prime motivators for the creation of words such as *mechatronic* (meaning *mechanical and electronic*). Both factors are relevant in the gaming environment as well.

Regarding typical changes in nouns within "e-English", Cheater (2006) gives five different **methods**: active nouns, noun/verb names, noun collapse, noun strings and reversing states. The first one refers to two methods of giving binary states to different adjectives such as *configurable* or *maintainable*. In order to talk about whether for example a piece of code is *maintainable*, one can talk about *maintainability*. If one wants to know whether a piece of software is built very rigidly to perform one task and one task only, or whether it can be easily modified to do something different, one can then talk about *configurability*. The other method which Cheater gives to activate nouns is first creating a new verb

with the use of the affixes *–ise* and *–ize*, and then deriving a noun from the resulting word form via the affix *–ation*. With this method you can get the neologisms *parameterise* and *parameterization* from the base word *parameter*. The second type, noun/verb names, refers basically to functional shift, where a single orthographical word functions as a word from two different word classes. Cheater lists several such "e-Neologisms", such as *codesign*, *diff* and *traceroute*. The third type, noun collapses (different from adjectival collapses, which referred to blends), refers here to compounding, where an adjective is "collapsed" into a noun, creating words such as *whitespace*, or two nouns are combined to create words such as *bitwidth*, *namespace* or *timestamp*. Cheater's method of referring to both blends and compounds as collapsing is, while logical, also confusing. While both compounding and blends do indeed combine two words resulting in a single word, for the purposes of my study, the clear division into blends and compounds instead of adjectival and noun collapses is preferable. The fourth method, noun strings, refers to a habit of naming new things by stringing together a number of nouns to describe what is being talked about, such as *face recognition committee machine*. According to Cheater, successful strings are then often shortened into acronym forms, again for the reasons of speed and efficiency. An example here would be *light amplification by simulated emission or radiation* or, nowadays more commonly known in realms of physics, ophthalmology and science fiction as, laser. The final method used with nouns, reversing states, is a rarer type, where instead of using traditional forms of negation, or even the "universal negator" discussed before, forms such as *notwork* are used to refer to a network that is in fact, not, working.

Adverbs are, according to Cheater (2006), used less in e-English due to the binarization of the *–ise* verbs, leading to a reduced need for graduated qualification of adjectives. Size prefixes (examples include *kilo-*, *mega-*, *giga-*, *solo-*, *duo-*, *tri-* and *multi-*) are also mentioned by Cheater as a common component in e-English words. In particular, *multi-* is mentioned as a particularly productive component, featuring in dozens of neologisms from

multiuser and *multiplayer* to *multiplatform* and *multiplexing*. Finally, Cheater describes a feature of e-English that she calls multiform neologisms created around core words. These are a group of neologisms, mostly compound words, which are created using a core word and then adding on a descriptive component in order to create a neologism. Cheater defines three different categories of such neologisms: neologisms with technical terms as their foundation, neologisms with common words to describe new technical applications and suffixed forms. In the first category she includes neologisms based around words such as *band* (*band-limited*, *bandwidth*, *narrowband*) and *bit* (*bitmap*, *bitwise*), while in the second category, the core words include *net* (*cheapernet*, *subnet*), *web* (*webcam*, *webpage*), and *code* (*pseudocode*, *codebook*). The third category uses suffixes like *-ware* to create neologisms such as *abandonware*, *bloatware* and *shareware*.

The language of the internet was seen for a long time as a niche variant, only used by the few and the nerdy. However this has been changing in the recent years as internet memes have started becoming mainstream with the popularity of services such as Youtube and Facebook. Cheater (2006) suggests that "e-English is not a perhaps-amusing 'dialect'. The internet is the future of virtually all forums of communication, written and verbal". At first such a claim seems rather extreme, but after one considers how big a part the internet can play in one's every day language use, the claim begins to gain credibility. At work, one communicates with others via email and instant messaging and after coming home from work, one opens up Facebook, different instant messaging programs, possibly IRC and starts watching content such as Youtube videos, different news streams and news articles from internet news services and listens to music and other content from different internet radios. And with the development of mobile phones into "smart phones" with internet access and in-built internet browsers, most of this content can be taken with you wherever you decide to go and in fact it is already possible to stay "online" throughout the day on one's smart phone, updating your Facebook and Twitter status and logging your whereabouts on services such as Foursquare. While it is unlikely

that all forms of communication would be transferred to the internet, it is much more plausible to think that a great deal of our daily language content will revolve around the internet in one form or another. And certainly when one considers how readily accessible content on the internet is and how many people use the internet, it is easy to imagine that language practices used on the internet can easily transfer into face to face communication. Consider for example the fact that at the time of writing, Facebook has over nine hundred million active users (Hachman, 2012), YouTube gets over 800 million unique users each month (YouTube, 2012) and Twitter has over half a billion users (Dugan, 2012). Putting those numbers into context, Facebook user numbers represent almost thirteen per cent of the world's population, whereas Twitter's numbers equal over seven per cent. When this many people frequent sites such as these it is no longer difficult to imagine that any changes in language use on the internet can be quickly reflected in the offline environment as well. Anecdotal evidence of this is young people using terms like *LOL* (meaning *laughing out loud*) or *trolling* (meaning *to elicit a hostile response from another user*) in face to face conversations.

2.4 Linguistic features of the language of gamers

In this chapter I will explain some of the characteristics of the language used by online gamers. I will describe relevant aspects of the games involved, the different types of communication used in the context of online gaming and offer some results of previous studies in this area.

In order to further explain how language is used by the gamers in the study, I should describe certain aspects of the type of game that is played on the forums that the data is collected from. The game is a Real Time Strategy game (RTS for short), where players control a varying number of different units on the screen using the mouse and keyboard. The game revolves around managing different resources found on the map to produce the units needed to counter whatever

units the opponent is producing with the resources he or she has acquired. The game is fast paced and there is practically no down time during which actions are not required by the player. The game also has inbuilt chat functions both for text-based communication and for Voice over IP (VoIP) communication. Due to the player having to keep his focus on the game at all times, any communication has to be as brief as possible in order to minimize the time away from taking actions in the game. This places demands on the language used such as brevity, succinctness and cost efficiency on the meaning conveyed to characters used ratio. Herring, Kutz, Paolillo and Zelenkauskaite (2009) conducted a study on the text chat in an Online First-Person Game which, while having a different genre of game, shares all the relevant features of RTS games. The results of the study that are relevant to this one were the findings about the kind of language used. These results were as follows: First of all, the chat messages in the data were only 1.4 words per message. Words within these messages were also short with an average length of 3.5 characters. The results support the hypothesis made in the study that "messages will be short and abbreviated, moreso than in other modes of recreational chat". Herring, Kutz, Paolillo and Zelenkauskaite (2009) note that "The use of abbreviation can be seen as an effective strategy to communicate under extreme time pressure using minimal keystrokes." The findings support the hypothesis of this study (discussed in chapter 4.1) that shortened forms such as abbreviations, blends and clippings will be in the majority of the words found in the data. Further supporting this hypothesis is the fact that Hatch and Brown (1995) also maintain that clipping is commonly used among close-knit communities as well as in computing. As internet gaming forums tend to combine both close-knit communities as well as computing, it seems logical to assume that clipped forms would be commonly found in the data as well.

Jargon is also a relevant concept as the gamer language could be considered a type of recreational jargon. Harley (2006) has so far the most concise explanation for jargon: "Jargon is just specialized terminology used by a particular group of people to serve its everyday communicational needs. It is

special to that particular group because other groups aren't thinking about or working with the same concepts on a daily basis." First of all, this definition shows that jargon is used for a reason, that is to communicate with others and secondly that jargon is specialized based on the needs of the group using it. The point that jargon is specialized between groups is also an important one as one of the two different topics of discussion examined in this study, the gaming subforum, is a very specific area of discussion whereas the other, general discussion subforum, is, as its name suggests, a much more general area of communication. As such, it could be expected that the more specialized area of discussion would exhibit more features commonly associated with jargon than the more general area of discussion. Whether this was the case or not will be explained later in the study.

3 DATA ACQUISITION

In this chapter I will go over the main questions relating to data acquisition for the study. The main issues in this chapter are twofold. First, data acquisition for the study will be explained in detail in section 3.1. What data was acquired, where it was acquired, what method was used in gathering the data and how much data was gathered. The reasons for why a particular forum over the thousands of forums found on the internet was chosen for the data gathering will be explained in section 3.1. Secondly, in section 3.2, there will be a closer look into what selection criteria was used when deciding which words to include in the data and which words to leave out. These selection criteria were already mentioned briefly in the section 2.1.1; however, here they will explored in more depth and detail.

3.1 Compiling the research materials

In this chapter I will explain how the data was gathered for this study. I will go over the basic details of data acquiring: what, when, how and why the data was

gathered. I will also give the numerical information on how much data was gathered altogether.

The data for the study was gathered from an internet forum called Relicnews (forums.relicnews.com). It is an internet forum that has a history dating back to 1999 and it has a range of subforums which include both highly specialized gaming forums as well as general discussion forums where thread topics can vary from current TV shows and literature to political topics. These forums were chosen for a number of reasons. First of all I was already familiar with the forums, having frequented them for over five years. This meant that navigation on the forums would be easy and the forum functionalities were already familiar, which would save time and effort when gathering the data. Another benefit of this familiarity was that I was already accustomed to the language used on the forums, which would ease the task of finding meanings for the different neologisms and in some ways also help with determining the etymology and the word formation processes used in the neologisms. For example, trying to decipher what the neologism *ATSKNF* (an abbreviation for 'And They Shall Know No Fear') meant would have been difficult without prior knowledge of what context it is used in. Similarly, determining the origin of the neologism *cults* (a semantic shift meaning 'heretics') would have been difficult without the prior knowledge that while the unit that *cults* referred to was called *heretics* in this iteration of the game, the corresponding unit in the previous iteration of the game had been called *cultists* and as people were familiar with the unit from the previous iteration, the clipped word carried on with the unit even if the units name changed between iterations. Familiarity with the forums had a downside as well, however. Being familiar with the language and the neologisms meant that it was quite easy to miss neologisms as they had become a part of my vocabulary. At times when going through the data, I would realize that I was concentrating on what was being said instead of what words were being used to say it and would have to go back and go through parts of the data again, concentrating on finding the neologisms. Nonetheless, all the advantages of being familiar with the language and the forums before hand far outweighed

the drawback of occasionally having to return back a line or two after getting too interested in the text.

The second reason for choosing these forums was that they contained both a general discussion subforum as well as a game specific subforum and both of them were in active use. While most gaming forums also include an "off-topic" subforum, it often sees very little use as people mostly frequent forums for specific topics of interest. For example, it would be strange for people who are interested in football to get active in discussing gardening (despite that possibly being of great interest to them) at the football forums when there are plenty of forums specifically for gardening available. As such, having both subforums active was preferable. Another perk of having both forums active was that there would be overlap between the users frequenting the different subforums. This would lend more credibility to any results found in the study as the data gathered would be at least partially from the same users and as such it would be more likely that differences in language use would in fact be a result of the topic and not just differences in personal language use.

Once the forums for the data gathering had been chosen, it was a question of actually gathering the data. The data gathering took place in two stages: the first stage was in spring 2010 and the second took place during the summer of 2012. The two stages were necessary as during the first, what was originally intended as the only stage of data gathering, the amount of data gathered was unequal with the general discussion subforum sample size being much larger than the gaming specific subforum's. Once this discrepancy was noticed at the beginning of the analysis phase, the second stage was implemented to equalize the sample sizes. For the raw data used in the study, a number of threads (a thread is a set of forum posts unified under a single topic and a starting post) were selected from both the gaming subforum and the general discussion subforum. There were no particular selection criteria for the threads, other than that they were selected from roughly the similar time period. The threads were,

however, gathered so that an equal amount of text would be available from both the general discussion threads and the gaming threads. Once a thread was selected, it was viewed in printable mode in order to eliminate some of the unnecessary text from the forum function buttons and links and then copied to a text file. Here equal amount of data is taken to mean an equal number of words in total, not just neologisms. Once enough raw data was gathered, it was possible to proceed to filtering out the regular words and concentrate on the core data of the study, the neologisms. The neologisms were sorted from regular words by hand and then categorized based on the word formation types used in creating them. When a neologism could be seen to have been formed using multiple word formation processes, all relevant processes were marked as being part of creating the neologism. For example, when coming across the word *shuriplat* which blends the words *shuriken* and *platform* and clips the latter word into *plat*, the word was marked as representing both blends and clippings. As a result, the sum total of the words representing different types of word formation processes appears to exceed the number of original neologisms examined.

The total number of words in the data was 74 776 words on the game specific subforum and 75 502 on the general discussion subforum. Combined, the word count is 150 278. These numbers also include some miscellaneous data such as time stamps, usernames and some forum functions which were transferred with the threads and which would have been too arduous to remove as this would have meant going through close to a thousand posts and removing individual time-stamps and other pieces of text. This miscellaneous data, however, constitutes only a small percentage of the data as a whole.

3.2 What is a neologism?

First of all it was stated that a neologism should not include words that dated back more than 30 years, as words that predate the internet would make poor examples of word formation on internet forums. Regarding the age of a word, it

was also mentioned that 20 years would be a mile stone that would count against a word being regarded as a neologism. To determine the age of a word, the online version of *Oxford English Dictionary* (www.OED.com) was used. In some cases the word could be found in the OED and in such cases it was necessary to find alternative ways of determining the age of the word. These alternative ways included finding out the meaning of the word and then if the word referred to something that could be linked with a specific date (for example *CODIS* for 'Combined DNA Indexing System', which was only white-papered in 1989 so it could safely be assumed that the word itself was not much older than that), it was possible to determine the age of the word as well. In order for a word to be included in the data, there had to be some reference to how old the word was proving that it was not over 30 years old.

