

Johanna Tolonen

The *fake news* phenomenon

A Corpus Linguistic Analysis of the term *fake news*

Faculty of Information and Communication Sciences

Bachelor's Thesis

August 2019

Tolonen, Johanna: *The fake news phenomenon: A Corpus Linguistic Analysis of the term fake news*

Kandidaatintutkielma, 22 sivua+liitteet

Tampereen yliopisto

Kielten kandidaattiohjelma, Englannin opintosuunta

Elokuu 2019

Tämä tutkimus käsittelee *fake news*-termin nopeaa kansainvälistä leviämistä ja sitä, minkälaisen sanojen yhteydessä termi useimmiten esiintyy. *Fake news* on nopeasti saavuttanut maailmanlaajuisen ilmiön maineen viimeisen kuluneen kolmen vuoden aikana. Termin ilmiömäinen kansainvälinen leviäminen näkyy esimerkiksi siten, että *fake news* valittiin vuonna 2017 vuoden sanaksi.

Tutkimuksen tavoitteena on selvittää News on the Web (NOW)- ja Corpus of Contemporary American English (COCA)-korpuksia hyödyntäen, mitkä verbit ja adjektiivit esiintyvät useimmiten *fake news*-termin välittömässä läheisyydessä. Tutkimuksessa keskitytään tarkastelemaan termin esiintymistä lauseessa verbien ja adjektiivien jälkeen. Tutkimuksen tarkoituksena on vastata seuraaviin kysymyksiin: 1) Miten termin *fake news* käyttö on muuttunut elokuun 2016 ja joulukuun 2017 välisenä aikana? 2) Mitkä tekijät ovat vaikuttaneet mahdolliseen muutokseen? 3) Mitkä adjektiivit ja verbit esiintyvät useimmiten termin *fake news* kanssa? 4) Mitä korpuksiin pohjautuvat tutkimukset voivat paljastaa kielenkäytön muutoksista?

Tutkimuksen aineistona toimi News on the Web (NOW) sekä Corpus of Contemporary American English (COCA)-korpuksien, joiden tarkoituksena on välittää mahdollisimman laaja ja reaaliaikainen kuva englannin kielen käytöstä ympäri maailmaa. Tältä osin tutkimus nojaa metodina viime vuosina yleistyneeseen korpuspohjaiseen diskurssianalyysiin.

Korpusaineisto osoittaa, että termi on levinnyt nopeasti vuodesta 2017 alkaen. Syynä termin käyttömäärän nopeaan kasvuun voidaan nähdä vuoden 2016 Yhdysvaltojen presidentinvaalit ja tästä johtunut termin kasvanut käyttö mediassa. Korpusaineistosta ilmenee lisäksi, että termin yhteydessä käytettyjen adjektiivien ja verbien voidaan nähdä heijastelevan asenteita ilmiötä kohtaan.

Asiasanat: corpus linguistics, neologism, etymology, compounding, corpus-based study

Table of contents

1. Introduction	1
2. Literary Review	3
2.1 What is fake news?	3
2.2 Dictionary meaning of <i>fake news</i>	5
2.3 Compounding, Neologism and Polysemy	6
2.4 Background of the phenomenon	7
3. Methods and Material	9
3.1 Corpus linguistics	9
3.1.1 The News on the web (NOW) corpus	10
3.1.2 The Corpus of Contemporary American English	10
3.2 Quantitative and qualitative analysis	11
3.3 Material	12
4. Results and Discussion	14
4.1 Results in COCA	14
4.2 Adjectives occurring with the term in the NOW corpus	15
4.3 Verbs occurring with the term in the NOW corpus	18
5. Conclusion	21
Bibliography	23
APPENDIX A	25
APPENDIX B	25
APPENDIX C	25
APPENDIX D	26
APPENDIX E	26

APPENDIX F	26
APPENDIX G	26
APPENDIX H	26

1. Introduction

The term *fake news* has become a global phenomenon over the past three years. *Fake news* can be seen all over social media, news and the Internet. Although the term itself is technically not new, nowadays it has lost some of its original meaning, as meaning only incorrect news, developed into something different and gained a further meaning. With the help of the News on the web (NOW) corpus and the Corpus of Contemporary American English (COCA), changes in the use of the term can be detected and examined in detail.

The topic of this BA thesis is the change and development in the use of the term *fake news*. This study will investigate which verbs and adjectives are most commonly used with *fake news*. The study will focus on verbs and adjectives occurring immediately before the term. Moreover, because of the vast number of words occurring with the term, the study will focus on the five most commonly used adjectives and verbs. Although this will not give an all-encompassing view of the use of *fake news*, it will provide a glimpse to the use of the phenomenon. Also, because this study focuses only on adjectives and verbs occurring right before the term, the results and conclusions of the study have been made from this point of view and are thus restricted to only cover the pre-phrasal place of the term. In addition, the verbs can occur in the corpora in any form, but the most common verb forms are infinitive, past tense and gerund. Interpretations on the use of the term will be made based on the results of the studies conducted. The study will focus on a specific point in time from the 1st of July 2016 to the 31st of December 2017.

Fake news was very current in the year 2018 and something that has been drawing attention to itself on the Internet over the past years. Although *fake news* has been very visible for the last three years or so, there are no previous studies covering the words occurring with the term. This is one of the reasons that influenced this study. The reason for the lack of previous studies might be that, although the term is not new, the phenomenon concerning the term is. This study will provide

an angle and a perspective on the use of the term and clarify how and in what context people tend to use the term. Moreover, studying the spreading of terms and language overall is always important. It is also beneficial to increase the knowledge of how people use language in real life. The study will include material from all around the world. A further study could focus for example on the use and proliferation of the term only in the United States.

The research questions for this study are: How has the use of the term *fake news* changed between July 2016 and December 2017 both in terms of frequency and the ways in which it is used? What factors can be argued to have contributed to this change? What are the most commonly used adjectives and verbs with the term *fake news*? And finally, what can corpus-based studies reveal from the changes that happen in languages?

