UNIVERSITY OF TARTU

DEPARTMENT OF ENGLISH STUDIES

CORPUS-BASED ANALYSIS OF IDIOMS DESCRIBING EMOTIONS IN BRITISH AND AMERICAN ENGLISH

BA thesis

SHARA HANNAMARY KULL

SUPERVISOR: JANE KLAVAN, PhD

TARTU

2020

ABSTRACT

This bachelor's thesis aims to analyse idioms describing three basic emotions – happiness, sadness, and anger – in the British and American varieties of English using Corpus of Contemporary American English and British National Corpus for frequency counts.

Idioms make language colourful and vivid. Idioms are frequently used to express emotions. However, learning and understanding idioms can be difficult. Furthermore, materials, such as textbooks and idiom dictionaries can often contain idioms that are not frequently used, making the studying of said idioms impractical. The findings of the thesis at hand show which emotion idioms are used frequently and which are not.

The first two chapters of the current thesis focus on describing and explaining what idioms, metaphors, emotions, and corpus linguistics are. Regular expressions and relative frequency are used in the current study – it is explained how and why they need to be used in the study along with examples. The third chapter contains the analysis of the idioms; each emotion has a separate section. The analysis of the spoken and written language is shown in a separate section as well. Following the analysis is the discussion, which emphasizes and examines the findings.

TABLE OF CONTENT

INTRODUCTION

Idioms are part of a language that makes it more vivid and colourful. When talking or writing about emotions and describing them, idioms can be used. However, it can be quite difficult to learn and understand idioms because the meaning of an idiom can often not be interpreted literally. Because there is a vast amount of idioms, it can be difficult to decide which idioms to learn and use. If one wants to learn the British variety of English, should they study the same idioms as someone wanting to learn the American variety? This thesis aims to find out which idioms of emotion occur more often in British and which in American English using corpus-based research on two corpora: Corpus of Contemporary American English (COCA) and British National Corpus (BNC). Furthermore, the thesis aims to find out whether some idioms occur more often in the spoken and some in the written language.

To begin with, the different definitions and descriptions of idioms need to be looked at. Some idiom dictionaries give quite a short definition for an idiom, pointing out that idioms have a different meaning to the meaning that is initially apparent – this is called non-compositionality. However, it is possible to focus on different aspects of an idiom with different dimensions and parameters. The first chapter focuses on the definitions, dimensions, and parameters of idioms to understand what an idiom is. Furthermore, it is shown why learning and understanding idioms can be troublesome and how idioms can often be described as fixed or semi-fixed.

The second section of Chapter 1 focuses on metaphors and language, showing why figurative language is important when it comes to expressing emotions and feelings. There are many emotion words and these words can be divided into expressive and descriptive ones, the difference between the two types of words is explained in the second section. Furthermore, the close link between metaphors, idioms, and the human thought process is demonstrated. Happiness, sadness, anger, fear, love, lust, pride, shame, and surprise are the

emotions in metaphors that have been most frequently analysed by different scholars. The second section explains which emotions are the basic ones. Furthermore, the section explains which emotions are used in the current thesis and why. Examples of the main metaphorical source domains for happiness, anger, and sadness are demonstrated, as well as the colours these emotions have often been associated with.

The third section of the first chapter focuses on the emotions and idioms that were chosen for the thesis at hand. The three chosen emotions – happiness, sadness, and anger – belong in the basic emotions group. The chosen emotions are often felt daily, which increases the possibility of idiom usage when describing them. The third section explains the criteria for finding the suitable lists of idioms for this thesis, acknowledging the importance of the length of the list of idioms as well as the relevance of the idioms. The lists of happiness, sadness, and anger idioms that are used in the research, are shown in three tables. In the fourth section the studies done on idioms by Liu (2003) and Václavíková (2010) are looked at.

There are many different corpora and the number, length, and content of the texts in a corpus can vary. The second chapter focuses on defining a corpus and shows how corpora can differ. In addition, the chapter discusses why it is important to know how to choose the right corpus for a research. Because the current thesis focuses on idioms in the British and American variety of English, the Corpus of Contemporary American English and British National Corpus were used. The description of the BNC and COCA are given, as well as a comparison between the two.

In the next two sections of Chapter 2, it is explained how and why searching an idiom from a corpus differs from searching a word. Regular expressions need to be used in corpora queries to get the results. BNC and COCA use different regular expressions. Examples of regular expressions used in the current thesis are given.

The last section in Chapter 2 focuses on relative frequencies. This thesis uses two different size corpora and that is why relative frequency needs to be used. What is relative frequency and why and how to use it is explained in the last section of Chapter 2. Furthermore, the formulas used in the thesis at hand are given as examples.

The analysis part of the thesis focuses initially on all the idioms together and then on each emotion separately. Happiness is the first category analysed. The most and least frequent idioms in BNC and COCA are presented along with graphs illustrating the findings. This process is repeated with sadness and anger idioms. Lastly, there is an analysis between spoken and written language. The analysis is done first between written COCA and spoken COCA and then between BNC and Spoken BNC:2014. Following the analysis is a discussion which focuses on the different findings of the analysis.

1. IDIOMS, METAPHORS AND EMOTIONS

1.1. What are idioms?

There are many different approaches to describing and defining an idiom. *Cambridge International Dictionary of Idioms* (CIDI, 1998: VI) describes idioms as "/…/ a colourful and fascinating aspect of English." Idioms are a way of saying something while making it more vivid. An example from CIDI (1998: VI): *look daggers at someone* is more emphasized than *look angrily at someone*. Similarly, *Longman American Idioms Dictionary* (LAID,1999: IX) focuses on the meaning of words and defines an idiom as a word sequence that has a different meaning when looking at the sequence together than when looking at the words separately. LAID (1999: IX) also points out that idioms make a language more colourful and help to emphasize meaning. Likewise, CIDI (1998: Introduction VI) mentions the aspect of making a language more colourful with idioms, also adding that idioms are used in both formal and informal, spoken, and written language. *The Oxford Dictionary of Idioms* (ODI, 1999: preface) gives a broader and more detailed definition:

> An idiom is a form of expression or a phrase peculiar to a language and approved by the usage of that language, and it often has a signification other than its grammatical or logical one. In practical terms this includes a wide range of expressions that have become in a sense fossilized within the language and are used in a fixed or semi-fixed way without reference to the literal meaning of their component words. (Oxford Dictionary of Idioms 1999)

Langlotz (2006: 2-3) agrees that defining an idiom is challenging and there are many aspects of which an idiom consists of. There are five dimensions to defining an idiom:

> a. semantic characteristics, b. structural peculiarities and irregularities and c. constraints or restrictions on their lexicogrammatical behaviour which cannot be explained by the general grammatical rules of the given language. Nevertheless, idioms are d. conventional expressions that belong to the grammar of a given language and e. fulfil specific discourse-communicative functions. (Langlotz 2006: 2-3)

Langlotz (2006: 3) shows the parameters for the definition of idioms as a table:

Semiotic dimension	Feature	Term
GRAMMATICAL STATUS	Degree of conventionalisation or familiarity	*institutionalisation*
FORM	Formal complexity of construction: multi-word unit	*compositeness*
	Lexicogrammatical behaviour: restricted syntactic, morphosyntactic and lexical variability	*frozenness*
MEANING	Meaning cannot be derived from constituent words but is extended/figurative.	*non-compositionality*

Figure 1. Parameters for the definition of idioms (Langlotz 2006: 3)

The term *institutionalisation* shows the extent of familiarity or conventionalisation of the idiom in a speech community (Fernando 1996: 3). *Compositeness* shows that idioms consist of two or more words and therefore consist of two or more lexical constituents. The term *frozeness* is used to capture lexicogrammatical restrictions. Langlotz (2006: 4) states that non-compositionality – words having a different meaning together than they do separately – is a primary feature of an idiom. This is supported by the aforementioned definitions and also by Katz and Postal (1963: 275), who emphasize the importance of non-compositionality in idioms.

Idioms are used in both written and spoken language and can be formal, informal, or very informal. Formal idioms are used in business documents, newspapers and books, lectures, and in polite conversations; informal idioms in more relaxed conversations, fictional books, and magazines; very informal idioms mostly in very relaxed situations, often involving a group of close members. (CIDI 1998: VI)

Non-compositionality makes the learning and understanding of idioms challenging for a language learner. Some idioms can have quite clear connections between the literal and the idiomatic meaning. For example, *in the blink of an eye* means *in an extremely short time* (McCarthy & O'Dell 2002: 108). The idiom describes an action that takes place in a very short amount of time. Therefore, the meaning of the idiom can be interpreted quite easily. However, other idioms may have no connection to the literal meaning of the words or the action. For example, *catch one's death* means *catch a severe cold or chill* (ODI 1999: 94).

However, out of context and in a literal way this idiom could be interpreted as someone physically catching their death.

Idioms are often described as fixed or semi-fixed. ODI (1999) describes idioms as expressions that are fossilized and are used in a fixed or semi-fixed way. McCarthy and O'Dell (2002: 6) claim that most idioms do not have variations and are fixed in their form (see Table 1).