In addition to these temporal limitations, there was one significant selection criterion which was included. This was a group of words on the gaming forums which came from the fictional universe, *Warhammer 40,000* that the game is based on. The universe is created as a backstory for a table top game, and it was created in 1987, meaning that many of these words would be well over 20 years old and as such it would be debatable whether they would be neologisms to begin with. Examples of this group include words such as *hormagaunt brood*, *genestealer brood*, *stikkbommaz, chaos shrine of Tzeentch* and *shuriken cannon weapon team*. Additionally, all of these words are coinages and including them as neologisms would have skewed the results in favour of first of all coinages and second of all in favour of the gaming subforums data in terms of the frequency of neologism use. Because of this, words that were directly imported from the game into the forum discussion, unchanged, were not included. Any derivatives from such words, however, were, as these derivatives are at the core of this study; words created on the internet.

Usernames, when they appeared in the main text of posts (as opposed to in the "posted by" field of the posts), were also present in the raw data. They were not

counted as neologisms, but any derivatives from usernames were, similarly to how names would not be considered neologisms but new nicknames would be.

4 ANALYSIS OF THE DATA

In this chapter I will detail the process to analyze the data of the study. I will introduce the research questions of the study and explain how they were chosen. I will go into the detail of the methodology relating to the data analysis and I will also explain what difficulties and problems were encountered during the data analysis and how they were handled.

4.1 Research questions for the study

In this chapter I will present the research questions for this study, provide the reasoning behind the research questions and also explain what initial hypotheses there were concerning those research questions when beginning this study.

When thinking about the research questions for this study, I needed to consider ways to expand the scope of research from just studying the different word formation processes used in the creation of neologisms on internet forums, which had been the primary, as well as the only, goal of my bachelor's thesis. Moving away from the idea of strictly descriptive analysis of the neologisms found on internet forums and the associated word formation processes, comparative study of the language used on internet forums and the effects of the topic of the conversation on the type of language used seemed like a logical next step. As such, I decided to study the differences between both the number of neologisms found on the gaming subforum compared to a general discussion subforum and the frequency of use of those neologisms within the two subforums. When thinking about how these issues could be studied, I decided

that the best way to proceed would be to gather equal amounts of data from both subforums, analyze both of these data sets and then compare the results found on each of the two for possible differences.

With a clear idea of what the study would be about, it was then a matter of formulating a set of research questions that would best help in delving into the topic. First, in order to find out about the differences in the neologisms and word formation processes between the two subforums, it would be necessary to first study them on their own. As such the first research question would deal with the word formation processes found on the two subforums. Once it had been ensured that the initial data necessary for the study had been covered, it was then necessary to define further research questions to deal with the frequency of use of the neologisms as well as the effect of the topic on their use. Thus, the second and third research questions were defined so that the second would deal with the frequency of the neologisms and the third would concentrate on the possible effects of the topic on the neologisms.

With the above in mind, the final research questions used in this study were defined as:

RQ1: What are the most often used word formation processes in creating the neologisms used on a gaming subforum and a general discussion subforum?

RQ2: Are there differences in the frequency of use of neologisms on different subforums based on the topics of discussion?

RQ3: Does the topic of the discussion influence the use of neologisms on internet forums?

Going into the study, there were some initial hypotheses concerning the results of the study, based on the results of the previous study. These hypotheses were:

H1: Acronyms and clippings will be the two most common word formation processes found in the data while borrowings will be rare among the neologisms, if they appear at all.

H2: Neologisms will be more common in the gaming subforum than on the general discussion subforum.

It should be noted that originally, the first research question was: "*How much are different word formation processes used when creating neologisms on internet forums?*". However, during the analysis of the data, it became clear that determining which word had been created on internet forums and which word had been created in another setting (such as in-game chat, IRC channels, face to face conversations etc.) would be impossible. As such, the research question was modified so that the focus of the study remained the same while still allowing neologisms found on the forums to be used in the data without confirmation of whether they had been created on the forums or if they had been created elsewhere. It was determined that the origin of the neologism was irrelevant as the fact that the neologism was being used on the forums was enough to justify its study as part of the language on internet forums.

4.2 Tools used with the data analysis

This section will detail the tools used in the data analysis, why they were chosen, how they were used in this study and what shortcomings there were with the tools.

In order to determine the frequency of neologism use on the forums, the data was entered into the word counter and frequency tool at http://rainbow.arch.scriptmania.com/tools/word_counter.html. After combining all the threads into a single text file and then copying the contents of the file into the tool, the tool gave an alphabetical list of the words found in the

data as well as the number of times they were used. Additionally, there were two different settings available with the tool; “count pure words” and “count everything as words”. The difference between the two was that the “count pure words” option listed strings of characters that fit certain parameters. An accurate list of these parameters was not available, but at least quotation marks were removed when using this option. Because using the “count pure words” approach would have excluded some of the neologisms found in the “count everything” list and the manual list of neologisms, it was decided that the “count everything” option would be used and then screened manually. This process allowed for double checking the data for any neologisms that might have been missed in searching the raw data for neologisms and several neologisms were in fact found only during this double checking process. In hindsight, it could have been more effective to only go through this data output from the analysis tool from the start and forego going through the raw data manually in the first place, as it could have been more time efficient to search for the neologisms from the list of unique character strings instead of reading through entire posts in the raw data, due to the amount of words being significantly lower in the data output list as it only included roughly ten thousand words per subforum whereas the raw data included roughly 75 000 words per subforum.

4.3 Methods of analyzing the data

In this section I will explain the methods used in analyzing the data. First I will explain the general process of the data analysis. The methods of analysis are then detailed, what aspects of the data were analyzed and how, including the categorization of the word formation processes used. I will also discuss the problems relating to the methods used in the study that were encountered during the analysis.

The data was analyzed in five main phases. The **first phase**, finding the neologisms, was a matter of manually searching through the data gathered from forum threads for the neologisms. Once a neologism was found in the text it was written down, containing the form in which it was found and its gloss. The second part of the analysis could, for some words, be completed at this point for if the word formation process (or processes, in cases where more than one word formation process was present) that was used in creating that neologism was immediately apparent (for example, it was clear that the word *CL* had been created by abbreviating the words *chaos* and *lord*), that word formation process would be written down as well. If the word formation process(es) were not immediately clear, it was added later when the neologisms were categorized according to word formation types. Once this list was complete, it was double checked when screening out the non-neologisms from the output received from the data analysis tool described above in section 4.2. If an assumed neologism was found in the output of the data analysis tool which was not included in the original list, it was first looked up in the raw data and then determined whether it was a case of omitting it when manually looking for the neologisms, or if it was simply a typo. For example, the output from the data analysis tool showed an expression *JST*, which on its own appears to be an abbreviation. However, when finding it in context within the raw data, it was discovered that it was simply a typo of the word *just* and as such was not added to the list of neologisms.

The **second part** of the analysis followed once the neologisms had been extracted from the raw data. At this point it was then a matter of categorizing the neologisms according to the word formation process used in creating them, in order to find out which word formation processes were used the most, what effects topic would have and if the word formation process affected frequency of use. When categorizing the neologisms, the following word formation types were decided as the main categories:

- abbreviations

- affixations
- alternative spellings (including corruptions and homophonic literations)
- blendings
- borrowings
- clippings
- coinages
- compounds
- conversions
- loan words
- reduplications
- semantic shifts

Of these, abbreviations include both initialisms and acronyms. Additionally, a new category was required for words (or word elements, rather) categorized as new affixes. Words where onomatopoeia was involved were categorized under coinages. There were no reduplications or borrowings found in the data so these categories will not appear in the results.

During this categorization, the words were marked up on each different category depending on whether that word formation process had been a part of the process of creating the word or not. As mentioned above, during this phase it was possible for a neologism to be marked as having multiple word formation processes being part of the creation process of that word. The problem at this stage was determining what counted as "being a part of the process", especially in cases where the neologism was formed by modifying another neologism. For example, with the word *zoner*, the base word is *to zone*, which means 'to keep enemy units out of a certain area', which in itself is a neologism formed by the word formation process semantic shift. When creating the word *zoner*, the base word *zone* was used and then via affixing the verb *zone* was transformed into the noun *zoner*, meaning 'a unit that can zone enemy units'. Now when looking at the word *zoner,* it is clear that affixing was used when creating the word. However, should semantic shift count as being part of the word formation process or not? It was certainly used when creating the base

word, without which the term would not exist. Sometimes in such cases, an issue that caused further problems was that there was a great deal of difficulty in determining which one of the two neologisms was the base neologism and which was the "new" neologism. Good examples of this are the two neologisms *debuff* (used as a verb) and *debuff* (used as a noun). When used as a noun, the term means 'a negative effect applied to a unit' and when used as a verb it means 'applying a debuff to a unit'. There was no way of telling which of these uses was used first. Both of them are also formed by adding the negative prefix *de-* to the neologism *buff*, which is also used both as a noun and as a verb, meaning the opposite of *debuff*, i.e. 'a positive effect applied to a unit' and 'applying a buff to a unit'. Thus, both *debuff* as a noun and *debuff* as a verb would have affixing included as a word formation process. As for the problem of which was first, the verb or the noun, it was ultimately inconsequential as for the purposes of the study, for as long as both words were included in the data, all the relevant formation processes would be represented in the data and the sum of the word formation processes used would remain the same, regardless of which of the two words was assigned to conversion.

Throughout this phase, neologisms gathered from the game specific forum were kept separate from the words gathered from the general discussion forum. Once this phase was over, the total number of instances of word formation processes used were totalled in such a way that words that had been formed using multiple processes would count as instances of every word formation process used in creating them. To clarify, the word *MoTDread* is a compound of the words *MoT* and *dread*, which themselves are an abbreviation of the words *Mark of Tzeentch* and a clipping of the word *dreadnought*. As such, *MoTDread* was counted as an instance of all three formation processes, compounding, abbreviation and clipping.

Once the words had been categorized, it was time to move on to the **third phase** of the analysis, determining the frequency of use of neologisms. At this point, the number of occurrences of the neologisms in the two different sets of data

were counted using the word counter and frequency tool (described in section 4.2). This process was somewhat more arduous than originally planned as first of all the tool could not separate between neologisms and regular words so the resulting list had to be manually screened for the neologisms. Once the regular words had been screened, the number of times used were then marked up alongside the already categorized word formation types as well as the meanings of the neologisms.

After the ground work, finding the neologisms, categorizing them based on word formation types and determining the frequency of use, had been completed, it was possible to begin the **fourth phase** of the analysis by comparing the data from the gaming subforum to the data from the general discussion subforum. First the two sets of data on word formation types used on the two different subforums were compared to one another in order to find out possible differences between the two. The comparison was done by comparing the number of neologisms formed with abbreviation on one forum to the number of neologisms formed with abbreviation on the other, then moving on to affixes, alternatives and so on until all the different formation types had been compared.

Once the comparison between word formation types was done, the final and **fifth phase** of the analysis could begin, where the frequency of use from the game specific forum was compared with results from the general discussion forum in order to determine what differences could be found.

4.4 Challenges during the analysis process

In this section I will explain some of the challenges that were encountered during the analysis process. Some of these challenges were orthographical remnants of forum functions impacting the data, the difficulty in separating

acronyms from initializations, determining the age of neologisms found in the data and whether or not to include words in common use as neologisms in the study.

As many of the words in the data were new, determining the etymology of words was often challenging and as etymology is a critical part of determining the word formation processes involved, many of the words in the data will be discussed here individually in order to clarify why and how they were categorized under specific word formation processes.

Another challenge was encountered when determining the age of words found in the data. Terms such as *DoD* (meaning *Department of Defense*), which, to the best of my knowledge is a common abbreviation, could not be located in the OED and as such would be candidates for neologisms in this study. However, the *Department of Defence* has obviously been in existence for longer than the internet, but as for the abbreviation *DoD*, it is unclear how long it has been in use. Typically the words that caused these types of problems were ones that were from fairly specific fields and it is possible that such words had been used for decades. They were however not mainstream words that would be included in the OED and this is where the problems with defining the age of the word rose. In some of these cases the word's age could be established with relative ease. For example when searching for *OPP* (for *Ontario Provincial Police*), a quick Google search revealed the website for the OPP and on the front page the logo was shown, prominently displaying the three letters in the centre of the logo and the history section detailing how the OPP was founded in 1909. Combining these two, it was clear that barring a very recent rework of the organization's logo, the abbreviation had been in use for a long time. However in other cases it was not quite so easy to determine the age of the word and as such a judgement call had to be made on whether to include the term as a neologism or not.

A problem with mapping out the neologisms used on the forums was encountered when looking in more detail at words which at first appeared as corruptions. As corruptions are in their most basic form "just" misspellings of a word with very little change in meaning, there was some difficulty deciding when a word found in the data was a corruption and when it was simply a typo, a case of lazy typing or possibly a clipping. Examples of this problem in the data were the words *cld* (for *could*), *cldnt* (for *couldn't*), *altho* (for *although*) and *gna* (for *going to*). Related to orthography, cases when two words were very similar in orthography and shared the same meaning, it was difficult to determine whether one was dealing with two separate neologisms or whether the two were alternative spellings and if so, which should be considered the "true" spelling of the word and which should be considered the alternative one. Examples of such cases were the words *HBDEV* and *hdev* (for *Heavy Bolter Devastator*), *knockback* and *knock back* (meaning an effect in the game which knocks units back, making those units unable to take action as they get back on their feet), *cos* and *coz* (for *because*) and *fckn* and *fn'*(for *fucking*).