The thesis is divided into five sections. The second section of the thesis will deal with the explanation of the term *fake news*. Neologism, polysemy, the etymology and background information of the term as well as explanation on the word formation process of compounding are covered in section 2. The third section of the thesis explains the methods and material used in the study. Section 3 also presents an introduction to corpus linguistics and quantitative and qualitative methods. In addition, the section describes the characteristics of the two corpora used in the study, the News on the Web (NOW) corpus and the Corpus of Contemporary American English (COCA). Section 4 will present the results and the discussion on the major findings of the study. Lastly, a concluding section summarizes the main findings of the study, together with discussion and views on the prospects of further study.

2. Literary Review

This section addresses the main concepts and theories central to the thesis. Concepts like neologism, polysemy and compounding are explained, and the etymology of the term *fake news* is discussed. Different interpretations and definitions of *fake news* will also be provided. The sources for these interpretations and definitions are English dictionaries including *Oxford English Dictionary*, *Merriam-Webster Dictionary*, *Collins English Dictionary* and *Cambridge Dictionary*.

2.1 What is fake news?

The name of the term *fake news* gives a good clue on the meaning of the term. *Fake news* is news that are simply incorrect. Moreover, *fake news* means news, information or stories that are written with the intention of harming other people or entities. The information, news or stories spreading as *fake news* are written with the same form as real news, following the correct structure and principles of real news. Therefore, it is sometimes difficult to identify whether the news is real or not. *Fake news* often deals with current topics and handles political issues and events like real news but can also deal with other kinds of themes, like completely falsified information of historical events. One of the problems that have arisen from the rapid spreading of *fake news* is that because anything and everything can nowadays be claimed fake news, the line between real and fake news has started to disappear. This then leads to a situation where people do not know which news is real and which is not.

Before dealing with the recent phenomenon of the term it is necessary to discuss the history or the etymology of the term. The etymology of *fake news* is, however, somewhat unclear and consists of many steps. Firstly, it is essential to explain what the term *etymology* means. “Etymology is the investigation of word histories. It has traditionally been concerned most especially with those word histories in which the facts are not certain, and where a hypothesis has to be constructed to account

either for a word's origin or for a stage in its history" (Durkin 2009, 15) and that "The term is also used more broadly to describe the whole endeavour of attempting to provide a coherent account of a word's history (or pre-history)" (Durkin 2009,16). The term *fake news* consists of two different words, an adjective *fake* and a noun *news*. To be able to understand the term properly, it is necessary to examine the two parts of the term separately. *Oxford English Dictionary* defines the word *fake* as something that is 'not genuine; imitation or counterfeit' (OED s.v. *fake*). The adjective *fake* according to *Oxford English Dictionary* has first appeared in the late 18th century. The second word of the term *news* is a noun that is a plural form but in construction it is singular. According to the *Oxford English Dictionary*, the first known use of *news* was in the late Middle English, in the 15th century (OED s.v. *news*). The definition of the word in *lexico.com* is 'newly received or noteworthy information, especially from the recent events' (*lexico.com* s.v. *news*).

Although the term *fake news* has existed in the English language for a long time, and has been used in the media regularly, it has started to appear more visibly in the media from the latter half of the year 2016 to the end of the year 2017. According to *Merriam-Webster Dictionary*, although there has always been fake news, the actual term *fake news* was first used around the end of the 19th century. One of the earliest occurrences of the term *fake news* is "Secretary Brunnell Declares Fake News About His People is Being Telegraphed Over the Country" [*Merriam-Webster Online Dictionary*, Cincinnati Commercial Tribune (Cincinnati, OH), 7 Jun. 1890]. Before the term *fake news* was invented, there were other words that described the same phenomenon of misinformation or incorrect news. One of the most commonly used word in place of *fake* was *false*, so *fake news* used to be called *false news*. The addition of the adjective *fake* into the English language generated the currently known term *fake news* (*Merriam-Webster Online Dictionary*, "the real story of 'fake news'").

During the past years, the term *fake news* has gone through a semantic change, in which the meaning of the word changes to something different. In today's world, *fake news* has become not

only a global phenomenon, but also a verbal political weapon. When earlier *fake news* had a fairly neutral tone, it now has a more negative connotation. The recent phenomenon and its development circle around the fact that nowadays *fake news* has been known to be used in situations where a person dislikes the information spread about him/herself. The news might not be incorrect, but the person the news is concerning, dislikes or disapproves it which leads to them calling the news *fake news*. This semantic change also known as *broadening* of meaning has brought the term to everyone's attention globally.

2.2 Dictionary meaning of *fake news*

The term *fake news* has not been incorporated into many dictionaries yet. Multiple dictionaries were examined for the definitions of the term. Although, as mentioned in the previous section, the term has existed in the English language a long time, only two online dictionaries had a definition for the term *fake news*. The dictionaries that had incorporated a definition of *fake news* were *Collins English Dictionary* and *Cambridge Dictionary*.

Collins English Dictionary defines the term *fake news* as 'false, often sensational, information disseminated under the guise of news reporting' (CED s.v. *fake news*). *Fake news* was also named as the word of the year 2017 by Collins English Dictionary. The word of the year is classified as the most important word in a specific year. This naming of the term as the word of the year supports the important role of the word in the English language.

Cambridge Dictionary defines *fake news* as 'false stories that appear to be news, spread on the internet or using other media, usually created to influence political views or as a joke' (*Cambridge Dictionary* s.v. *fake news*). Both online dictionaries define *fake news* similarly as news that is incorrect and spread as real news but the difference between the two online dictionaries is that *Cambridge Dictionary* has added the political aspect in its definition of *fake news*.