Table 1. Fixed aspects of idioms (McCarthy & O'Dell 2002: 6)

Variation	*Example*
Occasionally an idiom in the active voice can be used in the passive.	Government Minister always **pass the buck** if they are challenged about poverty. [blame somebody else / refuse to accept responsibility] The **buck has been passed** from Minister to Minister. No one seems prepared to accept the responsibility.
Some verb-based idioms also have noun-compound forms.	There is too much **buck-passing** in government nowadays. No one accepts the blame for anything.
One or more words can be varied in the idiom.	Stop **acting the fool/goat**! [Stop acting stupidly]

However, Dictionary of Idioms: Helping Learners with Real English (1995: IV) argues that idioms that are described as 'fixed' generally are not fixed at all. Some examples of idioms having alternative forms that do not change the meaning: *burn your bridges* and *burn your boats* or *up the ante* and *raise your ante*.

1.2. Metaphors and emotions

Metaphors and figurative language in general give people the possibility of expressing their emotions and feelings in a colourful and expressive way. Emotional language consists of a vast amount of words. A few examples of emotion language words: *anger*, *fear*, *joy*, *love*. Wilche (2009: 8) claims that emotion language is necessary to form social comprehension and internal process. Kövecses (2000: 2) divides emotion words into

expressive and descriptive ones. Expressive words are those which express emotions, such as anger, enthusiasm, disgust, and impressiveness. For example *wow!*. Descriptive emotion words describe the emotion, for example, joy, love, anger, and sadness. Kövecses (2000: 2) adds that in some cases descriptive emotion terms can also express emotions. The vertical hierarchical levels of emotion terms are: superordinate level – emotion, middle (basic) level – anger, subordinate level – annoyance. In emotion language terms the word "basic" can mean either that these words, in a vertical hierarchy of concepts, are in the middle level (above subordinate and beneath superordinate level) or that the emotion category at hand is more common than some other emotion category (Kövecses 2000: 3).

The third and the biggest subgroup of descriptive terms is metaphor and metonymy, which is the group of figurative terms and expressions (Kövecses 2000: 4). Because happiness, anger, and sadness idioms are analysed in the current thesis, the metaphor and metonymy subgroup was used. Fellbaum (2007: 13) emphasizes that metaphors are often closely linked to idioms and idioms often contain metaphors. Lakoff and Johnson (2003: 7) point out that the human conceptual system contains metaphors and the thought process itself is essentially metaphorical.

The most common emotions in metaphors that have been looked at and analysed by different scholars are the basic emotions: happiness, sadness, anger, fear, and love. In addition, lust, pride, shame, and surprise are also commonly analysed (Kövecses 2000: 20). Because these emotions – happiness, sadness, and anger – are felt almost daily, they were chosen for the analysis of the current paper. Happiness, sadness, and anger are considered to be basic-level categories (Kövecses 2000: 20). Both Lakoff and Johnson (2003) and Kövecses (2000: 20) agree that emotion metaphors are important. Kövecses (2000: 20) states that metaphors help to conceptualize both emotions and emotional experience. Lakoff and Johnson (2003: 246) agree by writing:

> The idea that metaphors are nothing but linguistic expressions, a mere matter of words, is such a common fallacy that it has kept many readers from even entertaining the idea that we think metaphorically. The fallacy is that metaphor is only about the ways we talk and not about conceptualization and reasoning. Countering this view is a huge body of empirical evidence gained from many different methods of inquiry that reveals the central role of metaphor in abstract thought. (Lakoff & Johnson 2003: 246)

Kövecses (2000: 21) suggests that cognitive semantics has studied anger more than other emotions. Kövecses (2013: 80-82) and Lakoff and Kövecses (1987: 201-204) give a list of the main metaphorical source domains for anger in English. In the list there are 12 source domains. Some examples from the list are: anger can be a hot fluid in a container (*She is boiling with anger*) and anger is fire (*He's doing a slow burn)*. Václavíková shows in her thesis *Idioms of Colour – A Corpus-based Study* (2010) that anger is strongly connected to the colour white (*white-knucle(d)*) and red (*See red*). The colour green (*Green with envy*) is also associated with anger but more precisely with envy and jealousy. Different metaphorical source domains can address different aspects of the concept of anger, the focus can be on the person that is angry (*That really got him going*), on the cause of anger (*He's a pain in the neck*), on the angry behaviour (*Don't snarl at me*), or on the aspect of control (*I was struggling with my anger*). (Kövecses 2000: 21-22)

The list of metaphorical source domains about happiness metaphors contains 15 domains taken from Lakoff and Johnson (2003). Some examples are: happy is up (*We had to cheer him up)* and happiness is being off the ground (*I was so happy my feet barely touched the ground)* (Kövecses 2000: 24-25).

In the sadness metaphor list there are 14 source domains (Kövecses 2000: 25-26). Some examples are: sad is down (*He brought me down with his remarks*) and sadness is a lack of vitality (*This was disheartening news)*. In her thesis, Václavíková (2010) associates sadness with dark colours: black (*Black mood*), blue (*Have the blues*), grey (*Grey world*). Steinvall (2007: 357-358) explains that in European and American cultures the colour black often symbolises death and mourning which is why dark colours in language are commonly

associated with sadness. Kövecses (2000: 26) points out that many sadness source domains are the opposites of the happiness source domains.

1.3. Idioms of happiness, sadness, and anger

There is a vast amount of idioms describing emotions, therefore a selection of emotions had to be made. The emotions that were chosen were happiness, sadness, and anger. These emotions were chosen because they are often a part of everyday life and because of that people are perhaps more likely to use idioms to describe them. These emotions are also categorized as part of the basic emotion category (the other two emotion categories remaining in the list are fear and love) in 11 languages based on cross-cultural research in *Metaphor and Emotion* by Zoltan Kövecses (2000), who conducted his research using the methodology by Fehr and Russell (1984). The criteria for finding the suitable lists of idioms for the current thesis were: 1) the lists had to be long enough to have variety but not too long in order to avoid repetition 2) the idioms had to be relevant and important to an English speaker. The vocabulary book *English Idioms in Use* by McCarthy and O'Dell (2002) was chosen. The aforementioned book contains 60 units and each unit is focused on a different category of idioms. In addition to a list of idioms, each unit also provides the definitions for idioms, examples of idioms in sentences, and exercises containing idioms. This vocabulary book was chosen for the current thesis because the books' lists of happiness, sadness, and anger idioms consist of 8-10 idioms. Furthermore, McCarthy and O'Dell (2002) chose the idioms for the vocabulary book using the CANCODE corpus of spoken English, ensuring that the idioms chosen are used by native speakers of English in conversations, newspapers, and novels. The idioms chosen for the study are given in Tables 2-4.

Table 2. List of idioms describing happiness (McCarthy & O'Dell 2002: 12-13)

Idiom	Meaning
Get a (real) kick out of something	Very much enjoy doing something
Do something for kicks	Do something because it is exciting, usually something dangerous
Jump for joy	Be very happy and excited about something that has happened
Be floating/walking on air	Be very happy about something good that has happened
Something makes you day	Something makes you feel very happy
Thrilled to bits	Extremely happy
Be/feel on top of the world	Extremely happy
Be on cloud nine	Extremely happy
Be over the moon	Extremely happy
Be in seventh heaven	Extremely happy

Table 3. List of idioms describing sadness (McCarthy & O'Dell 2002: 12-13)

Idiom	Definition
Out of sorts	Slightly unhappy or slightly ill
Down in the dumps	Unhappy
It is not the end of the world	What has happened won't cause any serious problems
Just grin and bear it	Accept the situation you don't like because you can't change it
A misery guts	Someone who complains all the time and is never happy (very informal)
Sour grapes	Being jealous about something you can't have
Puts a damper on	Stop an occasion from being enjoyable

Table 4. List of idioms describing anger (McCarthy & O'Dell 2002: 14-15)

Idiom	Definition
Drive someone up the wall	Make someone very angry (or sometimes very bored)
Drive/send someone round the bend/twist	Make someone very angry (or sometimes very bored)
Rub someone up the wrong way	Make someone annoyed
Get/put someone's back up	Make someone annoyed
Ruffle someone's feathers	Make someone annoyed
Put/send the cat among the pigeons	Do or say something that makes a lot of people angry or worried
Not be on speaking terms	Be so angry with each other that they refuse to speak to one another
Give someone an earful	Tell someone how angry you are with them
Give someone a piece of your mind	Tell someone how angry you are with them

1.4. Corpus linguistics and the study of idioms

Brezina (2018: 2) claims that corpus linguistics is a method that is used to confirm or contradict a statement that has been made about the language, using data drawn from a corpus as evidence. The results need to be replicable in follow-up studies. Statistics is important to corpus linguistics because it is used to work with quantitative information and corpus linguistics is a quantitative methodology where work is done with numbers that reflect word and phrase frequencies in corpora (Brezina 2018: 3).

Liu (2003) conducted a study *The Most Frequently Used Spoken American English Idioms: A Corpus Analysis and Its Implications*. Liu (2003) found that many idiom vocabulary books consist of idioms that have been chosen primarily on the intuition of the

author. Many of these materials can be unhelpful because they do not rely on empirical data. For example, a textbook may claim to cover all the essential idioms but in reality contain idioms that are rarely used. This makes the learning of idioms difficult. “Because of their rather rigid structure, quite unpredictable meaning, and fairly extensive use, idioms are “a notoriously difficult” but simultaneously very useful aspect of English,” states Liu (2003). Because of the importance of idioms and the lack of materials relying on empirical data, Liu (2003) conducted a study to identify the most frequently used spoken American English idioms. More specifically, Liu (2003) focused on idioms useful for students.