Some abbreviations proved to be problematic for the text analysis tool used in the study. As the tool listed all the occurrences of the words without paying attention to capitalization of the words, the abbreviation *BE* for *Battle Equipment* proved to be problematic as the sequence of the letters *b* and *e* was present in the data both as the neologism *BE* as well as the all too familiar verb *be*. In situations like this, it was necessary to manually go over the data and make a list of when *be* stood for *Battle Equipment* and when it was used in the more traditional sense. Examples of neologisms where this problem presented itself were: *BE* for *Battle Equipment* and *UP* for *underpowered*. A related problem was encountered with compound words where the two words were written out separately with a space in between, such as *focus fire*, *gen bash*, *global rep*, *HB dev* and *psi storm*. The analysis tool's definition of a word was just a string of characters with a space at the beginning and at the end. As such, it would count the word *global rep* as two words, *global* and *rep*. This meant that when looking at the frequency of that word, the tool could not tell how many times the word

had been used in the data and it was necessary to once again count occurrences of such neologisms manually.

When the threads were viewed in printable mode, some of the forum function indicators were still shown. Such forum function indicators included the date- and timestamps shown on each post, the *Quote* word shown for every use of the quotation function as well as the two letters, *PM*, shown on each post next to the user name of the poster noting the button which in normal viewing would allow the viewer to send a private message to the user. The presence of these will shift the ratio of neologisms to normal words in the normal words' favour. However, as this affects both the gaming discussion subforum data and the general discussion subforum in relatively equal manner, the ratios will be equally skewed for both of the subforums, so comparisons between the two will still be valid. There are some ways that could be used to effectively remove this problem. One is to edit out these indicators from the raw data entirely prior to the data analysis. The second solution would be to do an analysis of the average length of posts on the two different subforums, then determining how much the indicators contribute to the total sum of words in the data and then take this into account when doing the analysis. Both of these methods would, however, be so time-consuming that implementing either one would at least double the amount of work required for the analysis and as such were deemed impractical for this study.

Related to the typography of the words, words where conversion was involved, there was difficulty in determining the frequency of the words involved as the data analysis tool could not differentiate between verb forms and noun forms of words such as *nerf* (a semantic shift from the type of rubber used to manufacture children's toys and sports equipment, which received a new, though related, meaning of making something less effective) or *nade* (clipped from *grenade*). In such cases, the plural forms could also collide with the verb forms such as in the sentences "*if the Knob **nades** a HB team and then knifes one,*

the team is as good as dead" and "*stikkbommaz have nothing but their **nades**, that's all they can do"*. When words such as these were encountered, it was necessary to manually go through the data and separate the instances where the word was used in verb form and when it was used in noun form.

5 RESULTS

In this chapter the results of the study will be presented. First the results of the analysis in the two different subforums will be presented individually. Section 5.1 will report the results of the gaming subforum while section 5.2 will detail the results of the general discussion subforum. After the results of the two have been explained, the difference between the results found on the two different forums will be presented in section 5.3. These differences will be by comparing the two sets of results. The comparison points will be made between the following items: the most common word formation types in each of the different subforums and the frequency of neologisms used within each subforum in the full data. The two hypotheses presented in section 4.1 will be briefly discussed and it will be established whether they were proven correct or incorrect. Finally, section 5.4 will detail several points of interest found in forum language use.

5.1 Findings of the analysis on the data gathered on the gaming subforum

In this chapter I will present the findings of the analysis of the data gathered on the gaming subforum. First I will go through the numbers: the number of neologisms found on the gaming subforum, the types of word formation processes used on the neologisms as well as the number of times these neologisms were used. After the numbers, I will explain other findings on the gaming subforum such as the general tendency to use abbreviations in forming new words from the game content, the dominance of abbreviations and

clippings in the creation of the neologisms used and the internal regulation when creating new neologisms that would conflict with pre-existing abbreviations.

5.1.1 The numbers in the gaming subforum

The total number of unique neologisms found in the gaming subforum data that passed the screening was 331. Of these, abbreviating was seen as a factor in 131 neologisms and clipping in 104 neologisms. After these two dominant word formation types, the remaining word formation types and the frequency of use in order of magnitude were: semantic shift with 46 neologisms, affixing with 25 neologisms, compounding with 24 neologisms, alternative spelling with 15 neologisms, conversion with 14 neologisms, blending with 12 neologisms, coinages with 8 neologisms and finally the lone loan, *kekekeke*. Three new affixes were also found in the data: the suffix *-spam* as well as the prefixes *insta-* and *auto-*. The number of neologisms where multiple word formation processes were seen was 48 with four neologisms which had three identifiable word formation types and 44 neologisms which had two identifiable word formation types. The type frequencies of words representing different word formation processes as well as the percentages of the total can be seen below in table 1.

Table 1. Type frequencies of word formation types used in the gaming subforum.

Word formation type	Type frequency	Percentage
abbreviations	131	34.20 %
clippings	104	27.15 %
semantic shifts	46	12.01 %
affixes	25	6.53 %
compounds	24	6.27 %
alternative spellings	15	3.92 %
conversions	14	3.66 %
blends	12	3.13 %
coinages	8	2.09 %
new affixes	3	0.78 %
loan words	1	0.26 %
TOTAL	383	100.00 %

As hypothesized before in section 4.1, abbreviations and clippings were by far the most common word formation types in the neologisms found, having both a over a quarter of the total. Combined, the two comprised over sixty percent of all the neologisms in the data, more than the remaining nine formation types combined. This dominance is best shown in figure 1 below.

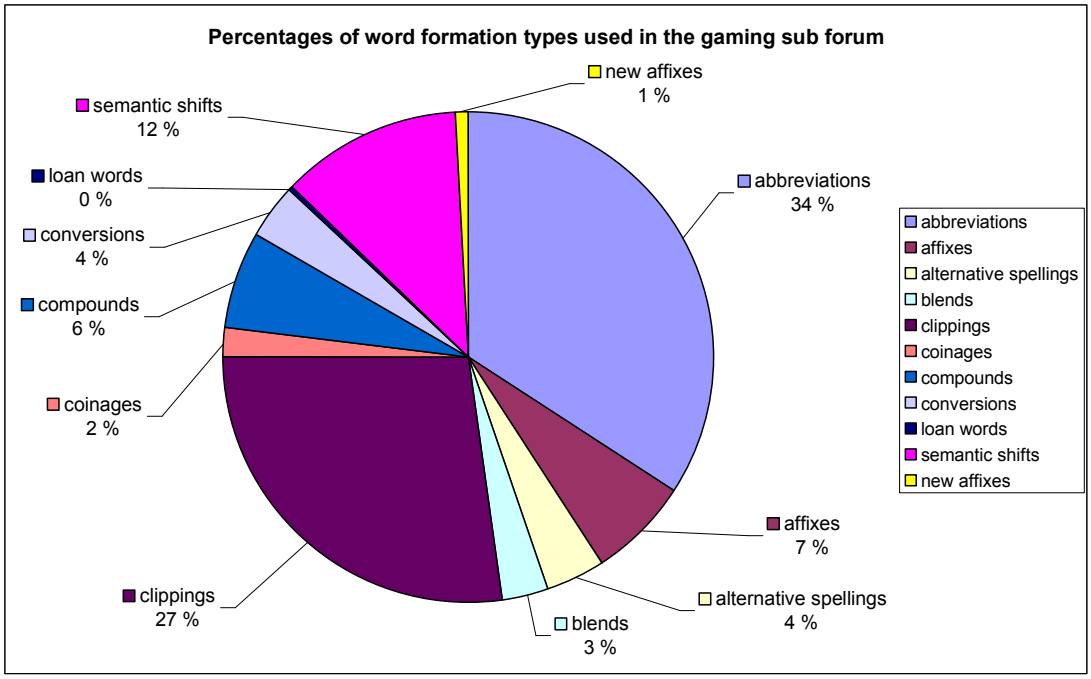

Figure 1. Percentages of word formation types used in the gaming subforum

In terms of the frequency of use of individual neologisms, abbreviations and clippings were also by far the most commonly used word formation types. Abbreviations were used 1287 times and clippings 490 times. The third type of words most often used was semantic shift, with the words used 234 times. Words used by other word formation types were all used less than a hundred times. The token frequencies representing the other formation types are: 80 conversions, 66 affixes, 52 compounds, 30 alternative spellings, 25 coinages, 21 new affixes, 12 blends and 1 loan word. These results are shown below in table 2. The most often used neologism in the data was *GFWL*, which is an abbreviation for 'Games for Windows Live' and it was used 72 times in the data. Other often used neologisms were *CSM* for 'Chaos Space Marines' (64

uses), *IG* for 'Imperial Guard' (55 uses), *PM* for 'Plague Marine' (57 uses), *T2* for 'Tier 2' (56 uses) and *DPS* for 'Damage Per Second' (53 uses).

Table 2. Token frequencies of word formation types used and averages of use per word in the gaming subforums.

Word formation type	Token frequency	Average of uses per word
abbreviations	1287	9.82
new affixes	21	7.00
conversions	80	5.71
semantic shifts	234	5.09
clippings	490	4.71
coinages	25	3.13
affixes	66	2.64
compounds	52	2.17
alternative spellings	30	2.00
blends	12	1.00
loan words	1	1.00
Total number of word uses	2298	-
Average times used of all words	-	4.02

As can be seen in table 2 above, the dominance of abbreviations continued when looking at the frequency of use. Neologisms where abbreviation was used were by far the most often used category in the data with an average of 9.8 uses per word. However, neologisms where clipping was used (3.9 uses on average) were not used as often on average as words formed with new affixes (7 uses on average), conversion (5.7 uses on average) or semantic shift (5.1 uses on average). The nonce expression loan word *kekekeke* was only used once and blends also had an average of 1 use per word. The averages of use per word are best shown below in figure 2.

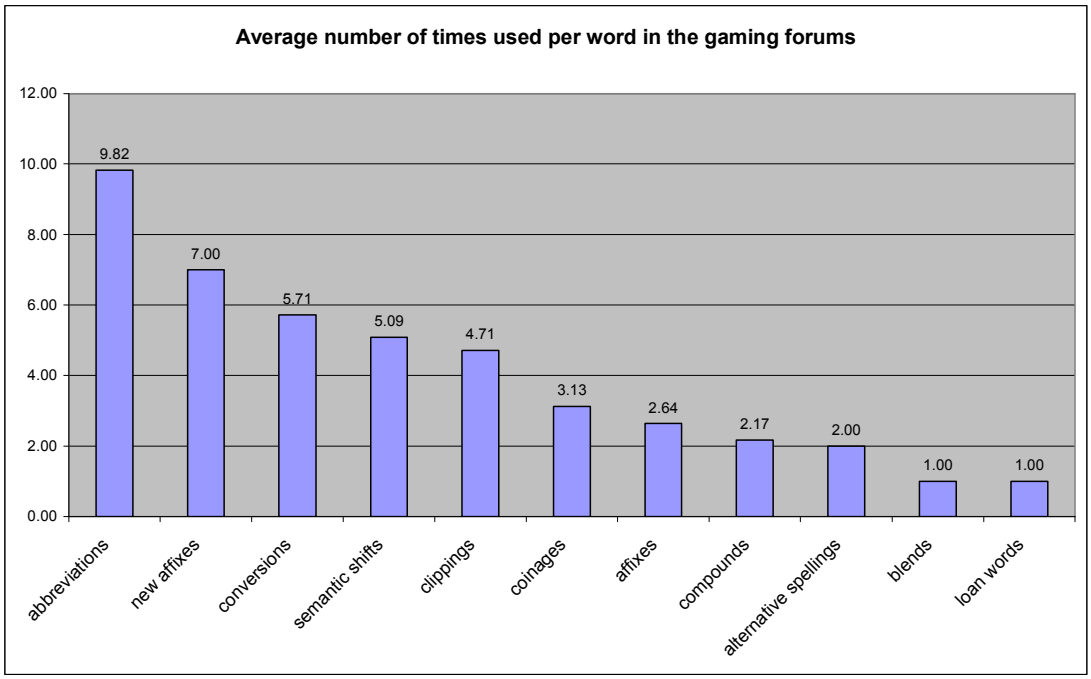

Figure 2. The average number of times used per word in the gaming forum.

5.1.2 Findings on the gaming subforum

As mentioned above, there was only one loan word found in the entire study and it was the word *kekekekeke*, which translates to 'hahahaha' for laughter. It appears to be a loan from a Korean onomatopoeic expression, which has then been written in the Latin alphabet.

When looking at the neologisms found in the gaming subforum data, it was apparent that there was a trend of creating abbreviations from existing game content words. The trend was, quite simply, to create new abbreviations by abbreviating any given word to its initial letters. This trend was however regulated (although not intentionally) in such a fashion that two letters rarely had two different meanings. It was also noticeable that in cases where a new abbreviation would have collided with a pre-existing abbreviation, the new form was clipped instead. For example when the weapon *heavy bolter* was abbreviated to *HB*, the unit *Howling Banshee* was not referred to as *HB* but as *shee* instead. A very clear example of this regulation is noticeable with the creation of neologisms for the units in the *Orks* faction. Many of their units,

such as *Shootaboyz, Sluggaboyz, Stikkbommaz* and *Stormboyz*, would have had to be abbreviated to *SB*. In order for a discussion on them to retain intelligibility beyond that of "SBs are clearly superior than SBs, however SBs and SBs beat both the SBs and the SBs", the words were clipped instead, resulting in forms such as *sluggas, shootas*, *stikks* and *stormies*. In situation where the context was different enough to separate between to identical abbreviations, such as in the case of *PM* for both 'private message' and 'plague marine' or *OP* for both 'overpowered' and 'original post'. A similar case can be seen with the clipping *mod* which refers both to 'a modification on the game' as well as to 'a moderator on the forums'. As the meaning of the clipping is easy to determine depending on what kind of context they are used in, the same abbreviation is widely used for both with very little risk of confusion.