2.3 Compounding, Neologism and Polysemy

The term *fake news* can be classified as a compound and a neologism. Compounding is a form of word formation where two individual words are combined to form a completely new word. “. . . a compound is a word that is made up of two or more elements, the first of which is either a word or a phrase, the second of which is a word” (Benczes 2006, 8). *Fake news* has been formed by compounding the adjective *fake* with the noun *news*. A compound is constructed from a “head” and a “modifier”. The “head” of a compound is the second word and the “modifier” is the first word of the combination. The “head” word determines the part of speech of the compound word. For example, because the “head” of the term *fake news* is “news”, the term is classified as a noun. Compounding is very frequently used in languages such as English and German, but not as commonly used in languages like Spanish and French. Words can be formed by compounding quite extensively and productively. Compounding is not only restricted to nouns but can also be used between adjectives or between an adjective and a noun, like *fake news* (Yule 2010, 55).

Michael Picone (1996) talks about neologism in his book *Anglicisms, Neologisms and Dynamic French*. He notes that “A neologism is any new word, morpheme or locution and any new meaning for a preexistent word, morpheme or locution that appears in a language” (Picone 1996, 3). *Merriam-Webster Dictionary* defines neologism as ‘a new word, usage, or expression’ (*Merriam-Webster Online Dictionary s.v. neologism*). *Fake news* has not been created recently, but because it has gained a new meaning and expanded from its original preexisting meaning, it can be classified as a neologism.

Not only is the term *fake news* a compound and a neologism but it is also a polyseme. When a word or a phrase has two or more meanings it can be classified as a polyseme. “Polysemy can be defined as one form (written or spoken) having multiple meanings that are all related by extension (Yule 2010, 120). In the case of *fake news*, as mentioned in section 2.1, the term can mean either

incorrect news or news that are true but unpleasant to the person the news is concerning. These two meanings are different to each other but still share a common ground.

2.4 Background of the phenomenon

Generally, a rapid proliferation and increase in the use of a term is the result of some major event that concerns a large number of people. There is one major political event that happened in the United States in the latter half of the year 2016 that probably influenced the increase of the use of the term *fake news*. The 58th American presidential election was held on the 8th of November 2016. The candidates in the last phase of the election were the Democratic candidate Mrs. Hillary Diane Rodham Clinton and the Republican candidate Mr. Donald John Trump. The winner of the 58th American presidential election was Donald Trump. However, the election of Donald Trump was not the part that influenced the proliferation of the term the most. One of the major points that highlighted the entire 2016 presidential election, and gained media attention all around the world, was the large amount of gossip and incorrect news also known as *fake news* spread through social media and online newspapers with the intention to harm and damage the election campaign of an opponent.

The 2016 presidential election was historic because of the two candidates in the final stages of the election. Both Mrs. Clinton and Mr. Trump had a scandalous history behind them. Hillary Clinton used the same email server when writing important state related emails as well as her own personal emails. Donald Trump was accused of sexual misconduct. The 2016 presidential election and especially the actions of candidate Donald Trump affected the proliferation of the term *fake news*. During the election, Donald Trump called all the negative information concerning him *fake news*, whether or not is really was false or true information. This persistent use of the term and wide interest of the scandalous presidential election led to the rapid proliferation of the term globally. Donald Trump has even claimed that he created the term (*The Washington Post*, October 26, 2017).

However, this is false, because as has been proven before the term was used already in the 19th century. Mr. Trump might not have invented the term, but he surely helped the term to gain its new meaning as news that are unpleasant to a person and spread the term around the world to everyone's awareness. Because the presidential election interested people all around the world and gained large attention in the media, the proliferation of the term was inevitable. Therefore, it was also more fruitful to study the phenomenon around the world and not only in the United States.

3. Methods and Material

In this section the methods and material used in the study will be introduced. First, a general introduction to corpus linguistics is presented. After that information on the two corpora, The News on the Web (NOW) corpus and The Corpus of Contemporary American English (COCA), used in the study will be provided. Section 3.2 consists of the aspects of the quantitative and qualitative methodologies used in the study, and finally section 3.3 will focus on the material used in the study.

3.1 Corpus linguistics

“It has often been said that, through corpora, we can observe patterns in language which we were unaware of before or only vaguely glimpsed” (Facchinetti 2007, 51)

Paul Baker and Tony McEnery define corpus linguistics as “. . . a powerful methodology- a way of using computers to assist the analysis of language so that regularities among many millions of words can be quickly and accurately identified” (2015, 1). Baker and McEnery continue by calling corpus “a body of language” (2015, 1). Corpus linguistics deal with corpora, which contain thousands of different kinds of texts. When a study is conducted, the selection of a corpus can be made depending on the information needed. Because different corpora are designed to serve different purposes, there are multiple different corpora available. For example, the main source of information in this study, the News on the Web corpus, consists of online news. The other corpus used in this study, the Corpus of Contemporary American English, has data from spoken language to academic texts. With the help of corpora, a certain way in which language is used can be investigated and examined. Many corpora are updated regularly to keep the corpora valid and to provide as much information as possible.

As Facchinetti argued in 2007, it would not be possible to study different kinds of phenomena in the language without the help of corpora. Or it would be possible, but extremely time consuming

and it would probably not cover as wide a context as with the help of corpora. With the help of corpora, it is possible to detect for example how and when words have appeared or disappeared, changed or developed into something else or when and where certain words have been used.

3.1.1 The News on the web (NOW) corpus

The News on the Web corpus is currently one of the largest structurally organized corpora, with 8.2 billion words of data (August 2019) available in full-text format on the Internet. The data in the NOW corpus is collected from various newspapers and online magazines. The corpus contains data from 2010 to the present day and is updated every day. The monthly intake of words is 140-150 million words per month (Davies, 2013). From 2017 it was also possible to upload the data from the NOW corpus for offline use.

The NOW corpus was chosen for this study because of its vast size and the fact that it has information and data from newspapers and online magazines, where the term *fake news* has mostly occurred.