Liu (2003) identified the most frequently occurring idioms across three large corpora samplings from spoken American English. These corpora were 1) Barlow’s (2000) Corpus of Spoken, Professional American English (CSPAE) 2) a corpus of Spoken American Media English (Liu, 2003) and 3) Simpson et al.’s (2002) Michigan Corpus of Academic Spoken English. To find the relative frequency, Liu (2003) set the frequency level of two tokens (words) per million and the results were separated into three bands – the first band consisted of the most frequently occurring idioms with 50 or more tokens per million words, the second band consisted of idioms that had 11-49 tokens per million words, and the third band consisted of idioms with 2-19 tokens per million words. The three most frequently used idioms in Liu’s (2003) study were: *kind of, sort of,* and *of course,* which were in the first band. The three least frequent idioms in the first band were: *have/keep in mind*, *call for*, and *in general*. In the second band the three most frequent idioms were *take care of*, *go over*, and *on the other hand* and the three least frequent idioms were *break down, put up*, and *take over*. The three most frequent idioms in the third band were *be open to ideas, rule out,* and *as for* and the three least frequent idioms were *once and for all, push the envelope*, and *with (keep) one’s eye on something*. (Liu 2003)

In conclusion, Liu (2003) stated that idioms in teaching materials need to be chosen with more care and empirical data should be looked at to ensure that the idioms included in a material are frequently used and authentic to the language. Furthermore, Liu (2003) added that including additional information about idioms, such as their usage frequency, could encourage students to learn those idioms.

Václavíková (2010) used corpus linguistics in her thesis *Idioms of Colour – A Corpus-based Study.* Different to Liu (2003), Václavíková did not analyse the idioms with language learners in mind but rather to see whether (colour) idioms appear more frequently in newspapers and magazines or in academic texts. Václavíková (2010) explained that she chose colour idioms for her study because colours are everywhere and therefore, they are bound to be used in language. Differently from Liu (2003), Václavíková (2010) used both British and American corpora. The three corpora used in the study were: British National Corpus (100 million words, UK 1980s-1993), Corpus of Contemporary American English (400+ million words, US 1990-2010) and the TIME Magazine Corpus of American English (100 million words, US 1923-2006).

The colours analysed by Václavíková (2010) were: red, orange, yellow, green, blue, violet, white, grey, and black. Václavíková (2010) searched for relevant expressions in general English dictionaries and from specialized dictionaries of idioms with the purpose of checking the expressions. Roughly 500 expressions were collected in the preliminary research.

Václavíková (2010) analysed each colour idiom in all three corpora to collect the data. One reason for that was to additionally see whether there is a difference in idiom usage between the British and American variety of English. To get the required data, regular expressions had to be used, and a context had to be chosen. Václavíková (2010) stated: “It is possible that my results will differ from those others may find, because there were usually

more ways of entering the expression into the interface." This shows the problem of analysing idioms: since there are many different ways of searching an idiom, there is always the possibility that the search does not give all the relevant answers and using a different regular expression might give different results.

Václavíková (2010) described each colour idiom and its frequency separately, showing the percentages in all subgroups (spoken, fiction, magazines). Normalized frequency was not used in Václavíkovás' (2010) study. Therefore, it is difficult to thoroughly compare the results of COCA and BNC because of their size difference. Examples of idioms that had a high usage frequency in Václavíkovás' (2010) study include: *black sheep of the family*, *out of the blue*, *green with envy*, *grey matter*, *see red*, *white-livered*, and *yellow journalism*. Overall, Václavíková (2010) found that colour idioms are used more frequently in American English than in British English and that idioms are used more in newspapers and magazines than in academic texts.

2.CORPUS LINGUISTICS AND IDIOMS

2.1.What is a corpus?

A corpus is essentially a collection of texts; the number, length, and content of the texts can vary greatly. Kennedy (1998: 3) describes a corpus as a collection of texts in an electronic database and adds that a corpus can be used both for linguistic analysis and to represent a language as a whole. Anderson & Corbett (2017: 4) define a corpus as a sizable body of texts that has been created for a specific purpose or purposes. Brezina (2018: 6) specifies that the texts can be either written texts or transcripts of spoken language. The texts in a corpus can be long or short, from many different authors or just by one, or from different categories – all these restrictions depend on the representativeness of the corpus. The current chapter focuses on identifying and defining a corpus in greater detail and mainly focuses on the two corpora used in the current study – COCA and BNC, describing and comparing the two. Additionally, explanations of corpus queries and relative frequency are given, as well as examples on how they are used in the study at hand.

The size of corpora can vary, ranging from having just a few texts (or even one text) to having millions of them. A corpus can be representative of a language as a whole or just represent the works of one author, works from a specific year, or works about a certain topic (Kennedy 1998: 4). Corpora can be designed for many different purposes. According to *An Introduction to Corpus Linguistics* (Kennedy 1998: 3-4) a corpus is designed to give answers to linguistic questions concerning, for example, lexis, grammar, and prosody of the language.

Kennedy (1998: 4) explains that in order to see how often and where exactly some phonological, grammatical, or lexical features occur, a corpus can be analysed distributionally. Furthermore, according to *Exploring English with Online Corpora* (Anderson & Gorbett 2017: 3), a careful corpus analysis can show how a language is commonly used opposed to how people might think it is.

When choosing a corpus, it is essential to take into consideration that there are many different corpora which represent different language populations. As mentioned before, the data in the corpus can vary, for example, the texts can be from one or various authors, from one specific year, or a specific time period. Therefore, it is important to know the original purpose for which the corpus was created when choosing it for analysis.

Because the purpose of the current thesis is to compare the usage of emotion idioms in the British and American varieties of English, it was crucial to choose corpora which represent both written and spoken language, and is restricted only by either British or American English and not by a year or an author. The size of the corpus was also a deciding factor because the corpora had to be large enough to provide valid data for comparisons. For example, a written corpus is considered to be small when it consists of less than five million words and a corpus of spoken language large when it consists of over a million words (Anderson & Corbett 2017: 7). The Corpus of Contemporary American English (COCA) and British National Corpus (XML edition) (BNC) were used in this thesis.

In COCA, phrases, synonyms, words, and substrings can be searched. There are also many different genres that can all be looked at either together or separately. When searching for a word or a collocation, one can choose a subsection, which can be a genre, subgenre, or a time period. COCA consists of genres, such as spoken, fiction, magazines, newspapers, academic texts, web (general), web (blogs), and TV/movies. The genres also have subgenres; for example, when choosing the *magazines* genre it is also possible to choose which category of magazines (choices include *children's*, *home & gardening*, *financial*). Due to this, COCA gives an extensive overview of different grammatical constructions, word frequencies, phrases, and usages across many genres. It is also possible to compare different words, collocates, genres, and time periods with one another. COCA is updated once or twice a year, and each year contains about 20 million words. (Davies 2008)

When creating the BNC, there were no field, genre, or register restrictions used. The BNC represents British variety of English and has both spoken and written language examples. To assure that the data in the written part of the corpus is not used only in books but is also authentic to the day-to-day life, there are also miscellaneous published (leaflets, manuals, advertisements), unpublished (letters, essays, memos), and to-be-spoken (play scripts, scripted television material) texts included in the corpus. When making a standard query in BNC, it is possible to choose the length of the match – for example, when limiting the search to the "longest-possible match," the longest text in the corpus with the matching query result will be presented. It is also possible to apply a restriction for the search when choosing written or spoken texts. However, there are no further genre-based restrictions available as are in COCA. The Spoken BNC:2014 was also used in the current study. There are 11,422,617 words in the aforementioned corpus. The texts in the corpus are mostly informal because the texts are the transcriptions of recordings done at home among family and friends. (Hardie 2012: 380-409, Love et al 2017)

2.2. Corpus queries

The thesis at hand used corpus queries to find out how frequently idioms occur in BNC and COCA. Each idiom had to be searched individually in both corpora for the purpose of finding the idiom's absolute frequency. Because looking up an idiom in a corpus is not as straightforward as looking up a word, regular expressions had to be used. See Appendix 1 for the full list of idioms and the corresponding regular expressions used in the current study. To find all the results for the query, it was also important to remember that each verb in an idiom can have various tense differences. The regular expressions vary for each corpus, therefore a formula for each idiom had to be found separately in both BNC and COCA. For example, in COCA, the necessary nouns, pronouns, verbs, adjectives, adverbs, and

prepositions could be chosen from a list of features located on the right of the search bar. Therefore, if an idiom consisted of a pronoun that can vary (*she, he*), the appropriate feature had to be inserted – in this case, for a pronoun: "_p*". In BNC, a different label for the feature had to be used – for example, "_{PRON}" had to be typed in the search to find a pronoun. When typing *Give someone an earful* in the COCA search only two results come up. However, when using regular expression "*_v* * an earful*" 144 results came up. From those 144 results the matching idioms had to be manually extracted. BNC is similar because there are no results when searching *Give someone an earful*. However, when using regular expressions "*{give/V} * an earful*" three results come up.