Conversion of words from nouns to verbs with a semantic shift was a fairly common phenomenon in the game-specific forums. Words such as *buff* (verb) *ninja* (verb), *nerf* (verb), *nade* (verb) and *spike* (verb) were found throughout the data alongside their noun counterparts.

Another practice used were clippings where all vowels were removed from the original word. Examples of such words were *cld* for 'could', *cldn't* for 'couldn't', *dmg* for 'damage', *fckn* for 'fuckin' and *lvl* for 'level'. *Plz* for 'please' could also be considered a word from this category. This practice is in keeping with the general idea of using as few characters in language as possible. The practice was used in cases where the original word was relatively easy to determine from the resultant clipping. For example, if one compares the word in the data, *cldn't*, to a made up word such as *nblvbl* for 'unbelievable', it is clear that not every word could be clipped in such a manner while still retaining legibility. Some characteristics for such words can be observed from the words found in the data. First of all, it seems that the ratio of vowels to consonants in the original words should be at least 1:1 (in *damage*) and preferably in favour of the consonants (*could* has a ratio of 3:2 while *couldn't* has a ratio of 5:2). Secondly,

the original word was also fairly short, ranging from five-letter words such as *could* and *level* to a maximum of six in *couldn't*, seven if counting the apostrophe. Attempting this clipping with very long words could quickly result in illegible words, like the example with *nblvbl* showed. Thirdly, all such words had a consonant as the first letter. This is only logical as changing the first letter of a word would make it much more difficult to deduce the meaning of the word.

5.2 Findings on the data gathered on the general discussion subforum

In this section, I will present the findings of the analysis of the data gathered on the general discussion subforum. First I will go through the numbers: the number of neologisms found on the general discussion subforum, the types of word formation processes used on the neologisms as well as the number of times these neologisms were used. After the numbers, I will explain some of the trends and language practices found on the general discussion subforum such as the use of clipping with user names.

5.2.1 The numbers in the general discussion subforum

The total amount of neologisms found in the general discussion data was 92. Abbreviation was by far the most common word formation type used and it was used in 45 unique neologisms in the data. The group with the second highest number of neologisms was clippings with 16 neologisms and the third was shared by alternative spelling and semantic shift with 10 neologisms each. The type frequencies for the other word formations were coinage with 9 words, conversion with 4 words, compounding with 3 words, blending with 2 words and affixing with a single word found. Borrowings, reduplications or new affixes were not found on the general discussion subforum. Out of the 92

neologisms in the data, 84 were formed using a single word formation process and 8 were formed using two word formation processes. The numbers and percentages can be seen below in table 3.

Table 3. Type frequencies of word formation types used in the general discussion subforum.

Word formation type	Number of uses	Percentage
abbreviations	45	45.00 %
clippings	16	16.00 %
alternative spellings	10	10.00 %
semantic shifts	10	10.00 %
coinages	9	9.00 %
conversions	4	4.00 %
compounds	3	3.00 %
blends	2	2.00 %
affixes	1	1.00 %
TOTAL	100	100.00 %

As shown above, abbreviations amounted to 45 per cent of the total while clippings, almost half of all the neologisms in the general discussion subforum. The following four word formation processes all fit within 7 per cent of one another, which is shown in figure 3 below. The figure also highlights how the remaining four formation types, conversion, compounding, blending and affixing combined only amounted to 10 per cent of all the neologisms.

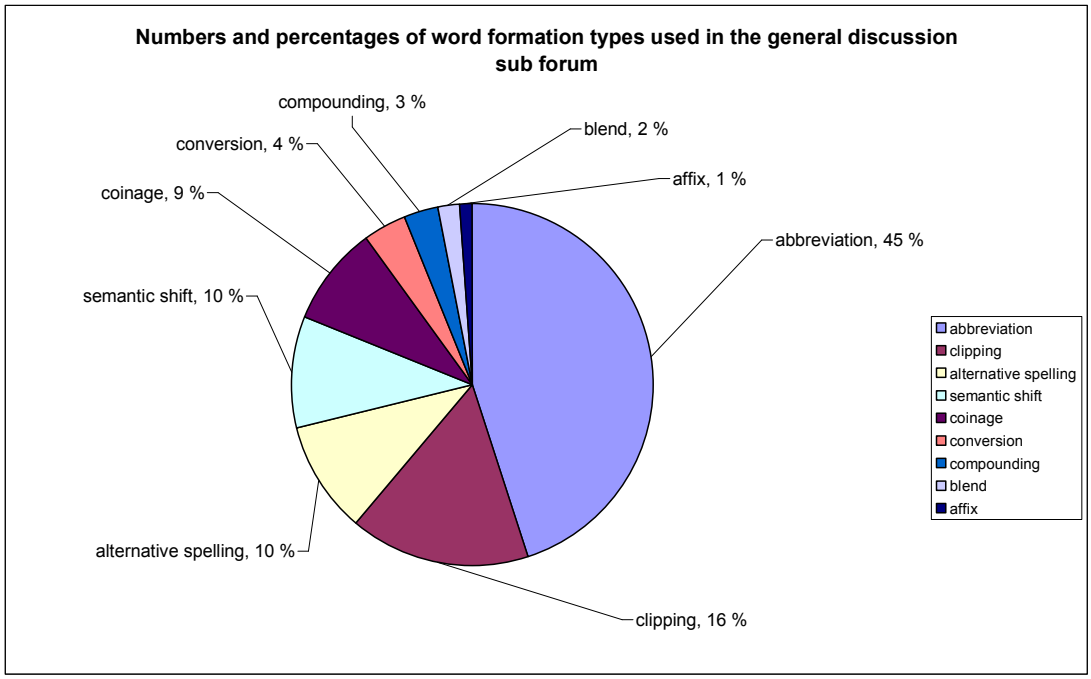

Figure 3. . Percentages of word formation types used in the general discussion subforum

When looking at the frequencies of use, the order remained largely the same with abbreviations and clippings together constituting 71 percent of the total number. While words with abbreviation were used on average two times per word in the data and cons, clippings had by far the highest average in the data; 3.5 times per word on average.

Table 4. Number of word formation types used and the average times of use per word in the general discussion subforum.

Word formation type	Token Frequency	Average of uses per word
clippings	56	3.50
abbreviations	92	2.04
blends	4	2.00
semantic shifts	19	1.90
alternative spellings	14	1.40
coinages	12	1.33
affixes	1	1.00
compounds	3	1.00
conversions	4	1.00
Average times used all words	-	1.69
TOTAL	205	-

When looking at the average times of use per word, clippings stand out above the others at 3.5 uses per word on average. Abbreviation, blending and semantic shift were the next two most often used formation types with averages of 2.04, 2.0 and 1.9 respectively. The remaining formation types and their averages were: alternative spelling with an average of 1.4, coinage with an average of 1.33 and compounding and conversion with both an average of exactly one use per word. The averages are better shown below in figure 4.

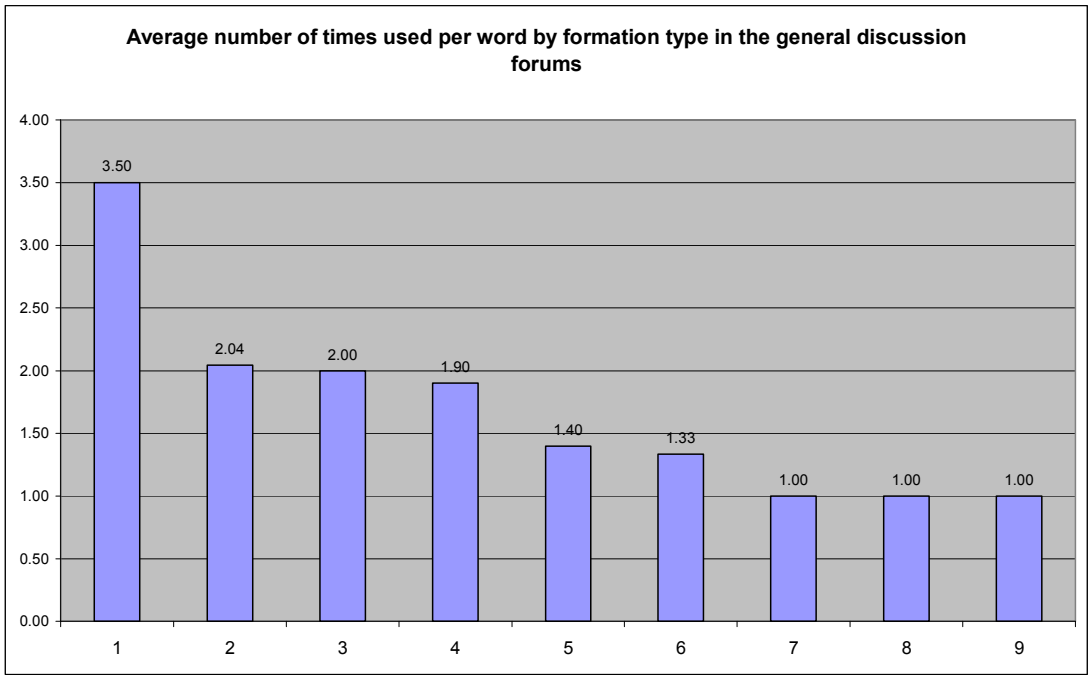

Figure 4. The average number of times used per word by formation type in the general discussion forum.

5.2.2 Findings on the general discussion subforum

One interesting feature that was observed in the general discussion data was the practice of shortening usernames of the people posting on the forums. The shortenings were most often clippings from the full usernames, such as *Vaarok* -> *Vaar*, *Octopus Rex* -> *Octo* or *Rex*, *Aron_DeTomado* -> *Aron*, *GodisanAtheist* -> *Godisan*, *Mac_Bug* -> *Mac* and *SquidDNA* -> *Squid*, however there were also abbreviations found in the data (*Troubleshooter* -> *TS* and *TheDividedGod* -> *TDG*). A particularly interesting case was the username *n0z3k1ll3r*, which, most likely due to having been typed in 133t, had four different types of clippings: *n0z*, *n0ze*, *noz3* and *no0z3*.

Many of the neologisms in the general discussion subforum data were related to technology, despite the fact that technology as such is not the main discussion topic of the subforum. Most of the technology related neologisms such as *CG* (Computer Graphics), *CS* (Computer Science), *DB* (Database), *dpi*

(Dots Per Inch), *HD* (High Definition) and *ISP* (Internet Service Provider) were also abbreviations.

5.3 Differences between the two sub forums

In this section, I will present the differences between the two subforums based on the analysis of the data. First I will discuss the differences in the vocabulary used in the two subforums. I will examine the size of the vocabulary by looking at the data from the data analysis tool. I will then move on to the differences in the amount of different neologisms used and the word formation processes used in creating them. Finally I will look at the frequency of use of the neologisms used in the two different subforums and the possible differences between them.

The total number of unique character strings found in the general discussions data was 11 281 and the total number of unique character strings found in the game specific data was 10 434. These numbers include all unique strings of characters. The term *unique character strings* is used here because these numbers include all unique strings of characters, which means that if the text included for example the character strings *IMO*, *IMO!* and *(IMO)*, each of these would count as a unique string of characters for the data analysis tool. These numbers are taken from the results of the data analysis tool using the "count everything as words" settings. For comparison, using the "count pure words only" setting, the numbers are 7113 on the general discussion forum and 5804 on the game specific forums. If this setting is used, the general discussion forum again holds a significantly larger vocabulary pool; however, the difference between the two is significantly increased. While with the "count everything" setting the general discussion pool is roughly 8 per cent larger than the gaming specific pool, the "pure words only" setting gives the general discussion forums approximately a 23 per cent larger vocabulary pool than the gaming specific forum. Regardless of the setting used, it is clear from the data that the general discussion forum

had a more varied vocabulary pool than the gaming specific forum. While I can not name a single and absolute reason as to why this is, I would hypothesize that the broader spectrum of topics would account for at least part of the reason for why the number of different words found was so much larger.

Despite the broader vocabulary on the general discussion subforum, the number of neologisms found was much higher in the gaming specific forum than on the general discussions forum. So much so, in fact, that the number of neologisms found on the gaming subforum outnumbered the number of neologisms found on general discussions more than three to one. This would seem to indicate that despite being given a narrower spectrum of topics to discuss, the gaming specific forum was much more active in the use of neologisms than the general discussion forum. This would also give credence to the idea that the topic of the discussion does in fact influence the number of neologisms used.

Comparing the average times of use per neologism between the two subforums, one can see that the total averages between the two are in favour of the gaming subforum. The total average on the gaming subforum was 4.02 whereas on the general discussion forum the total average was only 1.69. Looking at specific formation types, abbreviation, new affixes and conversion were the formation types which produced the most often used neologisms in the gaming forums with averages of 9.82, 7 and 5.71 uses per word on average each. Contrasting these to the general discussion forums, the respective averages for those formation types were 2.04 for abbreviation, 1.0 for conversion and new affixes were not used at all in the general discussion forums. The general discussion forums on the other hand preferred clipping above others (an average of 3.5 uses per word), with abbreviation and blending producing the most oft used neologisms (with averages of 2.04 and 2.0). This means that there was variance in the creation of neologisms both in terms of how often the neologisms were

used but also in terms of which word formation processes were used to create those neologisms.