3.1.2 The Corpus of Contemporary American English

The Corpus of Contemporary American English was also used in this study. The corpus has more than 560 million words of text, divided into different categories. There are five categories in which the texts are divided. These categories are spoken, newspaper, fiction, academic texts and popular magazines. The corpus is freely-available and “the only large and balanced corpus of American English” (Davies, 2008-). Unlike the NOW corpus, COCA is not updated every day, so it does not contain very recent language data.

3.2 Quantitative and qualitative analysis

This study has features from both quantitative and qualitative analysis. The following section will focus on defining the quantitative and qualitative features of the study.

“The quantitative approach is used when one begins with a theory (or hypothesis) and tests for confirmation or disconfirmation of that hypothesis” (Newman and Benz 1998, 3). In quantitative studies the results are often presented in a numerical form. This study focuses on the top five adjectives and verbs with most hits in numerical order. In addition, the frequencies with each word will be turned into percentages to see how big of a portion a word takes of the entire package. Because this study consists of large amounts of numerical data, it can be said that the numerical representation of the results of this study is quantitative.

“The qualitative, naturalistic approach is used when observing and interpreting reality with the aim of developing a theory that will explain what was experienced” (Newman and Benz 1998, 3). Also “Often in a qualitative design only one subject, one case, or one unit is the focus of investigation over an extended period of time” (Newman and Benz 1998, 9-10). Because one of the aims of this study is to understand and question the reasons behind the proliferation of the term and phenomenon, the study can be classified as a qualitative analysis as well. The study will also question why certain adjectives and verbs occur with *fake news* more often than others. If this study focused purely on the increase of the term the study would be purely quantitative but because the intention of this study is to not only search whether there is a change, but also understand the reasons behind the change, the study is a mixture of both quantitative and qualitative analysis.

3.3 Material

As mentioned before, the study was executed by using primarily the News on the Web (NOW) corpus. The Corpus of Contemporary American English (COCA) was also examined to provide a larger insight on the spreading of the term. COCA includes data exclusively from American sources, and thus can only be used to study the proliferation of the term in American English. Searches in the NOW corpus include data from newspapers and articles from all around the world. The searches were conducted during the spring and summer 2018. All searches in the corpora were made multiple times in different occasions to delete any possible errors that could influence the study.

First the possible proliferation of the term in COCA was examined. The searches in COCA were conducted by using the chart option and a search string 'fake news'. The results were examined in a vertical chart to investigate the results as clearly as possible. The results mostly included content from American news sources like CNN, FOX and PBS. The search hits were thus found in the subsections of spoken language (transcribed TV and radio shows on news and current events), magazine and newspaper texts while only a few hits were found on the sections of fiction and academic texts. The material covers all the occurrences of the term in American media from 1990 to 2017.

After the proliferation of the term was examined in COCA, the relevant hits of the term *fake news* were then retrieved for closer study. Firstly, the occurrences of the term were examined. The overall time span in which the study focused on was from July 1st, 2016 to December 31st, 2017. The time span was chosen because of the political events that had occurred and the possible impact they might have brought to the term. Also, the term started to appear more visibly on the media around this time. After deciding the time period, it was divided into three timespans. The first timespan was from July 1st, 2016 to December 31st, 2016, the second from January 1st, 2017 to June

30th, 2017 and the third from July 1st, 2017 to December 31st, 2017. This way the increase in the usage and the proliferation of the term could be examined biannually to see if and when a change had happened.

After dividing the time period in question into three separate timespans, the adjectives occurring before the term were investigated. The searches in the NOW corpus were conducted by using a search string `_j* fake news` and using the list search to provide a diagram of the words. The `_j*` in the search string marks adjectives. The same setting for the occurrence of verbs were made except that the search string was `_v* fake news` in which the `_v*` marks verbs. The placing of the `_j*` and `_v*` was crucial. Because the intention of this study was to study the use of adjectives and verbs before the term `_j*` and `_v*` had to be placed before the term *fake news*. The correct time period under examination was also selected each time to exclude occurrences of the term in any other time period. In each search, the five most used adjectives and verbs were selected for further investigation. The results were given in both numerical and percentual form. The results have been rounded to the nearest tenth.

It is also important to state that the material in the two corpora was not all-encompassing and could not cover all the possible uses of the term. It would be impossible to collect all of the usage of the term from all individuals all around the world. Although some usage of the term might be missing from the two corpora, the corpora give adequate enough evidence of the phenomenon and its vastness.

4. Results and Discussion

After all the material was gathered and all the searches conducted multiple times, the use of the term and the increase in the use was observable.

4.1 Results in COCA

SECTION (CLICK FOR SUB-SECTIONS) (SEE ALL SECTIONS AT ONCE)	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
SPOKEN	484	116.7	4.15	
FICTION	3	111.8	0.03	
MAGAZINE	228	117.4	1.94	
NEWSPAPER	109	113.0	0.96	
ACADEMIC	4	111.4	0.04	
<hr/>				
1990-1994	2	104.0	0.02	
1995-1999	2	103.4	0.02	
2000-2004	14	102.9	0.14	
2005-2009	18	102.0	0.18	
2010-2014	13	102.9	0.13	
2015-2017	779	62.3	12.50	
TOTAL	828			SEE ALL TOKENS

Table 1. Occurrences of fake news in COCA, all sections

Corpus of Contemporary American English

SEARCH
CHART
CONTEXT

[CHANGE TO HORIZONTAL CHART](#)

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT
2015	5	20.0	0.25	
2016	166	20.9	7.95	
2017	607	21.2	28.57	

Table 2. Occurrences of fake news in COCA, 2015-2017.