After the idiom had been searched and the number of occurrences of the idiom had been found, the results were recorded in Excel. The Excel file was created with various sheets, a sheet for each corpus with the collected data. For example, the query for *get a kick out of something* had 945 results in COCA – the number of results for the aforementioned idiom was then written into the COCA sheet. After all the results were recorded in Excel, an Excel formula was used to find the relative usage frequency of the idiom in the corpus (see the next section).

2.3. Relative frequency

According to *Statistics in Corpus Linguistics: A Practical Guide* (Brezina 2018: 42-43), there are two types of frequencies: absolute (raw) frequency and relative (normalized) frequency. Brezina (2018: 43) defines absolute frequency as the actual number of all the occurrences of a particular word in a corpus. Absolute frequency can be effectively used when analysing a single corpus or two corpora that are the same size. However, when comparing corpora (two or more) that have different sizes, relative frequency needs be used (Brezina 2018: 43, Anderson & Corbett 2017: 32). Because the current thesis focuses on

comparing two corpora (COCA and BNC) that have unequal size (1,001,610,938 and 112,102,325 words respectively), relative frequency is used to get results that can be compared to one another.

The absolute frequency of the word (or phrase) of interest and the total number of words in the corpus need to be known for calculating relative frequency (Brezina 2018: 43). The calculation for relative frequency was based on the following formula:

$$\text{relative frequency} = \frac{\text{absolute frequency}}{\text{number of tokens in corpus}} \text{ x basis for normalization.}$$

For example, for the idiom *Be over the moon* in the BNC, with an absolute frequency of 59 results, the relative frequency calculation is the following:

$$\frac{59}{112102325}\ x\ 1000000 = 0.53.$$

When comparing two or more corpora of different sizes, normalization is done. Anderson and Corbett (2017) state that "Normalized frequencies usually tell us the number of occurrences that there are, or that we can expect per thousand, or sometimes per million words." Furthermore, Brezina (2018: 43) states that normalization is useful for presenting findings or evidence in a way that is easier for a reader to grasp.

Normalization can vary depending on the size of the corpus. For smaller corpora, a normalization for 10,000 or even 1,000 words can be done. For the thesis at hand, the basis for normalization was chosen to be 1 million words in order to obtain comparable results from COCA and BNC.

3. ANALYSIS

3.1. General comments

There were 26 idioms taken from English Idioms in Use (2002: 12-15) for the analysis. Some idioms had different variations, for example *Be floating/Walking on air*, *Drive/Send someone round the bend/twist*. In these cases, each variation was first searched individually and later the variations were combined to get the final result. For example, for the idiom *... on top of the world – Be on top of the world* (280 results in COCA) + *Feel on top of the world* (47 results in COCA) the combined frequency is 327 results and the relative frequency of 0.33 per 1,000,000 words.

In order to get all the results from the corpora, the possible tense differences that an idiom can have needed to be taken into account. For this purpose, regular expressions were used. For example, in the idiom *Give someone a piece of your mind*, the word *give* can change. Therefore, the following regular expressions were used to allow tense differences: "{give/V} *" (in BNC) "_v*" (in COCA). For some idioms, the possessive forms had to also be found, for example for the idiom *Ruffle someone's feathers*. For some idioms that meant that the results had to be looked at and manually cleaned to make sure that the results included the idiom and not an expression with a different meaning. The full list of idioms together with the regular expressions used for both corpora is provided in Appendix 1.

3.2. Happiness idioms

As can be seen on Figure 2, the 10 happiness idioms included in the analysis are more frequent in COCA than in BNC. Idioms with highest frequency in COCA were: *Make your day* (1052 results, relative frequency 1.05), *Get a kick out of something* (945 results, relative frequency 0.94), and *Be/Feel on top of the world* (327 results, relative frequency 0.33). The least common happiness idiom in the COCA was *Thrilled to bits* (5 results).

In the BNC, the idioms with most results were: *Be over the moon* (59 results, relative frequency 0.53), *Make your day* (45 results, relative frequency 0.40), and *Be/Feel on top of the world* (36 results, relative frequency 0.32). The least common happiness idiom in the BNC was *Do something for kicks* (2 results).

In Figure 2 it can be seen that the idiom that had the biggest difference between COCA and BNC was *Make your day*, which had 1052 results in COCA with a relative frequency of 1.05 and 45 results in BNC with the relative frequency of 0.40. Idiom *Be/Feel on top of the world* had the most similar relative frequency in COCA and BNC with 327 results and the relative frequency of 0.33 in COCA and 38 results and the relative frequency of 0.32 in BNC. The two idioms with a higher relative frequency in BNC than in COCA were: *Be over the moon* with the relative frequency of 0.53 in BNC and 0.16 in COCA and *Be on cloud nine* with the relative frequency of 0.16 in BNC and 0.07 in COCA.

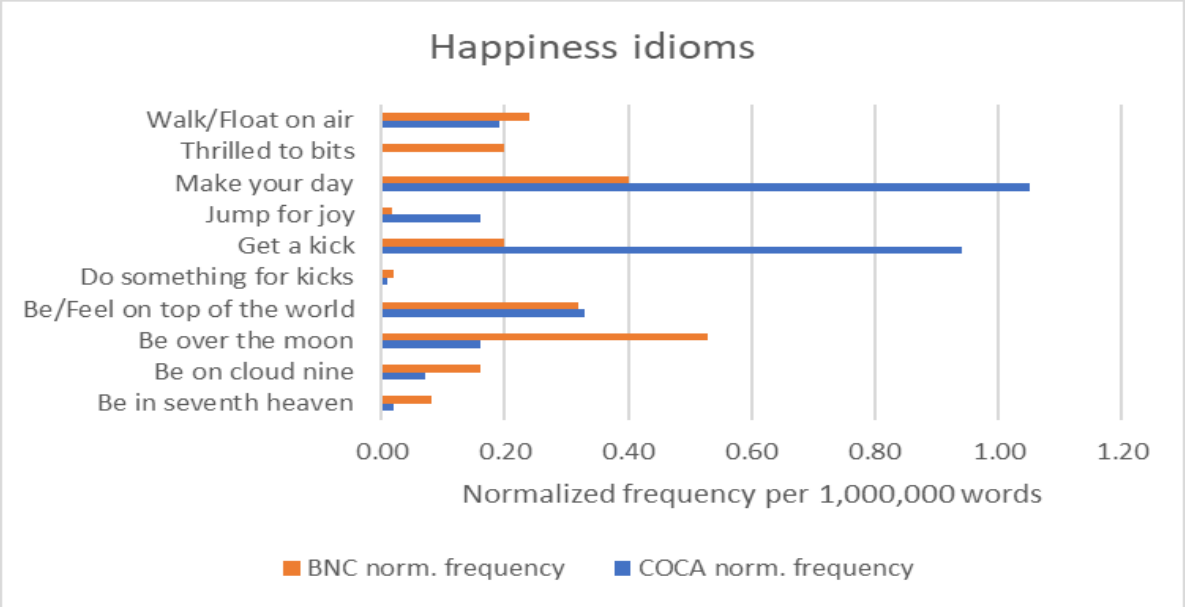

Figure 2. Normalized frequency counts of happiness idioms in COCA and BNC.

3.3. Sadness idioms

The relative frequency differences of the seven analysed sadness idioms were not as big as with happiness idioms (see Figure 3). The most common sadness idioms in COCA were: *Sour grapes* (489 results, relative frequency 0.49), *It's not the end of the world* (408 results, relative frequency 0.41), and *Out of sorts* (338 results, relative frequency 0.34). The idiom with the lowest frequency was *A/Misery guts* (5 results).

The most common sadness idioms in the BNC were: *Sour grapes* (48 results, relative frequency 0.43), *Out of sorts* (34 results, relative frequency 0.30), and *Just grin and bear it* (25 results, relative frequency 0.22). It can be seen in Figure 3 that the least common sadness idiom in the BNC was *A/Misery guts* (7 results).

The idiom *Put a damper on* had the biggest difference in COCA and BNC, in COCA it had 329 results and the relative frequency of 0.33 and in BNC 8 results with a relative frequency of 0.07. The idiom *Out of sorts* had the most similar frequency in COCA and BNC, 338 results and a relative frequency of 0.34 in COCA and 34 results and a relative frequency of 0.30 in BNC.