Looking at the differences, they confirm the second hypothesis put forward in section 4.1 that neologisms would be used more in the gaming forum. The first hypothesis about abbreviations and clippings being the most often used word formation types is also supported by the data as those two were the most commonly used word formation types in both subforums.

5.4 Further points of interest in the forum language use

This section will discuss in greater detail specific findings encountered when analyzing the data. These findings range from single neologisms to more general features of language use encountered with a multitude of neologisms. Some of these findings encountered were: the difficulty of determining the etymology of the words, the fine line between corruptions and misspellings as well as words that are similar in both orthography and meaning, quoting habits of forum posters impacting the frequency of neologisms used. These observations give great insight into the way language is used on the forums and they are perhaps the most important content of this study in terms of understanding the language used as they reflect the characteristic features of the language better than numbers can.

5.4.1 Nuances in meaning

The word pair *lolzy* and *lulzy* was a challenging one. The challenge was determining a possible difference in meaning between the two as *lolzy* appears to be formed from *LOL*, the well known abbreviation for *laughing out loud*, whereas *lulzy* shares some etymology as it seems to have been formed from the word *lulz*, which itself is a corruption of the abbreviation *LOL* and means the

act of laughing. The use of *lulz* is most commonly in the phrase *for the lulz*, which means that something is done just for the ensuing laughs, or so that you can LOL, so to speak. *Lulz*, however, also has this quality to it that the laughs are not simply amusing but also epic in stature somehow. This difference is difficult to explain precisely, but one could try to explain the difference by saying that while *LOL* would be the appropriate response to an amusing joke, you could get *lulz* if you somehow managed to, for example, defeat an overwhelmingly superior army with a single lowly worker unit by luring them onto a frozen lake and have the ice break under the vast army's weight. In essence then, *LOL* would see use with regular amusement, every day jokes and casual laughter, but *lulz* would be reserved for rarer situations. With the meaning of *lulz*, it is possible to determine the meaning of the word *lulzy* in the phrase "*Which is lulzy if you manage to spot it and have some hard hitting AoE nearby*", that is 'something that gives you lulz'. With all of that in mind, when looking at the word *lolzy*, it bears great resemblance both to the original *LOL* as well as the derived terms *lulz* and *lulzy*. The difficulty then arises when one has to determine whether these two should be considered alternative spellings of the same word, or whether the slight difference in meaning between *lulz* and *LOL* is present in the words *lulzy* and *lolzy*.

In some cases words found in the data were very close in meaning to pre-existing words but were used in a way that was not exactly the same as the original meaning of the word. In such cases, the difficulty rose in determining how much the meaning of a word needs to change in order to count as a neologism. For example, the word *murder* was found in the data to mean two slightly different things. Firstly, it was used with the meaning of 'killing other units easily', which is fairly close to the original meaning of the word. Secondly, however, the word found use in the meaning of 'defeating another player', which is moving a little further away from the original meaning of killing someone. It still retains some sense of the original meaning and it has already seen similar use in the context of sports where one team is said to completely murder another in football, for example. Another similar word was *butcher* and

in the phrase "*Low health is a major drawback with all the butchers out there*". The original meaning of the word is, of course, the trade of cutting live stock for meat and the meaning has also been expanded to include someone who kills in an indiscriminate manner (OED). Here the word was used as a noun for a unit that is very efficient and quick at killing other units in the game. On the surface, the meaning seems very close; in both instances the word refers to something that is proficient at killing things. However, as every unit in the game is proficient at killing other units, there has to be some difference in meaning here which separates these *butchers* from other units in the game. Determining the meaning in this case was fairly easy as the sentence provided almost a full explanation of the meaning. If low health (meaning 'small amount of hit points') is a problem due to the abundance of *butchers* then *a butcher* must be 'something that can quickly reduce an enemy units' hit points'. That being the case, I believe that the difference in meaning between how the word is used versus its original meaning is substantial enough that the word could be counted as a neologism with the formation type of semantic shift. Contrasting the word *butcher* to the word *murder,* I do not think that the change in meaning with the latter is significant enough for it to be considered a neologism. This is clearly a matter of interpretation and another analysis could have deemed neither of the two as neologisms while another might have included both as viable neologisms. In cases such as these there was no straightforward way of determining when the change in meaning was significant enough and the decision had to be made purely on a case by case basis.

One aspect that came up was the concept of onomatopoeia, which was present with the words *kekekeke* (meaning laughter, *hahaha*) and *pewpewpew* (meaning the sound of firing a laser gun -> 'to shoot'). While it is not categorized as a word formation type, it is nonetheless worth noting that onomatopoeia played a part in the creation of several words in the data. Such words were categorized under coinages in the results.

When dealing with semantic shift, the cause of most deliberation was determining how much shift could be considered enough for the word to be considered having a new meaning. For example, the words *snare* (noun) and *rush* (both noun and a verb) were both found in the data and they both could be considered words where semantic shift had taken place. Starting with the word *snare*, it was found in the data referring to abilities within the game that allowed them to immobilize vehicle units. The word *rushing* on the other hand was used to refer to a particular type of high-risk high-reward tactic where resources were used to get something as fast as possible, be it many military units or a new technology level, at the expense of other areas, such as economy or keeping up with technological advances. For example one could do a basic rush where they would forego economical development and instead invest in a mass of early units and try to overrun the opponent before their investment in their economy could compensate for the increased amount of military units. This type of a rush would get countered if the opponent could defend for long enough for their superior economy to begin replacing losses at a higher rate than the rusher could maintain the attack. Another example would be vehicle-rush, where one would forego early military (usually infantry) units and get to the second technological level so quickly that they could defeat the opponent with vehicles, which would quickly overwhelm the opponent's infantry units that did not yet have anti-vehicle abilities. The vehicle-rush would be completely countered by a normal rush by taking advantage of the lack of early military units and destroying the opponent before their vehicles became available. Now the word *snare* has been used since the 1300s with the meaning of a device that snares, ie. captures, something and the OED also mentions it as being used in figurative and allusive uses since the 1300s. *Rush* on the other hand, has no related noun meaning, however when used as a transitive verb with *into*, it has been used with the meaning of 'embarking headlong, rashly or hastily on a particular course of action' (OED, s.v. rush). Both words then have pre-existing meanings which are closely related to the new meaning and the trouble here is determining whether the difference in meaning between the pre-existing one and the new one is significant enough to warrant categorizing the

new meaning as a new word. In the case of *snare*, it was decided that as the use of the word in figurative and allusive uses was so established, the difference in meaning was not significant enough to consider it a new word formed with semantic shift. For the word *rush*, it was decided that there was a significantly big difference in meaning between "a type of tactic in a game where something is acquired early on at the cost of something else" and "doing something in a hurried manner" was enough to warrant categorizing the word as a new one.

When looking at the words *troll* (used as a verb) and *troll* (used as a noun), it was unclear which one was used first and as such it was difficult to determine which one of them should be labelled both as a conversion and semantic shift and which one just a semantic shift. Consulting the OED did not prove helpful as it offered both uses as having been first documented in 1992. Looking at the frequency of use showed that *troll* was used six times as a verb and only once in noun form. While this could be interpreted as proof that the verb use is more established and as such precedes the noun use, it does not conclusively prove this. However without other means to establish chronological order between the two, frequency of use was taken here as the deciding factor for determining that the verb was labelled with semantic shift and that the noun was labelled with both semantic shift and conversion.

In order to discuss a word found in the data, *TROLOLO*, a brief explanation of what is a meme is in order. A meme is defined in OED as "a cultural element or behavioural trait whose transmission and consequent persistence in a population, although occurring by non-genetic means (esp. imitation), is considered analogous to the inheritance of a gene". In the online environment, this most often means different types of image frames or catchphrases, which are copied and modified constantly. For example the advice animal called "philosoraptor" has a picture of a velociraptor deep in thought, and the accompanying text usually involves some kind of paradox, such as "What if Pinocchio said… / my nose will now grow" or "The sentence at the bottom is

true / The sentence at the top is false". The neologism in question, *TROLOLO*, draws its origins from the word *trolling* (meaning *posting something in order to get a reaction from other posters*) and it could be considered to have another component, *LOL*, added. The word is also related to the meme of the "trolling song", which is a video of a Soviet-era singer singing a song that repeatedly goes "tralalala-lal-lalal-lalalaa" and "lolololololoo". Due to the similarity to the word *troll*, the song was quickly labelled a trolling song and used with rickrolling (displaying a link which supposedly lead to something of interest, but instead lead to a music video of Rick Astley's "Never Going to Give You Up") attempts and other trolling activities. Combining this information about memes and the different components, we begin to understand what *TROLOLO* means. For more clarity, it is useful to look at the word in context. The sentence in which the word was used was "They retreat, but TROLOLO you teleport-chase and wipe the squad on retreat". From that, it is possible to deduce that the meaning in which the word was used was something akin to "surprise!", which ties in well with the idea of trolling and especially rick-rolling. With the meaning of the word established, it is then time to move on to determining how the word was formed and how it should be categorized. There are several possible interpretations here. The first possibility is that the word is a blend of *trolling* and *LOL*. The second possibility is that the word is a coinage. It is also possible to entertain a notion that it might be a loan, considering that the song is Russian in origin, but further inquiry into the meme reveals that the song does not in fact contain any lyrics, due to the artist himself removing the lyrics and performing the song with vocalisation alone. This limits the possible options to the first two; a blend or a coinage. While the blend is plausible, no evidence was found to support such an interpretation. Therefore it is more likely that as the word is used to refer to the meme with additional connotations to trolling. As the word refers to something new and is itself an entirely new word, it was ultimately categorized as a coinage.

5.4.2 Posting practices

The use of the quote function on the forums was a point that also required consideration. In the data, the same posts or parts of a post could be quoted multiple times within the same thread, potentially skewing the neologisms to normal words ratio as well as the frequency of certain word formation processes. For example, if a neologism created with a rare word formation process was used in an oft-quoted post, words representing that process would appear to be in more frequent use than what they actually were. On a positive note, as the quote function and its use is shared by both the posters on the gaming discussion subforum and the general discussion subforum posters, it should not make a notable difference to the results between the two subforums and as such, the results of the comparison should be relatively unaffected. Poster behaviour also added an extra consideration as sometimes the quoted parts were modified by the users in order to express an opinion of their own. For example a user could quote something like: "tacs are horribly underpowered for cost" and change *underpowered* to *overpowered* and add "Fixed." to the end of the post to make a counter point. Such edits could have included neologisms that would have been omitted if the quoted parts were removed automatically and screening them manually would have been even more time-consuming. There were three possible ways to deal with these considerations: 1) screening the data manually in order to remove the quoted parts, 2) gathering the data on a forum which allowed the user to hide the quoted parts when copying entire threads or 3) devising a script which allowed automatic deletion of quoted parts of posts once the data had already been collected. These methods were not used, however, due to the following reasons. 1) Manually going through the data would have been too time-consuming for this study, 2) the positive effects of being familiar with the forum and previous knowledge of the terminology outweighed the potential negative impacts on the study to gather the data from a different forum and 3) using a script to automatically remove the quoted parts would have required handling the data in source code form, which would have added more problems than it would

have solved, due to all the added material that would have been added when viewing the forums in source form (font, size, padding etc.).

Some words in the data had become so popular that it was, at first, difficult to see them as neologisms. Words such as *DVD* (common abbreviation for *digital video disc*) and *MB* (common abbreviation for *megabyte*) have become such integral parts of the English language that labelling them as neologisms seemed strange at first. However, when examining them closely, they were found to be relatively recent additions and they could be counted as neologisms, despite their deep integration with the language. For example, *DVD*, as a word, according to the OED, dates back to 1995 which gives it an age of 17 years. This is 3 years less than the 20-year time limit set in the definition of the term neologism for this study for when the age of a word would even begin to count against it to be considered a neologism and as such it is a perfectly good neologism for the purposes of this study. A word's popularity was not held against its neologism status and such words were counted as neologisms regardless of whether they were used all the time or hardly at all.

In a somewhat similar manner, new abbreviated coinages brought on by technological and scientific innovation and advances, the creation of something new like governmental organizations or TV-shows (such as *PCR-test* for *polymerase chain reaction test*, *CODIS* for *Combined DNA Indexing System, CSI* for *Crime Scene Investigation*) raised the question of whether these should be considered abbreviations or coinages or perhaps both. When these terms were first created, was there first something called a *Combined DNA Indexing System* which was only later abbreviated into *CODIS*, or was the abbreviation created simultaneously with the longer term. These terms also share the problems mentioned above, where they would be in common use and as such somewhat difficult to categorize as neologisms. Again, the decisions had to be made one neologism at a time as no two would share features so similar that the decision made on one would be instantly transferrable to another.