Searches conducted in The Corpus of Contemporary American English (COCA) revealed that the term *fake news* had mostly been used in the spoken language. Tables 1 and 2 above represent the findings in COCA. The term occurred also in magazine and newspaper language, but considerably less than in spoken language. The use of *fake news* in fiction and academic texts was rather scarce. An interesting observation can be made between the timespans 2010-2014 and 2015-

2017. Before the year 2015 the use of the term *fake news* was minor. Overall, between 1990 and 2009 there were altogether only 36 recorded uses of the term. Moreover, there were two recordings of the use of the term in 2010, two in 2011, one in 2012, five in 2013 and three in 2014. Altogether 13 hits recorded between 2010 and 2014. In the year 2015 only five hits were found. When the years between 2015 and 2017 were investigated in detail a clear increase between 2015 and 2016 could be seen, as can be observed in Table 2. There were 166 hits in the year 2016 and 607 hits in the year 2017. Altogether 778 hits were found between the years 2015 and 2017. It is necessary to note that in the wider search seen in Table 1, 779 hits of the term were found between 2015 and 2017. However, when the years were examined in detail (Table 2) one hit had disappeared. The reason to this disappearance was unknown.

4.2 Adjectives occurring with the term in the NOW corpus

After the wider picture of the proliferation of the term was known, it was important to focus on the most common adjectives occurring with the term.

	CONTEXT	FREQ
1	SO-CALLED FAKE NEWS	26
2	MALICIOUS FAKE NEWS	13
3	POST-ELECTION FAKE NEWS	8
4	VIRAL FAKE NEWS	6
5	PRO-TRUMP FAKE NEWS	5
6	OTHER FAKE NEWS	5
7	POPULAR FAKE NEWS	5

Table 3. Adjectives occurring with *fake news* in the NOW July 1st, 2016- December 31st, 2016

The searches focusing on adjectives were conducted in the NOW corpus. The five most common adjectives occurring right before the term *fake news* in the first timespan from July 1st, 2016 to December 31st, 2016 in the NOW corpus were *so-called*, *malicious*, *post-election*, *viral* and *pro-Trump*. The results are represented in Table 3. The adjective with most hits *so-called* had 26 hits out of 171 which equals to 15.2% of all the hits. The adjective with second most hits was *malicious* with 13 hits and 7.6%. The third adjective *post-election* had 8 hits and 4.7%, the fourth

adjective *viral* had 6 hits and 3.5% and the fifth adjective was *pro-Trump* with 5 hits and 2.9%. The adjective with the most hits *so-called* could be interpreted as a neutral adjective meaning something that is introduced for the first time. The corpus data thus supports the idea of the perceived novelty of the term in that the surrounding context has expressions denoting this quality. However, *so-called* can also mean something or someone that is not suitable for a specific position or situation. For example, “My so-called friend bought me nothing for my birthday” means that the friend might not be a good friend after all. The adjective *malicious* has a clearly more negative tone meaning something that is done with the attempt to harm others. The fourth adjective *viral* is neutral in tone and means something that is spreading quickly usually through the Internet. Two of the five adjectives *post-election* and *pro-Trump* had something to do with elections. When the hits were studied more closely, a connection to the election of the 58th president of the United States of America could be seen. Altogether, the 171 hits found proved that the use of the term in the latter half of the year 2016 was still rather moderate. Some of the adjectives had more of a negative tone, others were more neutral, so there was no clear predominance of the other.

	CONTEXT	FREQ	TOTAL 410 UNIQUE 182 +
1	SO-CALLED FAKE NEWS	57	
2	SPREADING FAKE NEWS	25	
3	ALLEGED FAKE NEWS	11	
4	RUSSIAN FAKE NEWS	10	
5	OTHER FAKE NEWS	8	
6	MAJOR FAKE NEWS	7	
7	ACTUAL FAKE NEWS	7	

Table 4. Adjectives occurring with *fake news* in the NOW corpus January 1st, 2017-June 30th, 2017

The five most common adjectives occurring right before the term in the second timespan from January 1st, 2017 to June 30th, 2017 in the NOW corpus and represented in Table 4 were *so-called*, *spreading*, *alleged*, *Russian* and *other*. *So-called* had 57 hits out of 410 which is 13.9% from all hits. The second was *spreading* with 25 hits and roughly 6.1%, the third adjective *alleged* had 11 and roughly 2.7%, the fourth adjective *Russian* had 10 hits and roughly 2.4% and the fifth adjective *other* had 8 hits and 2%. Altogether the number of hits had more than doubled from 171 to 410 in

six months (i.e. 0.20 to 0.48 per million words). Except from the adjective *so-called* which still has the most hits, the other adjectives in the top five had changed. Although the search was set to spot only adjectives, the verb *spreading* was included in the most adjectives used with the term. When the results of the word *spreading* were investigated in detail, it was noted that in most of the occurrences of the word *spreading*, the word was truly used as a verb and not as an adjective. Although this excludes the word *spreading* from the most used adjectives in this study, the word *spreading* gives a clue of what has been happening with the term. The adjective *alleged* is used to define something accused but not proven (*Merriam-Webster online s.v. alleged*). The adjective *Russian* implying someone or something from Russia had probably increased because of Russia's alleged influence on the outcome of the 2016 election. *Other* which is both an adjective and a pronoun is “used to refer to a person or thing that is different or distinct from one already mentioned or known about” (OLD s.v *other*). *Other* was often used with a noun *sites* as in “other fake news sites” referring to multiple sites dealing with fake news.

	CONTEXT	FREQ	TOTAL 383 UNIQUE 168 +
1	SO-CALLED FAKE NEWS	43	
2	SPREADING FAKE NEWS	32	
3	RUSSIAN FAKE NEWS	15	
4	INTENSE FAKE NEWS	14	
5	ALLEGED FAKE NEWS	11	
6	REAL FAKE NEWS	9	
7	PHONY FAKE NEWS	7	

Table 5. Adjectives occurring with *fake news* in the NOW corpus July 1st, 2017- December 31st, 2017

Table 5 represents the third and last timespan from July 1st, 2017 to December 31st, 2017. In this timespan the five most common adjectives occurring right before the term in the NOW corpus were *so-called*, *spreading*, *Russian*, *intense* and *alleged*. *So-called* had 43 hits out of 383 which is roughly 11.2% of all this, *spreading* had 32 hits and roughly 8.4%, *Russian* had 15 hits and roughly 3.9%, *intense* had 14 hits and roughly 3.7% and *alleged* had 11 hits and roughly 2.9%. Once again, the word *spreading* appeared in the search. The word was again used mostly as a verb and was thus excluded from this study. In the third time span a minor decrease in the use of the term could be

seen. The overall number of hits had decreased from 410 to 383. *Spreading* and *Russian* had numerically fewer hits than before, but in terms of percentages their use had increased. Also, a new adjective *intense* “having or showing a characteristic in extreme degree” (*Merriam-Webster online s.v intense*) had occurred on the five most used adjectives. *Alleged* had the fifth most hits when previously it had the third most hits.