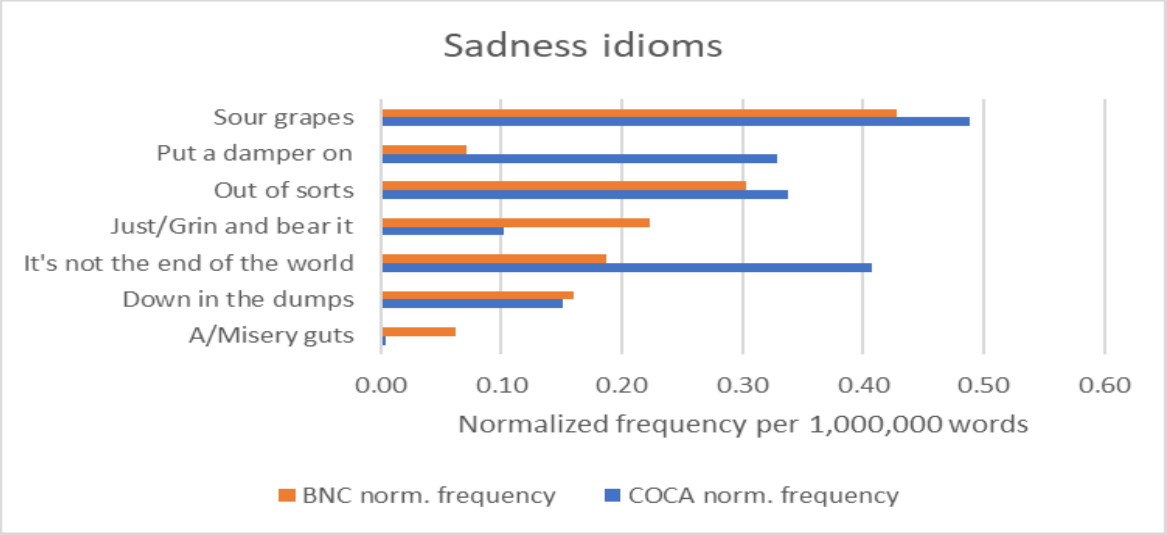

Figure 3. Normalized frequency counts of sadness idioms in COCA and BNC.

3.4. Anger idioms

It can be seen in Figure 4 that the nine anger idioms that were analysed were more frequent in BNC than in COCA. The most common anger idioms in COCA were: *Give someone an earful* (92 results, relative frequency 0.09), *Ruffle someone's feathers* (91 results, relative frequency 0.09), and *Drive someone up the wall* (89 results, relative frequency 0.09). The least common anger idiom in the COCA was *Put/Send the cat among the pigeons* (1 result).

The most common anger idioms in the BNC were: *Drive someone up the wall* (16 results, relative frequency 0.14), *Rub someone up the wrong way* (14 results, relative frequency 0.12), and *Give someone a piece of your mind* (13 results, relative frequency 0.12). The least common anger idioms in BNC were *Ruffle someone's feathers* (3 results) and *Give someone an earful* (3 results).

The idiom *Rub someone up the wrong way* had the biggest relative frequency difference, in COCA it only had 4 results and in BNC it had 14 results and the relative frequency of 0.12. *Give someone a piece of your mind* had the most similar relative frequency in COCA and BNC, with 58 result and the frequency of 0.06 in COCA and 13 results in BNC with relative frequency of 0.12.

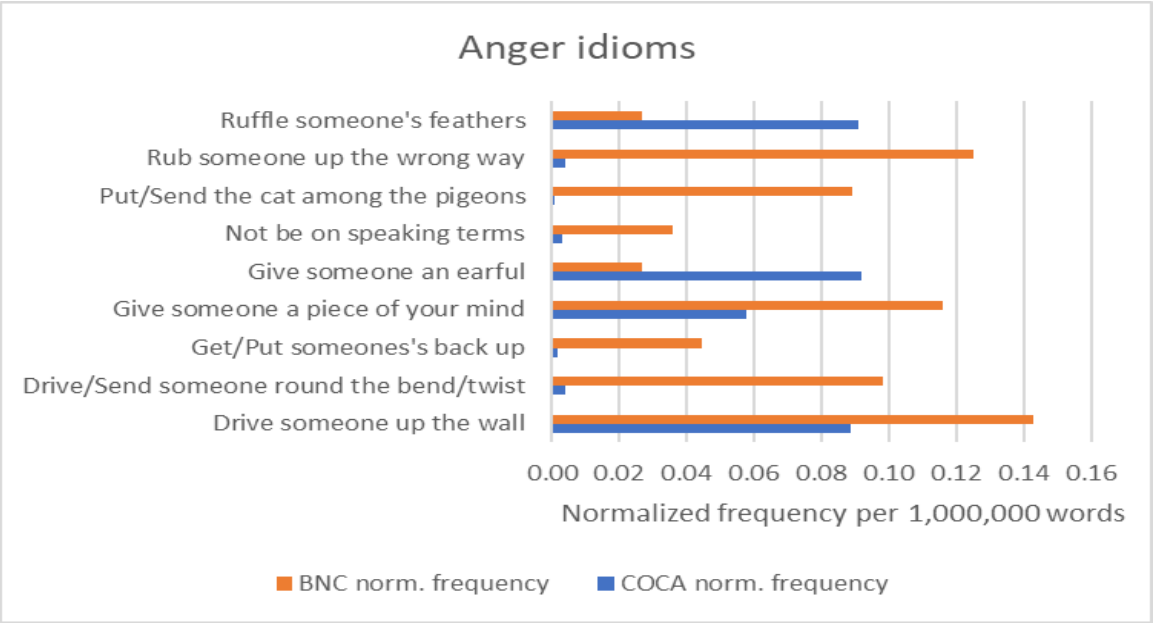

Figure 4. Normalized frequency counts of anger idioms in COCA and BNC.

3.5. Analysis between the spoken and written language

In order to see, whether idioms are used in everyday speech and if they are, then how often, the spoken and written part of a corpus had to be looked at and analysed separately. In COCA it was possible to choose between spoken and written language. Since there are 127,396,916 words in the spoken and 874,214,022 words in the written part of COCA, this number was used as the number of tokens when calculating relative frequencies. Example with *Be/Feel over the moon* idiom, which had 24 results in the spoken part of COCA is: 24*(1000000)/127396916=0.19. The sum of the usage frequency of each emotion category in both the written and spoken part of COCA was calculated and the frequency results can be seen in Figure 5. As can be seen from Figure 5 idiom usage was bigger in the written language; the difference was small, however (see Figure 5).

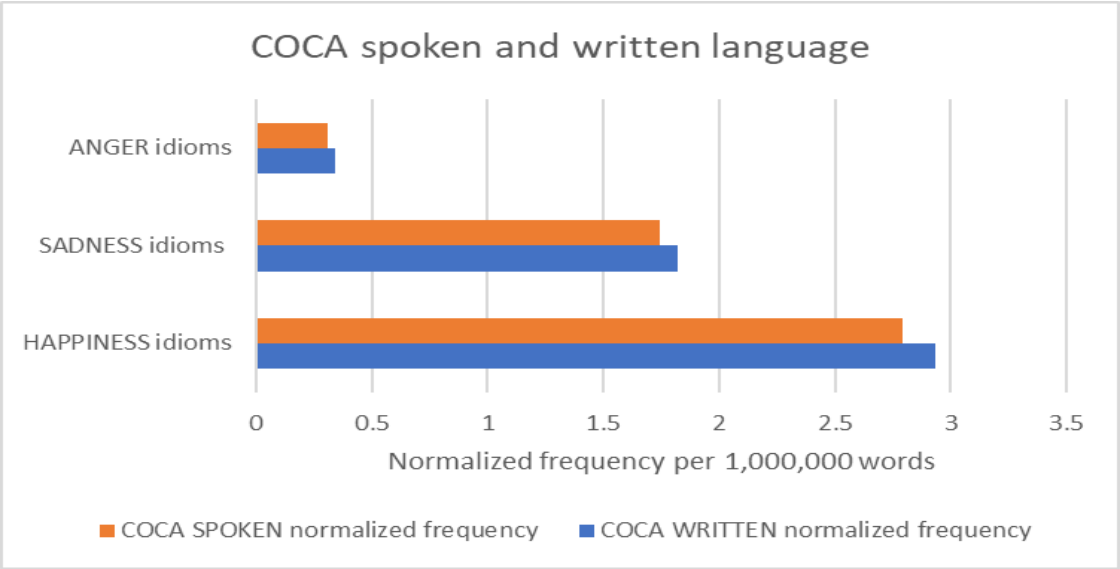

Figure 5. Normalized frequency counts of idioms in spoken and written part of COCA.

The Spoken BNC:2014 was analysed as well. However, because of the size of the corpus, there were very few results. In Spoken BNC:2014 there are 11,422,617 words in total. This number was used as the total number of tokens when calculating the relative frequency. Although the absolute frequencies were bigger in BNC, the relative frequencies were bigger in Spoken BNC:2014. (see Figure 6)

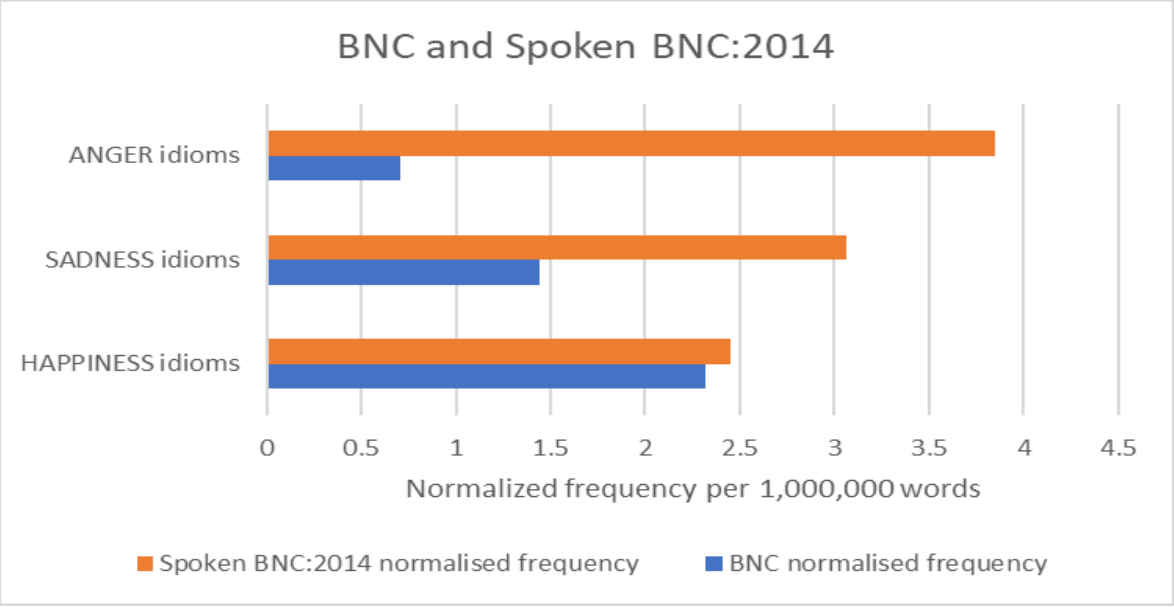

Figure 6. Normalized frequency counts of idioms in BNC and Spoken BNC:2014.