As the forums are a purely textual and informal form of communication, colloquialisms from spoken language appear on the forums in varying different written forms. To give examples in the data, one can look at the following words: *gna* for *going to/gonna*, *tho* for *though*, *altho* for *although*. Some of these words, such as *tho* and *altho*, could be categorized as homophonic literations of the written words as the –*ugh* at the end of the original forms is not pronounced. Other words in the above, such as *cos* and *coz* could be seen as clippings of the original word *because*, but they could also be considered as homophonic literations of the original form. Another point of view could even be to consider these words as just spoken language in written form and as such could not be categorized as neologisms at all. However, before taking such a point of view, one must consider that when thinking of how to write the spoken word unit *'cause*, the exact orthographical form is important in deciding whether a word can be categorized as a neologism or not. Technically, the spoken word form should be *'cause* and as such, *cos* and *coz* could be also viewed as homophonic literations of the spoken word form. Another aspect of language that must be considered here is the problem of how accurate and immutable the correspondence between the written form of a word and the spoken form of the word is. In English, especially, where the spelling of a spoken word form can vary wildly depending on what is meant, it is important to consider that the connection between a spoken word and its written counterpart is not always set in stone. As an example, one can look at the homophones *right* and *write*. If one encountered a sentence such as "I'm going to right a letter to my congressman about the unemployment in our state", it might take a few seconds, but any native speaker would quickly understand the meaning of the sentence. In such a case, the word *right* would be clearly seen as a misspelling as it is the written form of another word with a different meaning. In an online environment however, where homophonic literations are commonplace, such a form could be intentional and refer to the word *write*, relying on the context for understanding. In such a case, it could be argued that *right* could develop into an alternative spelling of *write*. In fact, the form *rite* can often be seen online as the spelling for *right* for example in "I know, rite?". As

such, it would be plausible to label *cos* and *coz* as new alternative spellings of *'cause* and include them in this study instead of dismissing them from the data as spoken language. What prevented these words from being counted as neologisms was the fact that they simply are not recent enough. While they could not be found in OED, *cos* and *coz*, even in written form, most likely are not new enough to be categorized as neologisms. With the word *gna*, meaning *going to*, there was some deliberation on whether it should be labelled as a clipping or a homophonic literation. On one hand, it does appear to have all the traits of a clipping as it is formed from a longer word by removing parts of the original word (*gonna*) while retaining more than just the initial. However in the other interpretation we again run into the problem with spoken language. When looking at how the word *gonna* is pronounced, it could considered that *gna* is a homophonic literation of a hastily pronounced *gonna*. It was decided that this word was labelled a clipping as even if the word was considered a homophonic literation, it would still clearly be a clipped form of *gonna* whereas it could not necessarily be considered a homophonic literation if it was first categorized as a clipping.

5.4.3 The bits and pieces used to make a word

The use of an apostrophe was also a point of interest in some abbreviations and clippings where one version would have an apostrophe at the end and another would not. In such cases there was some decision-making involved when choosing whether to include both as separate neologisms, whether to include one as the main neologism and another as an alternative spelling of the original and if so, which would count as the original and which the alternative spelling. Would the one that was used most frequently be considered the main neologism while the one that was used less frequently the alternative? What if both were used equally frequently? As any answers to the questions above would be arbitrary and could not be based on any actual justification other than that it was decided so, it was decided that such words would not be categorized

in terms of "main" and "alternative" but instead with both forms being categorized as an instance of the same neologism. For example, one could use forms such as *CSM's* and *CSMs* to refer to 'Chaos Space Marines' or *VP's* and *VPs* to refer to 'Victory Points'.

While it was assumed that there would be neologisms formed by affixing in the data, it was unexpected to find entirely new affixes. This was the case, however, when coming across words like *insta-gank*, *instawipe, insta-surrender* and *autowin*. After the first one or two words it became clear that there was a pattern with these neologisms, that being the use of a prefix in creating them. As affixes by themselves are not neologisms, there was not a category for them in the study. Additionally, as prefixes do not appear on their own, they would not have to be included in the results as a new word formation process. Nonetheless, these affixes are significant factor in the neologisms used in the gaming subforums and as such they will be explained in further detail here. The prefix comes from the word *instantaneous* and it is a clipped form of it. Adding it to the start of a neologism adds the meaning of happening instantly. With this in mind, *insta-gank* means 'to instantly kill', *insta-gibbed* means the same thing, as well as *insta-kill*. *Instawipe* also shares this meaning, but *insta-retreating* means 'pushing the retreat button instantly' and *insta-suppression* means 'to become suppressed instantaneously'. The other commonly used prefix is *auto-* which is a clipped form of *automatically*. It is used in neologisms such as *autowin*, *autolose* in the contexts such as "*Orks vs. Space Marines is an autowin for the Orks and an autoloss for the Space Marines*".

A form that was frequently used in many words was the form *spam*. Its use included the following words: *shoota-spam* (meaning to produce shootas to the exclusion of everything else), *spam shootas* (meaning the same as above, only this time used as a verb), *wraithguard spammers* (meaning people who spam wraithguard units), *spam-reinforce* (meaning to repeatedly reinforce one's units), *spammable* (meaning the quality of being possible to spammed), *Ork spam*

(meaning a large group of Ork units) and *spam-scale* (meaning how well the unit synergizes with more of the same units). Looking at these different neologisms, it could be argued that *spam* had varied uses. It was used as an affix in some words (*shoota-spam*, *stealerspam*, *nadespam*), it was an instance of semantic shift in others (*people just spam shootas vs. tacs for an easy win*) and as the base of an affixed word (*wraith guard spammer*). As productive a component as it was, it was never seen on its own so it was only present in the data as a part of its derivatives.

An atypical form of word formation was the case of extreme compounding, where multiple words were all compounded together. Words such as *NotWithJesusOrAmerica* or *OMGWTFCAN'TMOVEBLOWYOURWHOLEARMYUP* are orthographically single units without so much as a hyphen between them to separate the different parts from one another. Deciding whether such extreme compounds should be included was a difficult question to answer as on one hand one it is highly unlikely that such terms would be used more than once, but on the other hand nonce expressions are neologisms like any other so there should be no reason to exclude them. If one was to disqualify such neologisms based on their rather absurd length, where would one draw the line between absurd and reasonable? For example would *OMGWTFBBQ* be too long of a string? How about *WTFPWN*? And how would hyphenating impact such words, would it still count as a neologism if it appeared as *OMG-WTF-can't-move-blow-your-whole-army-up* and would it be categorized as a compound then? Other examples of character strings that caused problems of this kind were *'black/white/Hispanic/European/Asian/etc.'*, *'tell-it-like-it-is'*, *evil/conniving/lazy/stupid/fat/covetous/etc*, *zomgIamsogoingtopwnyousidewayswithIMBACHAOSRAPTORRUSH* and other long strings of words that had been written together or hyphenated. Ultimately the decision was made separately for each such string and in most cases the hyphenated cases were discarded while the non-hyphenated ones were accepted. The reasoning for this was that using hyphens or slashes is commonly

used as markers indicating a division between two separate bases that are linked together in order to explain a larger concept. When no such markers are given, it can then be assumed that the user had a reason to leave out spaces and other separation markers and chose to type the words together as a compound, which was most likely to try and give a single term to describe the events that took place instead of explaining it (albeit simply stringing the words together is most likely not the most concise way of describing such events). Another interpretation could be that where such compounding is used, there is usually a degree of urgency involved, which could then in turn impact the way the situation is described. For example with the term *OMGWTFCANTMOVEBLOWYOURWHOLEARMYUP*, it is clear from the constituent words that something in the game has happened to create a situation where one player is suddenly unable to move their units and is forced to helplessly watch as their army is destroyed. When looking at the term in context with the full sentence that it was used in, it is clear that this is indeed the case:

> (1) "They aren't the OMGWTFCAN'TMOVEBLOWYOURWHOLEARMYUP bullshit they were before, but the fact remains that they're still quite capable of killing retreating units, provided they either do it right as they retreat or intercept retreating units."

In such a situation it is conceivable that the player might blurt out something like the term above to define the situation they find themselves in. However, in this case there were individual words strung together in order to create something to describe this one specific event. It is difficult to see such a compound being used repeatedly and finally, the compound, as such was discarded. Nonetheless, breaking down the string there were still components which would qualify as neologisms. For example with *zomgIamsogoingtopwnyousidewayswithIMBACHAOSRAPTORRUSH*, the string could be broken down to the following constituents that were valid neologisms for the study: *zomg*, *pwn*, *IMBACHAOS* and *raptorrush*.

Related to the use of compounding, in some cases it was difficult to determine whether a word had been created using blending or compounding. For example, let's look at the word *hbdev*, which means *heavy bolter devastator*. This word was created by first abbreviating the words *heavy bolter* to *HB*, clipping the word *devastator* to just *dev* and then combining the two, the abbreviation *hb* and the clipped form *dev*. This results in the word *hbdev*, but was the end result reached by using blending or compounding? The question is a difficult one because the word shares qualities of both. In blending, two words are blended together by clipping parts of either one or both of the original words. In compounding, the two words are simply combined without taking anything away from either one. In this case, something has clearly been removed in order to get from *heavy bolter devastator* to *hbdev*. The problem, however, lies in determining when exactly letters started disappearing from the words. As both *HB* and *dev* are used on their own on the forums, it could be argued that *hbdev* is a result of compounding as the two words, *hb* and *dev*, are joined together without removing anything from either one. However it would be equally correct to argue that *hbdev* was formed directly from *heavy bolter devastators*, or even *hb devastators*, and then blending the two together, removing parts from either one or both. To complicate the matter even further, the form *hdev* was also found in the data, referring to the very same *heavy bolter devastators* as the form *hbdev*. In this particular case it was decided that the form *hdev* was categorized as a blend and the form *hbdev* was categorized as a case of compounding, but arguments for the opposite could also have been easily made.

In addition to the items above, which could be categorized as either compounding or blending, division into either blends or clippings was also quit challenging at times. As clipping and blending both utilize length reduction of words in creating new words there is likely to be some overlap. Bauer (1983) has noted that this particular category [blends] is not well defined and tends to blend with the other categories such as clipping, compounding, affixation and acronyms. Expanding on this problem, one could even say that all blends utilize

clipping as part of the creation process, as blending by definition involves shortening one or more component words in order to create a new form. With the formation type being so intertwined with other formation types, it is to be expected that when assigning elements of multiple formation processes to a single word, several overlapping formation types might be assigned to the same word. As an example, one can examine the word *plasdev*. The word means *plasma devastator* and it has been formed by taking the words *plasma* and *devastator*, clipping them both and then putting them together. Looking at the clipped forms of the component words, the word *dev* for *devastator* appears in the data on its own but *plas* for *plasma* does not. This leads to a possible argument that compounding can be ruled out as having been part of this word's creation process. This also lends credibility to the argument that clipping should be assigned to the word *plasdev* as being part of the creation process because *dev* is clearly a clipped form of *devastator* and as such, *plasdev* is a blend of the clipped form *dev* and the word *plasma*. In order to avoid assigning clipping to every blend found in the data, it was decided that only if a blend clearly combined a pre-existing clipped form or forms with at least one other component, which was reduced in length to create a new word, would the resulting word be labelled with both clipping and blending.

There were several minimal clippings, by which I mean words which were clipped only by a single letter or two, found in the data. Examples of such words in the data were *sik* (meaning *sick*), *loots* (meaning *lootas*) and *kno* (meaning *know*). In such cases it was not always certain whether such words were in fact clippings or just lazy writing or typos. These words fall into two different sets, ones that could be categorized as homophonic literations (*sik* and *kno*) and the ones that could not (*loots*). On one hand it could be argued that the ones that could be considered homophonic literations are more likely to be typos, due to them being more easily written incorrectly by non-native, or even native speakers, because the words are pronounced identically and if one only knows the spoken word *know*, it could easily be misspelled simply as *kno*. On the other hand, homophonic literations are commonly used on the internet, so it

could be that even if the poster was familiar with the correct spelling of the word *know*, they would use the shorter version *kno* as it would be quicker to type. As such, there is both an argument supporting dismissing them as typos and an argument for treating them as clippings which, in terms of plausibility, are equally plausible. With the words that were not homophonic literations, it could be argued that as the clipping is so minimal and there is no homophony with *loots* and original word *lootas*, it would be more likely that *loots* is just a typo. However, as there were no concrete reasons why such words could be dismissed as typos, they were included in the data as neologisms.

From the minimal clippings to the opposite end of the spectrum, words such as *d/l* for *download* and *w/e* for *whatever* began treading the line between clippings and abbreviations. The word *d/l*, which stands for *download*, is just two letters, d and l from *down* and *load*, separated by a slash. As such, it could easily be viewed as an abbreviation in the same way that *FBI* for *Federal Bureau of Investigation* or *cc* for *cubic centimetre* are. There are a couple of points worth considering here. The first point is the fact that the original word is *download*, a single word. Abbreviations are usually formed by abbreviating each individual word down to the initial letter and then combining these initials. In this case, following this formula would have yielded a simple *d* as the abbreviation for *download*. The second point to consider is the slash between the two letters. While punctuation, such as full stops in the abbreviation *U.S.A.*, is sometimes used in abbreviations, the use of a slash is very rare. However, as the original word is a compound which comprises of two individual parts, *down* and *load*, it could be argued that *d/l* would in fact be a abbreviation of the two base words. This would arguably be a more valid categorization as clippings usually involve shortening the original words at one end, whereas in these cases only the initial words were preserved. Another example of such a word was *w/e* for *whatever*, which is also a two-part word that has been reduced to the initials of the original two parts. These words were ultimately labelled as abbreviations.