4.3 Verbs occurring with the term in the NOW corpus

After the use of the term with adjectives was concluded, the research continued to the use of the term with verbs.

	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 865 UNIQUE 272 +
1	IS	IS FAKE NEWS	57	
2	SPREADING	SPREADING FAKE NEWS	43	
3	PREVENT	PREVENT FAKE NEWS	37	
4	TACKLE	TACKLE FAKE NEWS	24	
5	TACKLES	TACKLES FAKE NEWS	22	
6	COMBAT	COMBAT FAKE NEWS	21	
7	SPREAD	SPREAD FAKE NEWS	20	

Table 6. Verbs occurring with *fake news* in the NOW corpus July 1st, 2016-December 31st, 2016

As can be seen in Table 6, the five most commonly used verbs occurring right before the term *fake news* in the first timespan from July 1st, 2016 to December 31st, 2016. were *is* with 57 hits 6.6%, *spreading* 43 hits 5%, *prevent* 37 hits 4.3%, *tackle* 24 2.8% and *tackles* 22 hits 2.5%. If we count *tackle* and *tackles* together as one verb, the fifth most commonly used verb would be *combat* with 21 hits 2.4%. There were altogether 865 hits with *fake news*. The most commonly used verb *is* is neutral and the third person singular form of the verb *to be*. The second verb *spreading* and the gerund form of it implies that a proliferation is happening right now. Some hits with *spreading* were also present participles. *Tackle*, *prevent* and *combat* all relate to warfare and fighting which pinpoints the ongoing atmosphere surrounding the term in this timespan.

	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 2,370 UNIQUE 444 +
1	<input type="checkbox"/>	IS FAKE NEWS	269	
2	<input type="checkbox"/>	SPREADING FAKE NEWS	117	
3	<input type="checkbox"/>	ARE FAKE NEWS	109	
4	<input type="checkbox"/>	SPREAD FAKE NEWS	89	
5	<input type="checkbox"/>	COMBAT FAKE NEWS	81	
6	<input type="checkbox"/>	FIGHTING FAKE NEWS	70	
7	<input type="checkbox"/>	SPOT FAKE NEWS	62	

Table 7. Verbs occurring with *fake news* in the NOW corpus January 1st, 2017- June 30th, 2017

The five most common verbs occurring right before the term in the second timespan from January 1st, 2017 to June 30th, 2017 represented in Table 7 were *is* with 269 hits and 11,4%, *spreading* with 117 hits and 4.9%, *are* with 109 hits and 4.6%, *spread* with 89 hits and 3.8% and *combat* with 81 hits and 3.4%. If only diverse verbs are counted and *spreading* and *spread* are counted as one verb then the fifth verb would be *fighting* with 70 hits and 3%. It is, however, interesting to note that the verb form *spread* is used to prevent and educate people from believing fake news and to not spread them but the verb form *spreading* is used to refer to the actual phenomenon and proliferation of the term. For example, “Do not create or spread fake news or unverified information. . . “(SunStar June 3, 2017), in which the verb *spread* is used to prevent the use of fake news. “Journalists are spreading fake news “(Rappler June 16, 2017) is an example where the verb *spreading* is used to refer to the actual proliferation of the term. There were altogether 2370 hits where the term *fake news* occurred. Compared to the latter half of the year 2016 the use of *fake news* had increased vastly. The number of hits had increased from 865 to 2370 hits (1.02 to 2.75 per million words), which is almost triple the amount. The verb *combat* had both numerically and percentually more hits than in the first time span. The verbs *prevent* and *tackle* had disappeared from the top hits. This implied that the proliferation and the use of the term was already established and accepted.

	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 1,768 UNIQUE 395 +
1	<input type="checkbox"/>	IS FAKE NEWS	156	
2	<input type="checkbox"/>	SPREADING FAKE NEWS	125	
3	<input type="checkbox"/>	SPREAD FAKE NEWS	82	
4	<input type="checkbox"/>	'S FAKE NEWS	68	
5	<input type="checkbox"/>	WAS FAKE NEWS	62	
6	<input type="checkbox"/>	BE FAKE NEWS	41	
7	<input type="checkbox"/>	SPOT FAKE NEWS	38	

**Table 8. Verbs occurring with *fake news*
in the NOW corpus July 1st, 2017- December 31st, 2017**

In the third timespan from July 1st, 2017 to December 31st, 2017 represented in the Table 8 the five most commonly used verbs with *fake news* were *is* with 156 hits which is roughly 8.8% of all hits, the second verb was *spreading* with 125 hits and 7.1%, the third verb was *spread* with 82 hits and 4.6%, the fourth was 's which indicates "is" with 68 hits and 3.8% and the last verb was *was* with 62 hits and 3.5%. If only different verbs are considered and *is* and 's and *spreading* and *spread* are counted as one verb, the fourth most used verb would be *be* with 41 hits and 2.3% and the fifth would be *spot* with 38 hits and 2.1%. Altogether there were 1768 hits which used the term *fake news*. The verb *spread* was often used in to+infinitive form. In the last time span the tone of the verbs was clearly more neutral. The verbs related to warfare had disappeared from the top five. It is also important to note that the overall number of hits had decreased from 2370 to 1768.