The idiom *Drive someone up the* wall had the biggest relative frequency in Spoken BNC:2014 – 1.93 with the absolute frequency of 22. Other idioms with high relative frequency in Spoken BNC:2014 were: *It's not the end of the world* (absolute frequency 20, relative frequency 1.75) and *Drive/Send someone round the twist/bend* (absolute frequency 11, relative frequency 0.96). Idioms *Do something for kicks*, *Down in the dumps*, *Give someone an earful*, and *Not be on speaking terms* had the smallest relative frequencies (0.00) in Spoken BNC:2014..

There were more results for idioms in the written part of COCA, but the relative frequency was higher for some idioms in the spoken part of COCA. The happiness idioms had quite similar relative frequencies in both the spoken and written part. As can be seen in Appendix 3, the idiom *Make your day* had the highest results with relative frequency of 0.79 (with 101 results) in the spoken part and in the written part relative frequency of 1.05 (with 1052 results). Idiom *Get a kick* had the relative frequency of 0.94 (with 945 results) in the written part and the relative frequency of 0.91 (with 116 results) in the spoken part of COCA. The idiom *Be/Feel on top of the world* had the relative frequency of 0.33 (with 327 results) in the written part and the relative frequency of 0.39 (with 50 results) in the spoken part. Idioms *Do something for kicks*, *Thrilled to bits* and *Be in seventh heaven* had the lowest results and relative frequencies, all three having only one result in the spoken part of COCA (see Appendix 3).

The sadness idioms also had quite similar relative frequencies in both the written and spoken part of COCA (see Appendix 3). The idiom with the highest relative frequency was *Sour grapes*, which had the relative frequency of 0.55 (with 70 results) in the spoken part and relative frequency of 0.49(with 489 results) in the written part of COCA. The idiom *It's not the end of the world* had the relative frequency of 0.41 (with 408 results) in written part and the relative frequency of 0.34 (with 43 results) in the spoken part. The idiom *Out of sorts*

had the relative frequency of 0.34 (with 338 results) in the written part and the relative frequency of 0.25 (with 32 results) in the spoken part of COCA. Idioms *Down in the dumps*, *Just/Grin and bear it* and *A/Misery guts* had the lowest results and relative frequencies with *A/Misery guts* having only five results in the written part and no results in the spoken part of COCA.

There were less anger idioms in general in the spoken part of COCA than in the written part. However, two idioms had a higher relative frequency in the spoken part (see Appendix 3). The idiom *Give someone an earful* had the relative frequency of 0.15 (with 19 results) in the spoken part and the relative frequency of 0.09 (with 92 results) in the written part of COCA. The idiom *Give someone a piece of your mind* had the relative frequency of 0.09 (with 12 results) in the spoken part and the relative frequency of 0.06 (with 58 results) in the written part. The idiom *Ruffle someone's feathers* had a higher relative frequency in the written part, which was 0.09 (with 91 results) and the relative frequency of 0.06(with 8 results) in the spoken part. The idioms *Drive/Send someone round the bend/twist*, *Rub someone up the wrong way*, *Get/Put someone's back up*, *Put/Send the cat among the pigeons* and *Not be on speaking terms* had no results in the spoken part of COCA.

4. Discussion

The aim of this thesis was to analyse idioms describing three basic emotions – happiness, sadness, and anger – in COCA and BNC and to see how the idiom usage differs between the British and American varieties of English. In addition, how it differs between the spoken and written language. The idioms for the analysis were gathered from the book *English Idioms in Use* (McCarthy & O'Dell 2002). The aforementioned book had lists of idioms that each consisted of seven to ten idioms. Idioms that were significant to the native speaker were chosen by McCarthy and O'Dell's (2002) for their vocabulary book. These idioms from *English Idioms in Use* were chosen for the current study. According to the results of the analysis, happiness idioms are more frequent in COCA than in BNC. The idiom with the highest relative frequency difference is *Get a kick*, which has considerably more results in COCA than in BNC, showing that this idiom is used more in American English than in British English. *Thrilled to bits* and *Be over the moon* are used more in British English. Out of the three emotion categories, happiness idioms have the highest absolute frequencies both in COCA and BNC. This shows that people tend to describe happiness with idioms more than feelings of anger or sadness.

The difference in usage frequencies between sadness idioms is not very significant. The idiom *It's not the end of the world* is used significantly more often in American English and the idioms *Just grin and bear it* and *A/Misery guts* more often in British English. Idioms *Down in the dumps* and *Out of sorts* have a similar usage frequency in COCA and BNC.

It is surprising to see that the anger idioms have higher relative frequencies in BNC. The biggest difference in idiom usage is with the idiom *Rub someone up the wrong way*, which is used in British English more than in American English. Idioms *Ruffle someone's feathers* and *Give someone an earful* are used more in American English. The idioms *Drive someone up the wall* has the smallest difference in usage frequency in COCA and BNC.

The difference between the relative frequency of spoken and written language in COCA is not very substantial, showing that idioms are used frequently in spoken as well as written language. The fact that McCarthy and O'Dell (2002) chose the idioms for their vocabulary book from the CANCODE may explain the small difference. The written part of COCA consists of newspaper and academic texts but also of blog, web, fiction, and magazine texts. Therefore, the majority of texts in the written part of COCA are informal. The spoken part of COCA consists of different TV and radio programs representing authentic language use. The BNC consists of texts from newspapers, journals, and fiction books. The Spoken BNC:2014 consists of transcribed texts that have been gathered from real-life contexts. Because Spoken BNC:2014 is considerably smaller than BNC and COCA, the results were also limited.

Idioms can be formal, informal, and very informal. The idioms used in the current analysis were largely informal, meaning that they are mostly used among family and friends in a non-formal setting. Some idioms were very informal, such as *A/Misery guts*. When searching an idiom in a corpus, each result can be looked at separately to further identify the setting in which it occurs (for example in a blog or a magazine, in a show or recorded conversation).

Many idioms analysed in the current study varied, some more than others. Idioms that had less variation in grammar or vocabulary and were more fixed, were shorter idioms, such as S*our grapes*. The longer the idiom, the more possibilities for variations and the less frozen they were. For example, the idioms *Drive/Send someone round the bend*/*twist*, *Put/Send the cat among the pigeons* and *Be floating/walking on air* varied to considerable extent. There were also vocabulary variations that were not included in the analysis because it would have been too difficult and time-consuming to include all the possibilities and

variations for each idiom. For example, *Put/Send the cat among the pigeons* can also start with the word *Set* or *Throw*.

When doing a corpus-based study, as does the thesis at hand, it needs to be kept in mind that corpora sizes often differ and the regular expression used might not be inclusive enough to get all the results. The idioms for the study at hand were carefully selected. However, it is possible that there are happiness, sadness, and anger idioms that are frequently used in either the British or American variety of English and which would have been frequent in a corpus but did not get selected for the analysis of the current study. Using different idioms could give different results. Replicating the current study with different idioms could be an interesting topic for future research. The results of the current study can also be affected by the fact that BNC (XML Edition) is a closed corpus released in 2007 and COCA an open corpus that is updated once or twice a year. Therefore, the data collected from BNC may not give an accurate representation of the current language use.

Similarly to Václavíková's (2010) study, the results of the current thesis show that idioms are used more frequently in the American variety of English. The results of the current study can not be directly compared to the results of Liu's (2003) or Václavíková's (2010) studies due to the difference in idioms that were analysed. However, the results of all three studies show which idioms are used considerably more than others. These insights can help learners of English choose which idioms to study first. A student focusing on British variety of English could focus on learning the idioms that have a high relative frequency in the BNC first. Additionally, since the results show that anger idioms are often used in British English it could be useful to study anger idioms. On the other hand, a student focusing on American variety of English could, at first, focus on learning idioms that have a high relative frequency in COCA. Further corpus-based studies could be done on idioms to find out which

idioms are useful to study. Further research could focus on other emotions and compare their usage difference in British and American English as well as their overall usage frequency.