Finally a feature seen on the gaming forums that was not reviewed in this study was the use of emoticons. Emoticons are strings of letters, numerals and other characters that express facial expressions. Most common (Western) emoticons use either a colon or a semicolon to indicate eyes (colon for regular eyes, semicolon for winking) and a multitude of other characters to indicate mouth, tongue and nose. The Western style emoticons are tilted 90 degrees, so they in order to understand them at first, one needs to tilt one's head to the left. Common examples of Western style emoticons are :-) for smiling, :-p for showing your tongue, : - D for laughing and : - O for surprise. With emoticons, it needed to be decided whether they should be categorized as words. Among the data, the first character-based emoticon found was *oO*, which at first seems like an abbreviation as it was added from the data analysis tool, which capitalized the first o, making it appear as *OO* instead. This particular emoticon, *oO*, is a Japanese-style emoticon and indicates confusion and wonder by resembling one's eyes with one eye-brow raised. The Japanese style emoticons are, contrary to the Western style emoticons, horizontally level, so they can be "read" without tilting one's head. Examples of Japanese-style emoticons are (^_^) for smiling, (O_O) for surprise and (^_~) for winking. Often times the underscore representing the mouth is left out, as well as the parenthesis representing the edges of a person's face. Another expression which appeared to be an emoticon was *QQ* which could be interpreted as a Japanese-style emoticon used to indicate crying (and often times it is in fact now used to express crying). Studying the etymology for this word, however, reveals that originally its meaning was not *to cry* but instead was related to exiting a game and in fact it was not an emoticon to begin with. The etymology for the term is that in the game *Warcraft II* (released in 1995), one could press the alt key and then the Q key twice to exit, or quit, the game in progress, so the term *QQ* became used for quitting. When the original meaning was used in phrases like *what are you going to do about it? QQ?*, the term started being associated and used with the new meaning of crying, even though the original meaning was something different. The problem with emoticons was whether they could be categorized as neologisms for the purposes of this study in the first place and if

so, what word formation type would they be counted as. Coinage is, in my opinion, the only formation type that one could categorize emoticons under but ultimately they were not included in this study as neologisms. Nonetheless, as *QQ* is not an emoticon, it was included as a neologism and categorized under coinages. *oO* was not included as a neologism in this study.

6 CONCLUSION

First of all it seems right to begin the conclusion by going over the research questions of the study and answering them. As such, the first research question was:

"What are the word formation processes used in creating the neologisms used on a gaming subforum and a general discussion subforum?"

In order to answer this question, we can list the word formation types used in each subforum in order of productivity. For the gaming subforum, neologisms were created with the following word formation processes: abbreviation, clipping, semantic shift, compounding, affixing, alternative spelling, conversion, blending, coinage, using new affixes and borrowing words. In the general discussion subforum, neologisms were formed using the following word formation processes: abbreviation, clipping, alternative spelling, semantic shift, coinage, conversion, compounding, blending and affixing.

The second research question was: "Are there differences in the frequency of use of neologisms on different subforums based on the topics of discussion?"

The short answer to this question would be a resounding yes. To expand on that answer, one could say that neologisms were used much more frequently in the gaming subforum with the average times used per neologism being over two times as high in the gaming subforum than in the general discussion forum. In terms of individual word formation processes, abbreviations in the gaming subforum were used almost three times more frequently than any

neologisms in the general discussion subforum. The only individual word formation type to produce words of more frequent use in the general discussion subforum than in the gaming subforum was blending with an average of two uses per word compared to an average of only one use per word.

The third and the final research question of the study was: "Does the topic of the discussion influence the use of neologisms on internet forums?" and the answer to that was left somewhat open. While in this study the evidence does support the idea that the topic of discussion influences the use of neologisms, the comparison was only made between two different topics. However, the data from the gaming subforum had over three times as many individual neologisms as the general discussion subforum and those neologisms were also used more often by more than two times as often. As the differences are so significant, they could be taken as evidence that if the sample size was increased, the results between different topics would be different as well. Thus, for the purposes of this study, it can safely be stated that based on the results of this study, the topic of discussion does influence the use of neologisms on internet forums in such a manner that if the topic is internet gaming, a greater variety of neologisms will be used and with greater frequency than if the topic is a less specific one such as politics or movies.

In hindsight, one can usually see room for improvement in how the study was conducted and that was the case here as well. One of the things that should have been done differently in this study was the method of data analysis used. In this study, the data was processed by hand, reading through the data and manually writing down the neologisms found in the data. This was both highly time-consuming as well as prone to human error. In order to hasten the process of data analysis, it would have been preferable to use a more advanced analysis software that would have enabled the use of more advanced filters. By being able to apply customized filters, strings of characters such as "*able* or *(disregarding* could have been listed under *able* and *disregarding*, removing the

need to sift through essentially the same words twice. If a study such as this one is attempted in the future, it would be highly advisable to use an already existing piece of software for the data analysis. If the use of such a program is, however, not possible either due to expenses involved in acquiring such a program, limited access to such software or simply inexistence of such software, it would be advisable to elicit the help of a capable programmer to program a script to automatically process the data.

The data gathering method used in the study could also have been improved by removing the time stamps, quotes and other miscellaneous remnants of the forum software functions which added unnecessary clutter to the data. Without this clutter, the data would have been more concise, screening the neologisms from mundane words would have been faster and more raw data could have been gathered and used within the same time constraints. Without this clutter, the results would have also been more precise and as a single mention, analyzing the word *PM* would have been much easier as that was one of the words which appeared in the remnant text and as such, analyzing its frequency in the actual data was extremely difficult as it was necessary to screen through hundreds of false uses of it in the data.

Naturally there can never be too much data to study, other than when its analysis becomes too time-consuming. As such, additional data from different sources, be it from other websites or other subforums of the same website, could have been used and with more resources available to a study such as this, should be used. This would provide more reliable results and could possibly reveal further differences in how the topic affects the vocabulary used. Perhaps sports attracts more coinages, or perhaps gardening involves more Latin borrowings. In order to determine whether the results of this study represent language use on this one forum, or further yet just two subforums of this one forum, more data would have been needed. An excellent addition would have been another subforum from the same forums, for example one dedicated to a

different game. There were a host of such subforums available on the Relicnews forums; subforums for a WW2 era RTS (Real Time Strategy) game, an FPS (First Person Shooter) in the same sci-fi setting as the game in this study's data and another RTS game focused on a very different sci-fi setting. Also available were subforums for table-top gaming, computer hardware and software topics, artistic interests, user introductions and tech support. Gathering data from for example the FPS game subforum and the other RTS subforums would have given valuable information about whether the differences between the general discussion subforum and the gaming subforum were a result of the gaming vs. non-gaming division or if for example there were differences between FPS gaming and RTS gaming, between a sci-fi setting and a WW2 setting or between general discussion and artistic discussion.

When thinking of how and what further study should be conducted, the first thing to consider is obviously to look at what the study failed to explain and what questions were left unanswered. In this vein, the questions that remained unanswered in this study were at least the following: why were the results between the two subforums so different and if the sample size had been increased to include the entire subforums, would the results have remained the same. There is a clear method to solving at least the second of those questions and that would be to simply increase the sample size. That would however require a much greater effort than what the scope of this study was, but nonetheless the results would be interesting. As for solving the reasons for the differences between the two subforums, a detailed explanation would most likely require input from experts in organizational communication, but as Hatch and Brown (1995, 210) reported, some fields make greater use of certain types of neologisms and the full answers would most likely be something in that vein.

One interesting aspect worth researching is the longevity of neologisms formed on the internet and whether the neologisms found on the general discussion

subforum last longer in use than their counter parts from the gaming subforum. This would require a long term research which would either begin by charting out the neologisms currently in use and then doing periodical checks on which of the words are still in use and which have fallen into disuse and can not be found on the forums in newer threads and posts. Another way to map out the longevity of neologisms would be to first chart out the neologisms used, then do a survey of the user base of the forums and then poll the user base on which of the neologisms they recognize. It might be the case that while there are a great many neologisms found on the forums, many of them are nonce expressions which are not even recognized by majority of the forum users. This could very well be the case with many alternative spellings, as they could easily be simple typos.

Another possible avenue of future research would be to do an analysis of the neologisms on the forums based on how many of the neologisms used are actually created on the forums. My estimate would be that out of the neologisms used on the gaming forums, at least half would have been created on the forums. On the general discussion forums, I would estimate that practically none of the neologisms in use, apart from the abbreviated and clipped usernames, would have been created on the forums. This would be an important area of research as it would help determine whether the increased use of neologisms in the gaming forums is a result of the linguistic productivity of the users there, or whether the increased numbers of neologisms are seeping in from other aspects of language use related to gaming. A complication with this area of study would be the difficulty of tracking down exactly where a word is created. In order to do this study, it would be necessary to follow a game from the start and monitor both the forums as well as the in-game communication between players. The best method to achieving this would be to co-operate with a game developer in order to get access to the early beta development process of a game, where the player pool would still be relatively small and the amount of data would be more easily manageable. Ideally one should also get logs of the in-game chats automatically submitted into a

database with an automated system to log new expressions within. Similar software for monitoring the forum environment would also be highly beneficial.

In order to gain a better understanding of how topic affects word formation, one should do a comparative study between several different gaming forums. As mentioned in chapter 6.2, valuable information would be revealed by studying first of all different subforums of these same forums and secondly entirely new forums and comparing the results. The study of other subforums would be important because there the effects of individual language use could be minimized. When the users remain the same but the topic of discussion changes, then it is more probable that the changes in vocabulary and word formation are due to the effects of topic and not from different people using different vocabulary. In order to completely eliminate the effects of individual language one would need to do a controlled experiment where the researchers started threads on different topics and the participants in a study would be asked to reply in each of the different threads, preferably in equal length. Doing this would ensure that the differences would not be due to individual differences as the participants in the threads would be identical. This, however, would mean that the experiment would be quite artificial and as such it would not be a sample of natural language use.

In conclusion, this study examined how neologisms are formed and used on internet forums and what differences there are between two subforums with different topics of discussion. It was discovered that the gaming forum saw a greater variety in neologisms and those neologisms were used more often. While this type of research, categorizing words and word formation types, can be seen as an older form of internet language research, I believe that there is still a need for such basic level research to be conducted alongside other modes of research, such as neologism longevity research and communicational research with regards to non-text-based modes of communication like VoiP.

7 BIBLIOGRAPHY

Adams, V. 2001, *Complex Words in English*. Harlow: Pearson Education Limited.

Algeo, J. "Where do all the Words Come From?", 1980. *American Speech*., Vol 55, No 4, 264-277. Duke University Press.

Androutsopoulos, J. 2006. *Introduction: Sociolinguistics and computer-mediated communication*. Journal of Sociolinguistics, 10(4), 419-438. (19 Apr 2012) http://search.ebscohost.com.ezproxy.jyu.fi/login.aspx?direct=true&db=ufh&AN=21980034&site=ehost-live

Aronoff, M. 1976. *Word Formation in Generative Grammar*. Cambridge, Massachusetts: MIT Press.

Barnes, S. B. 2003. *Computer-Mediated Communication : human to human communication across the Internet*. Boston (Mass.):Allyn and Bacon.

Bauer, L. *English word-formation*, 1983. Cambridge: Cambridge University Press.

Cheater, A. P., "Beyond meatspace – or, Geeking out in e-English." *English Today* Vol 22. No. 1., 18-28.

Crystal, David. 2001. *Language and the internet*. Cambridge: Cambridge University Press.

Crystal, D. 1997. *English as a Global Language*. Cambridge: Cambridge University Press.

Driscoll, D. 2002. *Ubercool Morphology of internet Gamers: A Linguistic Analysis*. Undergraduate Research Journal for the Human Sciences, California University of Pennsylvania (14 Dec 2011) http://www.kon.org/urc/driscoll.html

Dugan, L. 2012. "Twitter to Surpass 500 Million Registered Users On Wednesday." *AllTwitter*, 21 Feb 2012. (2 July 2012) http://www.mediabistro.com/alltwitter/500-million-registered-users_b18842

Hachman, M. 2012. Facebook Now Totals 901 Million Users, Profits Slip. *PCMag.com*, 23 April 2012. (2 July 2012) (http://www.pcmag.com/article2/0,2817,2403410,00.asp

Harley, H. 2006. *English Words: A Linguistic Introduction*. Oxford: Blackwell Publishing.

Hatch, E. and Brown, C. 1995. *Vocabulary, Semantic, and Language Education*. Cambridge: Cambridge University Press.

Herring, S. C. 2002. "Computer-mediated communication on the internet." *Annual Review of Information Science and Technology*, 36, 109-168. (20 Apr 2012) http://ella.slis.indiana.edu/~herring/arist.2002.pdf

Herring, S. C. 2004. "Slouching toward the ordinary: Current trends in computer-mediated communication." *New Media & Society*, 6 (1), 26-36. (11 Apr 2012) http://ella.slis.indiana.edu/~herring/slouching.pdf

Herring, S. C. 2008. "Language and the internet." In: W. Donsbach (Ed.), *International Encyclopedia of Communication*. Blackwell Publishers. (17 Apr 2012) http://ella.slis.indiana.edu/~herring/lg.inet.pdf

Herring, S. C., D. O. Kutz, J. C. Paolillo and A. Zelenkauskaite 2009. *Fast talking, fast shooting: Text chat in an online first-person game. Proceedings of the Forty-Second Hawai'i International Conference on System Sciences (HICSS-42)*. Los Alamitos, CA: IEEE Press. (Apr 18 2012)
http://ella.slis.indiana.edu/~herring/hicss.bzflag.pdf

Herring, S. C. 2010. "Computer-mediated conversation Part I: Introduction and overview". *Language@internet, 7*, article 2. (11 Apr 2012)
http://www.languageatinternet.org/articles/2010/2801

Herring, S. C. 2011. In S. C. Herring, D. Stein, & T. Virtanen (Eds.), *Handbook of pragmatics of computer-mediated communication*. Berlin: Mouton. (11 Apr 2012) http://ella.slis.indiana.edu/~herring/commentary.2011.pdf

Jackson H. and E. Zé Amvela 2001 (2000). *Words, Meaning and Vocabulary : an introduction to modern English lexicology*. Trowbridge: The Cromwell Press.