5. Conclusion

Fake news and the phenomenon with it spread through social media and news all over the world between the years 2016 and 2017. However, based on the results, it can be claimed that the use of the term *fake news* has slowly started to decline. The peak in the use of the term was in the earlier part of the year 2017 when the number of *fake news* occurring with adjectives had more than doubled and the number of *fake news* with verbs had more than tripled. Despite the vast increase in the use of the term in the first part of the year 2017, in both adjectives and verbs the number of hits declined in the latter half of the year 2017.

There were no huge differences between the tones of the adjectives used with the term in any time span. However, a difference in the tones of the verbs used with the term could be seen. The verbs had a more neutral tone in the latter part of the year 2017 than in the latter part of the year 2016. In addition, more forms of the verb *be* were used and verbs connected to warfare were used less. In one year, the world had gotten used to *fake news* and instead of fighting against it concentrated on teaching people how to deal with it. Although the use of the term has started to decline *fake news* still continues to spread around the world but in a less prominent manner. The use of the term has now decreased to a point where it cannot be spotted daily or even weekly on online magazines or social media anymore.

The meanings of words can be narrowed or broadened, words are borrowed and lent from other languages and sometimes entire languages disappear. This study proves that corpus-based studies can detect changes in the use of terms, phrases and words. Moreover, corpus-based studies can provide a statistic view of a change in the use of words, phrases and terms. Corpus-based studies also enable the examination of large amounts of texts and material that in other ways would be either impossible or extremely time-consuming. The purpose of this study was to examine *fake news* in two different corpora and study the proliferation of the term. It is always useful to study phenomena and this way broaden people's knowledge of certain aspects of language. There are 7.53

billion people on Earth. Approximately 600-1700 million people speak English. These people do not speak English the same way. There are different dialects, accents, jargons and slangs. In addition, although standard English is taught in schools all over the world, the language people actually use varies depending on time, place and situation. For example, people born in the 1960's speak differently to people born in the 21st century. Corpora can reflect attitudes and other aspects of language use that may not seem evident purely on an intuitive basis. Without corpora the way people actually use language would be difficult to study. It is however important to note that corpora are not impeccable and can contain flaws. When a corpus-based study is executed, the results must be studied carefully and cannot be trusted blindly.

Although conclusions can be made based on the results, there were some restrictions in the study that affected the results and thus must be noted. The study was conducted with a limited set of material and data, including only adjectives and verbs occurring with *fake news* and using only two corpora. Moreover, the time period in question was restricted to only cover the latter half of 2016 to the end of 2017. This deliberate selection of certain word classes and time period leaves room for further studies. The topic of further studies could be nouns or adverbs occurring with *fake news* or different words occurring after the term. Also, the use of the term only in or outside United States could be interesting to examine.

The proliferation of *fake news* has had some interesting effects on people. *Fake news* as a concept and the recognition of fake news has become an important subject in schools where students of all ages are taught to identify and thus avoid incorrect news. The phenomenon has brought the media awareness and consciousness to everybody's attention. It is very important to teach today's Internet generation what to believe and what not to believe when they read news. All things considered some of them are true and some of them are truly fake news.

Bibliography

- Baker, Paul and McEnery, Tony, eds. 2015. *Corpora and Discourse Studies: Integrating Discourse and Corpora*. Basingstoke: Palgrave Macmillan.
- Benczes, Réka. 2006. *Creative Compounding in English: The Semantics of Metaphorical and Metonymical Noun-noun Combinations*. Amsterdam: John Benjamins Publishing Co.
- Borchers, Callum. 2017. Trump falsely claims (again) that he coined the term 'fake news'. [Internet] The Washington Post. Available from https://www.washingtonpost.com/news/the-fix/wp/2017/10/26/trump-falsely-claims-again-that-he-coined-the-term-fake-news/?noredirect=on&utm_term=.fd98f931719a [Accessed 11.01.2019]
- Cabaero, Nini B. 2017. Cabaero: Fake news as a business. [Internet]. SunStar. Available from <https://www.sunstar.com.ph/article/149474/Lifestyle/Cabaero-Fake-news-as-a-business>. [Accessed 24.8.2019]
- Cambridge Dictionary*. 2018. [Internet] Cambridge University Press. [Accessed 18.4.2018]
- Collins English Dictionary*. 2018. [Internet] Collins [Accessed 18.4.2018] (CED)
- “Fake news” Def. 1. Dictionary.cambridge.org. Cambridge Dictionary. n.d. [Accessed 18.04.2018]
- “Fake news” Def. 1. Collinsdictionary.com. Collins Dictionary. n.d. [Accessed 18.04.2018]
- Cornelio, Jayeel Serrano. 2017. The decline of critical thinking. [Internet]Rappler. Available from <https://www.rappler.com/thought-leaders/173064-decline-critical-thinking> [Accessed 24.8.2019]
- Davies, Mark. 2013. *Corpus of News on the Web (NOW): 3+ billion words from 20 countries, updated every day*. Available online at <https://corpus.byu.edu/now/>. [11.01.2019]

- Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Available online at <https://corpus.byu.edu/coca/> [11.01.2019]
- Durkin, Philip. 2009. *Oxford Guide to Etymology*. Oxford: Oxford University Press.
- Facchinetti, Roberta. 2007. *Corpus Linguistics 25 years on.*, Amsterdam/New York, NY: Rodopi B.V.
- Hunt, Julia. 2017. 'Fake news' named Collins Dictionary's official Word of the Year for 2017. [Internet]Independent. Available from <https://www.independent.co.uk/news/uk/home-news/fake-news-word-of-the-year-2017-collins-dictionary-donald-trump-kellyanne-conway-antifa-corbynmania-a8032751.html> [Accessed 18.4.2018]
- Lexico.com*. 2018. Oxford University Press. [Accessed 24.8.2019]
- Merriam-Webster online Dictionary* .2018. *The real story of 'fake news'*. [Internet] Available from: <https://www.merriam-webster.com/words-at-play/the-real-story-of-fake-news> [Accessed 18.4.2018]
- Merriam-Webster online Dictionary*. 2018. Merriam-Webster, Incorporated. [Accessed 18.4.2018]
- Newman, Isadore and Benz, Carolyn R. 1998. *Qualitative-quantitative Research Methodology: Exploring the Interactive Continuum*. Carbondale, Ill: Southern Illinois University Press.
- Oxford English Dictionary Online*. 2017 Oxford University Press. [Accessed 18.4.2018] (OED)
- Oxford Living Dictionary Online*. 2018 Oxford University Press. [Accessed 18.4.2018] (OLD)
- Picone, Michael D. 1996. *Anglicisms, Neologisms and Dynamic French*. Ed. Jean-Claude Chevalier. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Yule, George. 2010. *The Study of Language 4th edition*. Cambridge: Cambridge University Press.