CONCLUSION

Idioms make language use more vivid and they are frequently used to express emotions. However, learning and understanding idioms can be one of the hardest things for a language learner. Furthermore, materials such as textbooks or idiom dictionaries can oftentimes contain many idioms that are not frequently used in daily conversations. This makes it hard for a language learner to decide which idioms to learn. It would be useful to know which idioms are frequently used because then the more frequently used idioms could be learned first. The thesis at hand analysed the usage frequency of idioms with emotional connotations in British and American variety of English. Idioms regarding happiness, sadness, and anger emotions were chosen for the analysis from the book *English Idioms in Use* by McCarthy and O'Dell (2002). The Corpus of Contemporary American English and British National Corpus were used for frequency counts.

The first chapter of the current thesis focused on idioms. Different definitions for an idiom were given as well as the dimensions that help to define an idiom. The second section of the first chapter focused on metaphors and emotion, explaining why metaphors are so important and how the thought process is metaphorical. Furthermore, examples of metaphorical source domains for idioms were given, such as *happy is up*: *We had to cheer him up*. The third section in the first chapter outlined which idioms were used in the current thesis; additionally explaining, from where and why these idioms were chosen. The fourth section focused on previous corpus-based studies that have been done on idioms. The works of Liu (2003) and Václavíková (2010) were described. The conclusions of their studies were given along with examples of the most frequently occurring idioms they found.

The second chapter described the methodology that was used in the current study. The first section focused on corpora, describing, and explaining what a corpus is. Descriptions of the two corpora – COCA and BNC – were given as well as the comparison

of the two. The second section explained why searching idioms from a corpus is more complicated than simply searching a word from a corpus. Examples of regular expressions used in the current thesis were given. The final section of Chapter 2 focused on absolute and relative frequencies. Furthermore, it was explained why and when to use relative frequencies as well as how to calculate relative frequency.

26 emotion idioms were analysed in the thesis at hand. Out of these 26 idioms ten were happiness idioms, seven sadness idioms, and nine anger idioms. Happiness idioms with the highest relative frequency in COCA were *Make your day* and *Get a kick out of something*. This suggests that a person, who wants to learn American variety of English, should learn the aforementioned idioms first instead of first learning an idiom that is not frequently used, such as *Thrilled to bits*. Idioms *Be over the moon* and *Make your day* had the highest relative frequency in BNC, suggesting that it would be prudent to learn these idioms first when endeavouring to learn the British variety of English. It would be unhelpful to learn the idiom *Do something for kicks*, which had the lowest relative frequency in BNC.

The results of the analysis suggest that a person interested in learning American English should learn the idioms *Sour grapes* and *It's not the end of the world*. The results of the current thesis show that the idiom *A/Misery guts* could be studied later because it is not frequently used in American English. A person interested in learning British English could start with learning the idioms *Sour grapes* and *Out of sorts*.

Lastly, out of the seven sadness idioms, the idiom *A/Misery guts* could be learned by a person who is interested in learning the British variety of English because this is the most frequent idiom in BNC. Anger idioms were largely more frequent in BNC. A person wanting to learn the British variety of English could first focus on anger idioms *Drive someone up the wall* and *Rub someone up the wrong way* because these two idioms had the highest relative frequency in BNC. Idioms that could be learned later are *Ruffle someone's feathers*

and *Give someone an earful* which had the lowest relative frequencies. A person wanting to learn American English could focus on anger idioms *Give someone an earful* and *Ruffle someone's feathers* which had the highest relative frequencies and later the idiom *Put/Send the cat among the pigeons* because the latter had the lowest relative frequency in COCA.

The analysis between the spoken and written part of COCA showed that idioms are used with similar frequency in both the written and spoken language. The most frequent happiness idiom in the spoken part of COCA was *Make your day*. Out of the sadness idioms, *Sour grapes* had the highest relative frequency in the spoken part of COCA. Out of the anger idioms, the idiom with the highest relative frequency was *Give someone an earful*. Overall, there were fewer anger idioms in the spoken part of COCA supporting the finding of the current study that anger idioms are less frequent in American English than in British English. In Spoken BNC:2014 the idiom *Drive someone up the wall* had the highest relative frequency.

The current study focused on emotion idiom frequencies, analysing the data collected from BNC and COCA. There were idioms that were frequently used both in COCA and BNC but there were also idioms that had a significant usage difference. Future research could focus more on emotion idioms, for example, a longer and more thorough list of happiness idioms could be analysed. Furthermore, research about overall idiom frequency could be helpful for language learners in order to make it easier for them to decide which idioms to study first. Such research could help a language learner, depending on whether they want to learn the British or American variety of English, to choose which idioms to learn first.

LIST OF REFERENCES

Primary sources

Davies, Mark. 2008. The Corpus of Contemporary American English (COCA): 1990-present. Available at http://corpus.byu.edu/coca/.

Hardie, Andrew. 2007. The British National Corpus (XML Edition): 1994- present. Available at https://cqpweb.lancs.ac.uk/bncxmlweb/.

McCarthy, Michael and O'Dell, Felicity. 2002. *English Idioms in Use*. Cambridge: Cambridge University Press.

Secondary sources

Anderson, Wendy and Gorbett, John. 2017. *Exploring English with Online Corpora- Second Edition*. United Kingdom: Palgrave.

Cambridge International Dictionary of Idioms. 1998. United Kingdom: Cambridge University Press.

Dictionary of Idioms- Helping Learners with Real English. 1995. London: HarperCollins Publishers.

Fehr, Beverly & Russell, James A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology*: General, 113: 3, 464–486.

Fellbaum, Christiane. 2007. *Idioms and Collocations- Corpus-based linguistic and Lexicographic Studies*. London: Continuum.

Fernando, Chitra. 1996. *Idioms and idiomaticity*. Oxford: Oxford University Press.

Hardie, Andrew. 2012. CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17: 3, 380–409.

Katz, Jerrold J. and Postal, Paul M. 1963. The semantic interpretation of idioms and sentences containing them. *MIT Research Laboratory of Electronic Quarterly Progress Report*, 70, 275-282.

Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. England: Addison Wesley Longman Limited.

Kövecses, Zoltán. 2000. *Metaphor and Emotion*. United States of America: The University of Cambridge.

Kövecses, Zoltán. 2013. The Metaphor-Metonymy Relationship: Correlation Metaphors Are Based on Metonymy. *Metaphor and Symbol*, 28: 2, 75-88.

Lakoff, George and Johnson, Mark. 2003. *Metaphors we live by*. London: The university of Chicago Press.

Lakoff, George and Kövecses, Zoltán. 1987. The cognitive model of anger inherent in American English. In Dorothy Holland and Naomi Quinn (ed). *Cultural Models in Language and Thought*, 195-221. Cambridge: Cambridge University Press.

Langlotz, Andreas. 2006. *Idiomatic Creativity*. Amsterdam: John Benjamins Publishing Company.

Liu, Dilin. 2003. The Most Frequently Used Spoken American English Idioms: A Corpus

Analysis and Its Implications. TESOL Quarterly, 45: 4, 661-688.

Longman American Idioms Dictionary. 1999. England: Longman Publishing.

Love, R., Dembry, C., Hardie, A., Brezina, V. and McEnery, T. (2017). The Spoken BNC2014: designing and building a spoken corpus of everyday conversations. In International Journal of Corpus Linguistics, 22(3), pp. 319-344

Oxford Dictionary of Idioms. 1999. Oxford: Oxford University Press.

Steinvall, Anders. 2007. Color and emotions in English. In Robert E. MacLaury, Galina V. Paramei and Don Dedrick (ed). *Anthropology of Color: Interdisciplinary multilevel modelling*, 347-362. Sweden: Department of Modern Languages.

Vaclav, Brezina. 2018. *Statistics in Corpus Linguistics- A Practical Guide*. United Kingdom: Cambridge University Press.

Václavíková, Eva. 2010. *Idioms of Colour – A Corpus-based Study*. Master's Diploma Thesis. Masaryk University: Faculty of Arts.

APPENDIX 1: List of idioms and regular expressions

IDIOM	Regular expression used in COCA	Regular expression used in BNC
A misery guts	misery guts	misery guts
Be floating/walking on air	walk* _i* air + float* _i* air	{be/V} float* on air + {be/V} walk* on air
Be in seventh heaven	v* _i* seventh heaven	{be/V} in seven* heaven
Be on cloud nine	v* on cloud nine	{be/V} on cloud nine
Be over the moon	v* over the moon	{be/V} over the moon
Be/feel on top of the world	v* on top of the world	{be/V} on top of the world + {feel/V} on top of the world
Do something for kicks	do* for kick*	{do/V} * for kicks
Down in the dumps	down in the dumps	down in the dumps
Drive someone up the wall	dr*v* _p* up the wall	{drive/V} * up the wall
Drive/send someone round the bend/twist	dr*v* _p* round the bend + dr*ve _p* round the twist + sen* _p* round the bend + sen* _p* round the twist	{drive/V} * round the bend + {drive/V} * round the twist + {send/V} * round the bend + {send/V} * round the twist
Get a (real) kick out of something	g*t* a kick	{get/V} a {kick/N} out of +{get/V} a real {kick/N}out of
Get/put someone's back up	g*t * 's back up + put * 's back up	{get/V} * 's back up + {put/V} * 's back up
Give someone a piece of your mind	_v* * a piece of * mind	{give/V} * a piece of * mind
Give someone an earful	_v* * an earful	{give/V} * an earful
It's not the end of the world	_p* _v* not the end of the world + _nn* _v* not the end of the world	not the end of the world
Jump for joy	jump* _i* joy	{jump/V} for joy
(Just) grin and bear it	grin and bear _p*	grin and bear *
Not be on speaking terms	not _v* on speak* term*	not * on speak* term*
Out of sorts	out of sort*	Out of sorts
Put/send the cat among the pigeons	_v* the cat amon* the pigeon*	{put/V} the cat among* the pigeons + {send/V} the cat among* the pigeons
Puts a damper on	v* a damper on	{put/V} a damper on