Kalima, J. 2007. *Word formation on Internet Gaming Forums*. University of Jyväskylä, Department of Languages. (24 Apr 2012)
https://jyx.jyu.fi/dspace/bitstream/handle/123456789/18948/Kalima_Juhamatti_kandi.pdf?sequence=1

Katamba, F. *English Words. Structure, History, Usage*. 2005. London: Routledge.

Maybin, J. and J. Swann (eds.) 2006. *The art of English: everyday creativity*. Basingstoke: Palgrave Macmillan in association with the Open University.

Plag, I. 2003. *Word-Formation in English*. Cambridge: Cambridge University Press.

Weiner E. S. C. 1983. *Oxford guide to English usage*. Oxford: Oxford University Press.

YouTube. (2 Jul 2012) http://www.youtube.com/t/press_statistics/?hl=en-GB

Yule, G. 2006. *The Study of Language*. Cambridge: Cambridge University Press.

Appendix 1: A list of neologisms from the general discussion forum

AFAIK	as far as I know
arcadyness	arcade -> arcady -> arcadyness
aron	Aron_DeTomado (username)
ATM	At The Moment
B5	Babylon 5
batshit (crazy)	mad
BTW	By The Way
CG	Computer Graphics
CS	Computer Science
cybersurfer	someone who frequents the internet
DB	Data base
dems	democrats
derp	EXPLAIN
diss	disparage someone
doc	document
DoD	Department of Defense
dpi	Dots Per Inch
DVD	Digital Video Disk
emote	emoticon
FFS	for fuck's sakes
FTW	For The Win
FWIW	For what it's worth / for whoever is wondering
GD	General Discussions
gitmo	Guantanmo bay
Godisan	GodisanAtheist (username)
google-fu	one's ability to use the Google online search engine
GWOT	Global War On Terrorism
hacking (v)	gaining access to a network / databse without authorization
HD	High-definition
high-res	high resolution
HS	high school
hurr	laughter
hype	build-up of excitement prior to something
IIRC	If I remember Correctly
IMHO	In My Honest Opinion
IMO	In My Opinion
InsanoBitch	Insane bitch
interwebs	internet
ISP	Internet Service Provider
IT	Information Technology
JPG	Join Photographic Experts Group
lgbtf	Lesbian Gay Bisexual Transgender and Friends
lol	laughing out loud
lulz	laughs
Mac	Mac_Bug (username)
mcjob	low-paid job with few prospects
meh	expression of indifference
MMORPG	Multiplayer Massively Online Role Playing Game
n0z3	n0z3k1ll3r (username)
NDIS	National DNA Indexing System
ninja	to ninja-post, to post a comment while another was writing their own
NotWithJesusOrAmeric	not with Jesus or America (summarising a politician's stance)

a	
NS	NoSurrender (username)
Nuri	Nurizeko (username)
Octo	Octopus Rex (username)
OMFG	Oh My Fucking God
OP	Original Post
OST	Original Soundtrack
PDA	Personal Digital Assistant
POC	Proof of Concept
ppl	people
pron	porn
reach-around	sexual act
redonkulously	ridiculously
Retro	Retroboy (username)
Rex	Octopus Rex (username)
RL	real life
roflmao	rolling on the floor laughing my ass off
sand 'N	sand nigger
scissor	sexual act
screencap	screen capture
screengrab	screen capture
smilie	smilie
spoilered	put under spoiler tags
Squid	SquidDNA (username)
SSN	Social Security Number
ST	Star Trek
STR	Short Tandem Repeats
TDG	TheDividedGod (username)
troll (n)	a person who trolls (see below)
troll (v)	to elicit a response by posting antagonalistically or counter-factually
TS	TroubleShooter (username)
Vaar	Vaarok (username)
vidya	video game
w	with
w/	with
wut	what
WTC	World Trade Center
WTF	what the fuck
xbl	Xbox Live
yoof	youth
$$$	money

Appendix 2: A list of neologisms from the game-specific subforum

´net	internet
+1	agreed with the post above, signed
1 shotted	to kill in one shot
AC	Aspiring Champion
AC	Assault Cannon
altho	although
AOE	Area of Effect
Apoc	Apothecary
Apoth	Apothecary
army-wipe	destroying the opponent's entire army
arty	artillery

ASM	Assault Space Marine
ass-termies	assault terminators
ASTKNF	And They Shall Know No Fear (misspelled)
auto-	automatically
autowin	automatic win
BC	Bloodcrusher
BE	Battle Equipment
beastmode	extremely good
BL	Bloodletter
blob	big group of units in one place
blobbing	the act of creating blobs
blobby	adjectivized 'blob'
BLRs	Bloodletters
blue	requisition (the resource shown in blue numbers)
boss	good
boss	Warboss
BS	Barbed Strangler
BS	Big Shoota
BS	bullshit
BTW	by the way
bubble	an ability with the visual effect of a shining bubble
buff (v)	improves
buff (n)	an improvement, a positive bonus
bug out	not work, stop working
butcher	unit that kills others quickly
buttrape	defeat easily
camp	remain stationary
Cap Spire	Capital Spire
capper	a unit that can capture a point on the map
to cap	to capture
cat	Catachan
cc	close combat
CD	cooldown
CF	community forums
chain-heal	heal in rapid succession
chaosraptorrush	a rush strategy with chaos raptors
CL	Chaos Lord
cld	could
cldnt	couldn't
cockblock	an obstacle
CoH	Company of Heroes
configurable	something that can be configured
co-op	co-operative
CoT	chains of torment
CR	Chaos Rising
crit mass	Critical mass
critical mass	a mass of units that is so big as to be hard to beat
crusher	blood crusher
CS	Chaos Sorceror
CS	Community Site
CSM	Chaos Space Marine(s)
cult-bomb	Doom Blast
cults	cultists, meaning heretics in the game
d/l	download
dancer	a unit that is keeping away from close combat by moving
debuff (n)	the effect applied when debuffing something

debuff (v)	to reduce in effectiveness, opposite of 'buff'
decap	de-capture
dev	devastator
dev	developer
DLC	Downloadable content
dmg	damage
dodgeable	able to be dodged
DoT	Damage over Time
DoW	Dawn of War
DoW1	Dawn of War 1
DoW2	Dawn of War 2
DoWI	Dawn of War 1
DoWII	Dawn of War 2
DoW2:R	Dawn of War 2: Retribution
DPS	Damage Per Second
DPSer	a unit that does dps
dread	dreadnought
Ekko	Ekko Tek (username)
Eldritch	eldritchweather (username)
Eldritch	Eldritch Storm
Empyreal	Empyreal Abyss
Ex:	For example
exarch'd	upgraded with an exarch
eyecandy	visually very appealing
FC	Force Commander
fckin	fucking
Fex	Carnifex
ff	focus fire
FFA	Free For All
float	to have excess amounts of resources unspent
fluff	the backstory of the wh40k universe
fluffy	in keeping with the 'fluff'
fn'	fucking
focus fire	term for concentrating all your damage on a single enemy unit
FoF	Fleet of Foot
FTE!	For the Emperor!
FTW	For The Win
FX	effects
gank	kill something unexpectedly
gaunts	hormagaunts and termagaunts
gen	generator
gen bash	an attack on the opponent's generators
generator bashing	playstyle involving destryoing the opponent's generators
GFWL	Games for Windows Live
GFX	graphics
GG	Good Game
GL	Grenade Launcher
Global	global ability, affecting entire playing field
Global rep	Global repair
GM	Guardsmen
gna	going to
gu	guardian
GUO	Great Unclean One
HB	heavy bolter
HB Dev	Heavy Bolter devastator
hdev	heavy bolter devastator

heavies	heavy infantry units
heavy inf	heavy infantry
hi-jack	to hi-jack a thread by changing the topic of the discussion
HMG	Heavy Machine Gun
horma	hormangaunt
horms	hormangaunts
HotW	Hammer of the Witches
howl	Demonic Roar
HP	hit point(s)
HT	Hive Tyrant
HW	heavy weapon
HWT	heavy weapon team
IG	Imperial Guard
IIRC	If I remember Correctly
IMBA	imbalanced
IMO	In My Opinion
IMHO	In My Honest Opinion
inf	infantry
insta-	doing something instantly
IRC	Internet Relay Chat
jihadist	cultist
kb	keyboard
KCSM	Khornate Chaos Space Marines
kekekekeke	hahaha, laughter
kewl	cool
kite	to move your units around an enemy unit, avoiding damage
kno	know
knob	Kommando Nob
knock back (n)	an effect that knocks units back
knockback (n)	an effect that knocks units back
la	Lictor Alpha
lab (something) (v)	test somethin
lazcannon	las cannon
LC	Lightning Claw
Lib	Librarian
Lock	Warlock
logins	game account information
logons	acts of logging on
LOL	Laughing out loud
lolzy	LOL-worthy, funny
loots	lootas
LoS	Line of Sight
LtGB	Let the Galaxy Burn
lulz	laughs
lulzy	see above
lvl	level
manti	Manticore
mega cult-bomb	an improved cultist bomb
meh	expression of indifference
meta	metagame
metagame	the current state of the game, what strategies and units are used
micro	Micro management
micro-intensive	requiring a lof of micro management
ML	Missile Launcher
mod	moderator
mod	modification

modder	someone who makes modifications
MoK	Mark of Khorne
MoN	Mark of Nurgle
MoT	Mark of Tzeentch
MotDread	A dreadnought with the Mark of Tzeentch upgrade
MP	Multiplayer
MU	match-up
murder	defeat easily
MWB	Merciless Witchblade
nade	grenade
nade (v)	to throw a grenade at, ie. "grenade" something
Nade-spam	Grenade spam = using a lot of grenades
nerf (v)	to make weaker
nerf	a change that makes weaker
nid	Tyranid
ninja (v)	to do someting quickly and stealthily
NM	Noise Marines
Non-repair	Not to do with repair
noob	new player, derogatory
np	no problem
nuke	ability that does a lot of damage on a big area
ofc	of course
OMFG	Oh my fucking God
OMGWTF OP	Oh My God What The Fuck Overpowered
OMGWTFBBQ	very good. (Oh my God, what the fuck, barbeque)
OP	overpowered
OP	original poster
Opness	overpoweredness
ORLY	Oh really
pathing	pathfinding
PC	Plague Champion
PDEV	Plasma Cannon Devastators
pewpewpew	shoot
pfist	power fist
plagues	plague marines
plasdevs	plasma devastators
Plasma	plasma gun
plat	platform
plz	please
PM	Private Message
PM	Plague Marine
PMC	Plague Marine Champion
ppl	people (p-pl)
pred	predator
Psi storm	Psionic Storm
pub	public
PukeLauncher	Bile Flamer
pumped	made bigger (got pumped from small to medium)
pwn	own, win against someone easily
QQ	cry (originally 'to quit' from Warcraft alt + q + q)
RA	Ravener Alpha
Ralph	Ravener Alpha
rape	to defeat badly
Ravs	Raveners
RB	Razorback
red	resource shown with red numbers

regen	regenerate
rep	repair
req	requisition
req bleed	loss of requisition over time
resetup	to set up again
RL	Rocket Launcher
RL	real life
RN	Relicnews
rolfstomping	beating easily
RR	Rocket Run
RTS	Real Time Strategy
rush	a strategy where you attack or get tech very quickly
Sab	Sabulum (a user name)
shuriplat	shuriken platform
sik	sick
SM	Space Marine
Sorc	Sorceror
SP	Single Player
spam (n)	the act of producing many units of the same type
spam (v)	to produce many units of the same type
spammable	easy to 'spam'
spike	To grenade spike, to throw a grenade at one's feet
splody	something that explodes (explodey -> splodey -> splody)
squishy	weak, easy to kill
sry	sorry
SS	Soulstorm
stealer	genestealer
Stealerspam	”spamming” genestealers
sticky	highlight a thread
stormies	storm boys
strat	strategy
Supp	suppression
supp team	suppression team, suppression + heavy weapon team
SvS	Sturm von Stahl
T1	Tier 1
T1.5	Tier 1.5
T2	Tier 2
T2.5	Tier 2.5
T3	Tier 3
tac	tactical marine
tactical	tactical marine
tact	tactical marine
Tank Hero	a commander unit that has a lot of health points
TB	Tankbusta
TBH	to be honest
TCSM	tactical chaosmarine
teamie	team mate
tech 2	tier 2
teching	advancing to the next tier, upgrading your units
tele	teleport
terma	termangaunt
termi	terminator
TG	Tyrant Guard
tic	heretic
TIOW	There Is Only War
TLS	The Last Stand

TM	Techmarine
tourny	tournament, tourney
TROLOLO	exclamated expression of successful trolling
trukk	war trukk
TS	True Skill
TT	Table Top
turrent	turret
turtling	strategy where a player builds mostly defensive units
TzDread	Tzeentch Dread
UP	Underpowered
upversioning	when new versions of the game are made
UR	your
UYC	Use Your Choppas
vanilla	without any upgrades
VC	Venom Cannon
vDoW2	Vanilla Dawn of War 2
vDread	Venerable Dreadnought
VenDread	Venerable Dreadnought
venombroods	Warrior brood upgraded with Venom Cannon
vent	Ventrilo
VP	Victory Point
w'	with
w/e	whatever
WB	Warboss
WG	Wraith Guards
WH40k	Warhammer 40,000
wit	with
WL	Warlock
WS	Warp Spider
WSE	Warp Spider Exarch
WTF	What the fuck
WTFPWN	What The Fuck own, win easily
XP	experience
xpac	expansion pack
zerkers	Chaos Space Marines with Mark of Khorne
Zero-upgrade	an upgrade which does not does not improve the unit
Zoan	Zoanthrope
zomg	ridiculed OMG
zone (v)	to keep enemy units out of a certain area
zoner	a unit that can zone (see above)