APPENDIX A

Occurrence of fake news in COCA, all sections

SECTION (CLICK FOR SUB-SECTIONS) (SEE ALL SECTIONS AT ONCE)	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
SPOKEN	484	116.7	4.15	
FICTION	3	111.8	0.03	
MAGAZINE	228	117.4	1.94	
NEWSPAPER	109	113.0	0.96	
ACADEMIC	4	111.4	0.04	
1990-1994	2	104.0	0.02	
1995-1999	2	103.4	0.02	
2000-2004	14	102.9	0.14	
2005-2009	18	102.0	0.18	
2010-2014	13	102.9	0.13	
2015-2017	779	62.3	12.50	
TOTAL	828			SEE ALL TOKENS

APPENDIX B

Occurrence of fake news in COCA, 2015-2017

Corpus of Contemporary American English

SEARCH CHART CONTEXT

CHANGE TO HORIZONTAL CHART

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT
2015	5	20.0	0.25	
2016	166	20.9	7.95	
2017	607	21.2	28.57	

APPENDIX C

Adjectives occurring with *fake news* in the NOW July 1st, 2016- December 31st, 2016.

	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	
1	<input type="checkbox"/>	SO-CALLED FAKE NEWS	26	
2	<input type="checkbox"/>	MALICIOUS FAKE NEWS	13	
3	<input type="checkbox"/>	POST-ELECTION FAKE NEWS	8	
4	<input type="checkbox"/>	VIRAL FAKE NEWS	6	
5	<input type="checkbox"/>	PRO-TRUMP FAKE NEWS	5	
6	<input type="checkbox"/>	OTHER FAKE NEWS	5	
7	<input type="checkbox"/>	POPULAR FAKE NEWS	5	

APPENDIX D

Adjectives occurring with *fake news* in the NOW corpus January 1st, 2017-June 30th, 2017.

	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 410 UNIQUE 182 +
1	<input type="checkbox"/>	SO-CALLED FAKE NEWS	57	
2	<input type="checkbox"/>	SPREADING FAKE NEWS	25	
3	<input type="checkbox"/>	ALLEGED FAKE NEWS	11	
4	<input type="checkbox"/>	RUSSIAN FAKE NEWS	10	
5	<input type="checkbox"/>	OTHER FAKE NEWS	8	
6	<input type="checkbox"/>	MAJOR FAKE NEWS	7	
7	<input type="checkbox"/>	ACTUAL FAKE NEWS	7	

APPENDIX E

Adjectives occurring with *fake news* in the NOW corpus July 1st, 2017- December 31st, 2017.

	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 383 UNIQUE 168 +
1	<input type="checkbox"/>	SO-CALLED FAKE NEWS	43	
2	<input type="checkbox"/>	SPREADING FAKE NEWS	32	
3	<input type="checkbox"/>	RUSSIAN FAKE NEWS	15	
4	<input type="checkbox"/>	INTENSE FAKE NEWS	14	
5	<input type="checkbox"/>	ALLEGED FAKE NEWS	11	
6	<input type="checkbox"/>	REAL FAKE NEWS	9	
7	<input type="checkbox"/>	PHONY FAKE NEWS	7	

APPENDIX F

Verbs occurring with *fake news* in the NOW corpus July 1st, 2016- December 31st, 2016.

	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 865 UNIQUE 272 +
1	<input type="checkbox"/>	IS FAKE NEWS	57	
2	<input type="checkbox"/>	SPREADING FAKE NEWS	43	
3	<input type="checkbox"/>	PREVENT FAKE NEWS	37	
4	<input type="checkbox"/>	TACKLE FAKE NEWS	24	
5	<input type="checkbox"/>	TACKLES FAKE NEWS	22	
6	<input type="checkbox"/>	COMBAT FAKE NEWS	21	
7	<input type="checkbox"/>	SPREAD FAKE NEWS	20	

APPENDIX G

Verbs occurring with *fake news* in the NOW corpus January 1st, 2017-June 30th, 2017.

	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 2,370 UNIQUE 444 +
1	<input type="checkbox"/>	IS FAKE NEWS	269	
2	<input type="checkbox"/>	SPREADING FAKE NEWS	117	
3	<input type="checkbox"/>	ARE FAKE NEWS	109	
4	<input type="checkbox"/>	SPREAD FAKE NEWS	89	
5	<input type="checkbox"/>	COMBAT FAKE NEWS	81	
6	<input type="checkbox"/>	FIGHTING FAKE NEWS	70	
7	<input type="checkbox"/>	SPOT FAKE NEWS	62	

APPENDIX H

Verbs occurring with *fake news* in the NOW corpus July 1st, 2017- December 31st, 2017.

	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 1,768 UNIQUE 395 +
1	<input type="checkbox"/>	IS FAKE NEWS	156	
2	<input type="checkbox"/>	SPREADING FAKE NEWS	125	
3	<input type="checkbox"/>	SPREAD FAKE NEWS	82	
4	<input type="checkbox"/>	'S FAKE NEWS	68	
5	<input type="checkbox"/>	WAS FAKE NEWS	62	
6	<input type="checkbox"/>	BE FAKE NEWS	41	
7	<input type="checkbox"/>	SPOT FAKE NEWS	38	