Rub someone up the wrong way	rub* _p* up the wrong way	{rub/V} * up the wrong way
Ruffle someone’s feathers	ruffl* _app* feathers + ruffl* * ‘s feathers	{ruffle/V} * 's feathers
Something makes your day	ma* _app* day	{make/V} _{PRON} day
Sour grapes	sour grapes	sour grapes
Thrilled to bits	thrill* to bits	thrilled to bits

APPENDIX 2: List of idioms, absolute frequency, and normalized frequency

	COCA		BNC	
Idioms	**Absolute frequency**	**Normalized frequency**	**Absolute frequency**	**Normalized frequency**
Happiness idioms				
Be floating/walking on air	191	0.19	27	0.24
Be in seventh heaven	19	0.02	9	0.08
Be on cloud nine	66	0.07	18	0.16
Be over the moon	161	0.16	59	0.53
Be/feel on top of the world	327	0.33	36	0.32
Do something for kicks	10	0.01	2	0.02
Get a (real) kick out of something	945	0.94	22	0.20
Jump for joy	158	0.16	19	0.17
Something makes you day	1052	1.05	45	0.40
Thrilled to bits	5	0.00	22	0.20
Sadness idioms				
A misery guts	5	0.00	7	0.06
Down in the dumps	152	0.15	18	0.16
It is not the end of the world	408	0.42	21	0.19
Just grin and bear it	102	0.10	25	0.22
Out of sorts	338	0.34	34	0.30
Puts a damper on	329	0.33	8	0.07
Sour grapes	489	0.49	48	0.43
Anger idioms				
Drive someone up the wall	89	0.09	16	0.14
Drive/send someone round the bend/twist	4	0.00	11	0.10
Get/put someone's back up	2	0.00	5	0.04
Give someone a piece of your mind	58	0.06	13	0.12
Give someone an earful	92	0.09	3	0.03
Not be on speaking terms	3	0.00	4	0.04
Put/send the cat among the pigeons	1	0.00	10	0.09
Rub someone up the wrong way	4	0.00	14	0.12
Ruffle someone's feathers	91	0.09	3	0.03

APPENDIX 3: List of idioms, absolute frequency, and normalized frequency in written and spoken part of COCA

	COCA WRITTEN		COCA SPOKEN	
Idioms	**Absolute frequency**	**Normalized frequency**	**Absolute frequency**	**Normalized frequency**
Happiness idioms				
Be floating/walking on air	191	0.19	20	0.16
Be in seventh heaven	19	0.02	1	0.01
Be on cloud nine	66	0.07	10	0.08
Be over the moon	161	0.16	24	0.19
Be/feel on top of the world	327	0.33	50	0.39
Do something for kicks	10	0.01	1	0.01
Get a (real) kick out of something	945	0.94	116	0.91
Jump for joy	158	0.16	31	0.24
Something makes you day	1052	1.05	101	0.79
Thrilled to bits	5	0.00	1	0.01
Sadness idioms				
A misery guts	5	0.00	0	0.00
Down in the dumps	152	0.15	25	0.20
It is not the end of the world	408	0.42	43	0.34
Just grin and bear it	102	0.10	12	0.09
Out of sorts	338	0.34	32	0.25
Puts a damper on	329	0.33	40	0.31
Sour grapes	489	0.49	70	0.55
Anger idioms				
Drive someone up the wall	89	0.09	1	0.01
Drive/send someone round the bend/twist	4	0.00	0	0.00
Get/put someone's back up	2	0.00	0	0.00
Give someone a piece of your mind	58	0.06	12	0.09
Give someone an earful	92	0.09	19	0.15
Not be on speaking terms	3	0.00	0	0.00
Put/send the cat among the pigeons	1	0.00	0	0.00
Rub someone up the wrong way	4	0.00	0	0.00
Ruffle someone's feathers	91	0.09	8	0.06

APPENDIX 4: List of idioms, absolute frequency, and normalized frequency in BNC and Spoken BNC:2014

	BNC		BNC:2014	
Idioms	**Absolute frequency**	**Normalized frequency**	**Absolute frequency**	**Normalized frequency**
Happiness idioms				
Be floating/walking on air	27	0.24	1	0.09
Be in seventh heaven	9	0.08	0.00	0
Be on cloud nine	18	0.16	1	0.09
Be over the moon	59	0.53	9	0.79
Be/feel on top of the world	36	0.32	1	0.09
Do something for kicks	2	0.02	0	0.00
Get a (real) kick out of something	22	0.20	2	0.18
Jump for joy	19	0.17	1	0.09
Something makes you day	45	0.40	11	0.96
Thrilled to bits	22	0.20	2	0.18
Sadness idioms				
A misery guts	7	0.06	1	0.09
Down in the dumps	18	0.16	0	0.00
It is not the end of the world	21	0.19	20	1.75
Just grin and bear it	25	0.22	2	0.18
Out of sorts	34	0.30	1	0.09
Puts a damper on	8	0.07	0	0.00
Sour grapes	48	0.43	3	0.26
Anger idioms				
Drive someone up the wall	16	0.14	22	1.93
Drive/send someone round the bend/twist	11	0.10	11	0.96
Get/put someone's back up	5	0.04	3	0.26
Give someone a piece of your mind	13	0.12	1	0.09
Give someone an earful	3	0.03	0	0.00
Not be on speaking terms	4	0.04	0	0.00
Put/send the cat among the pigeons	10	0.09	3	0.26
Rub someone up the wrong way	14	0.12	3	0.26
Ruffle someone's feathers	3	0.03	1	0.09

RESÜMEE

TARTU ÜLIKOOL

ANGLISTIKA OSAKOND

Shara Hannamary Kull

Corpus-based analysis of idioms describing emotions in British and American English.

Korpuspõhiline emotsioone kirjeldavate idioomide analüüs Briti ja Ameerika inglise keeles.

Bakalaureusetöö

2020

Lehekülgede arv: 47

Annotatsioon:

Käesolev bakalaureusetöö uurib õnne, kurbuse ja viha emotsioone kirjeldavate idioomide kasutussagedust Briti ja Ameerika inglise keeles kasutades sageduste leidmiseks kahte korpust – Corpus of Contemporary American English (COCA) ja British National Corpus (BNC). Töö eesmärgiks on kindlaks teha kas ja kui palju erineb idioomide kasutussagedus Briti ja Ameerika inglise keeles. Samuti analüüsida, kas idioome esineb rohkem suulises või kirjalikus keeles.

Töö jaguneb kahte ossa: esimene, teoreetiline osa selgitab, mis on idioomid ning mis on korpus, andes ülevaate antud bakalaureusetöös kasutatud korpuste kohta. Samuti on välja toodud analüüsitavate idioomide nimekirjad ning töös kasutatud regulaaravaldised ja suhtelise sageduse arvutused. Töö teises osas on välja toodud analüüsi tulemused, mida illustreerivad joonised.

Töös analüüsiti kokku 26 emotsiooniidioomi; nendest kümme olid õnne, seitse kurbuse ning üheksa viha idioomi. Esmalt leiti mõlemast korpusest regulaaravaldisi kasutades iga idioomi absoluutne sagedus. Leitud sagedused kanti Exceli tabelisse. Seejärel arvutati iga idioomi suhteline sagedus. Kahe korpuse suhtelisi sagedusi võrreldi omavahel. Selgus, et õnne ning kurbuse emotsiooni kirjeldavate idioomide suhteline sagedus oli suurem COCAs, mis viitab sellele, et neid idioome kasutatakse sagedamini Ameerika inglise keeles. Viha emotsiooni kirjeldavate idioomide suhteline sagedus oli suurem BNCs, mis viitab sellele, et viha idioome kasutatakse sagedamini Briti inglise keeles. Võrdlus COCA suulise ja kirjaliku osa vahel näitas, et idioomide kasutussagedus on nii kirjalikus kui suulises keeles üsna sarnane; kirjaliku osa suhtelised sagedused olid vaid veidi suuremad. Võrdlus BNC ja Spoken BNC:2014 vahel näitas, et idioomide suhteline kasutussagedus on suurem Spoken BNC:2014 korpuses.

Märksõnad:

Inglise keel, korpuspõhine võrdlus, idioomid

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Shara Hannamary Kull,

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

Corpus-Based Analysis of Idioms Describing Emotions in British and American English,

mille juhendaja on Jane Klavan, PhD,

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

[allkiri]

Shara Hannamary Kull

Tartus, 26.05.2020

Autorsuse kinnitus

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

[Autori allkiri]

Shara Hannamary Kull

Tartus, 26.05.2020

Lõputöö on lubatud kaitsmisele.

[Juhendaja allkiri]

Jane Klavan

Tartus, 26.05.2020