



Article

Text Analysis Methods for Misinformation–Related Research on Finnish Language Twitter

Jari Jussila ^{1,*} , Anu Helena Suominen ² , Atte Partanen ¹ and Tapani Honkanen ³

¹ HAMK Smart Research Unit, Häme University of Applied Sciences, P.O. Box 230, 13100 Hämeenlinna, Finland; atte.partanen@hamk.fi

² Faculty of Management and Business, Tampere University, P.O. Box 300, 28100 Pori, Finland; anu.suominen@tuni.fi

³ School of Entrepreneurship and Business, Häme University of Applied Sciences, P.O. Box 230, 13100 Hämeenlinna, Finland; tapani.honkanen@hamk.fi

* Correspondence: jari.jussila@hamk.fi; Tel.: +358-504-657-285

Abstract: The dissemination of disinformation and fabricated content on social media is growing. Yet little is known of what the functional Twitter data analysis methods are for languages (such as Finnish) that include word formation with endings and word stems together with derivation and compounding. Furthermore, there is a need to understand which themes linked with misinformation—and the concepts related to it—manifest in different countries and language areas in Twitter discourse. To address this issue, this study explores misinformation and its related concepts: disinformation, fake news, and propaganda in Finnish language tweets. We utilized (1) word cloud clustering, (2) topic modeling, and (3) word count analysis and clustering to detect and analyze misinformation-related concepts and themes connected to those concepts in Finnish language Twitter discussions. Our results are two-fold: (1) those concerning the functional data analysis methods and (2) those about the themes connected in discourse to the misinformation-related concepts. We noticed that each utilized method individually has critical limitations, especially all the automated analysis methods processing for the Finnish language, yet when combined they bring value to the analysis. Moreover, we discovered that politics, both internal and external, are prominent in the Twitter discussions in connection with misinformation and its related concepts of disinformation, fake news, and propaganda.

Keywords: disinformation; misinformation; fake news; propaganda; Twitter; Finland; Finnish language; infodemic; social media



Citation: Jussila, J.; Suominen, A.H.; Partanen, A.; Honkanen, T. Text Analysis Methods for Misinformation–Related Research on Finnish Language Twitter. *Future Internet* **2021**, *13*, 157. <https://doi.org/10.3390/fi13060157>

Academic Editor:
Arcangelo Castiglione

Received: 20 April 2021
Accepted: 10 June 2021
Published: 17 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Across disciplines, there is an increasing interest in misinformation, which is an umbrella term referring to false information circulating online [1]. However, there is also limited understanding of why certain individuals, societies, and institutions are more vulnerable to disinformation, i.e., the intentional spread of misinformation [2]. Furthermore, fake news is considered one of the information disorders of misinformation and disinformation, a notably effective vehicle for disinformation [3]. Recently, the dissemination of disinformation and fabricated content, e.g., in the form of fake news on social media such as Twitter, is a growing concern, especially due to the lack of awareness of the existence of such false information [4–7]. The severity of this concern has expanded as younger generations choose social media sources over journalistic ones for their information [5]. Different disciplines have approached the subject of misinformation from various viewpoints of human behavior and communication, as well as the spreading enabled by information systems and social media. As examples of those presented approaches are research on cognitive biases related to information literacy [8], disinformation processing, e.g., concerning sentence comprehension and semantic decision making [9,10]. In the field of communication research, disinformation is studied primarily from the perspective of

political communication because its functions are particularly visible and effective, especially in election campaigns. The study of propaganda has a long history in communication research, but disinformation was rarely a topic of primary analysis before 2017 [11]. Since then, communication research on disinformation has widely increased, and the study of social media has played a special role in this [12]. In information systems, research in big data is a popular trend that deals with high volume, high velocity, high variety, and high veracity information assets that require new forms of processing for enhanced decision making and insight discovery [13]. In social media research, it has been identified that conversations, actors involved in the conversations, and the interactions between the actors are relevant dimensions for the dissemination of disinformation [14,15]. Furthermore, the term “fake news” serves as a discursive device for ordinary citizens to consolidate group identity in everyday political utterances on Twitter [16]. The rapid spread of false information on Twitter has especially raised concerns [3,17,18]. So far, research is still theoretically scattered, and there is a need for a data-driven understanding of the phenomena. The research has especially focused on the role of social media in facilitating the dissemination of fake news. However, it has been shown that social media also serves as a networked public space for opposing parties to define, contest, and strategically leverage the “fake news” label to their respective interests, thus challenging democracy with a weaponized discourse of fake news [16]. A total of 33% of Finns tend to trust or totally trust news and information they receive through online social networks and messaging apps [19]. Thus, investigating the spread of misinformation, disinformation, fake news, and propaganda in the discourse of Finnish Twitter is a relevant research topic. Previous research has also identified that users have difficulty in recognizing manipulation of information in news [20], which supports the need for further research on the topic. Methods to study and analyze social media data, also regarding misinformation, are currently evolving alongside social media applications. Thus, we focus on Twitter data, especially in Finnish Twitter. Therefore, in this article, we aim to study misinformation and its related concepts in the context of Twitter, test the functionality of a few analysis methods for Twitter data in the Finnish language, examine the discussed themes linked with misinformation and its related concepts in Finnish tweets, and propose some promising future research directions on misinformation in social media. This study is limited to the investigation of the text content of tweets [21]. Limiting the investigation only on the text content of tweets was chosen to uncover the inherent challenges of text mining Finnish tweet contents and to understand what kind of insights could be drawn from text only. The Finnish language has special grammatical characteristics since word formation occurs with the addition of endings (bound morphemes, suffixes) to word stems. In addition, derivation and compounding are two ways of forming new words from existing words and stems, which makes the analysis of words and text, as well as machine learning of the language, more complex than with English, for example. Therefore, our goals are to find functional methods of analysis as well as discussion themes that are linked with misinformation-related concepts in Finnish Twitter data. To reach our goals, we seek to answer the following research question: “What are the functional methods to analyze Finnish Twitter data to find out discussion themes that are linked with misinformation and concepts related to it?”

The article is formulated as follows: First, in the introduction, we give background regarding our research problem and present the research question. Second, we describe the key concepts related to misinformation. Third, we discuss our methodological choices for our research. Fourth, we illustrate our research results. Fifth, we discuss our results in light of the literature and present practical implications together with further research suggestions. And sixth, we draw our conclusions.

2. Overview of Key Concepts

Simultaneously with the COVID-19 pandemic, the world is struggling with an “infodemic”. Infodemic is information overload about a problem, which makes discovering a solution cumbersome, and it can spread misinformation and disinformation [22,23].

“Misinformation” and “disinformation,” together with “fake news,” are related concepts, now excessively utilized both by academics and other information providers. The proliferation in the use of the concepts and the lack of generally agreed definitions, has led to interchangeable use [23]. Misinformation is defined as false, mistaken, or misleading information [24], and as a type of claim that can be verified and that has been confirmed to be false [23]. It refers to forms of factually inaccurate information that is propagated inadvertently [25], often considered as an honest mistake [4]. Some studies use misinformation as an umbrella term when referring to false information circulating online [1]. Misinformation, disinformation, and fake news are three distinct concepts, although similar characteristics are detectable in the definitions used. Other potentially conceptually overlapping with misinformation are propaganda, conspiracy theories, and false rumors [23]. Disinformation, in addition to the falseness of the information, includes the purposeful intention by the sender or information provider, and thus distributed, asserted, or disseminated—often online—to mislead, deceive, or confuse [1,24–27]. The communication of misinformation has a social function: (1) it is comprehended as a type of collective sense-making, and (2) it subverts outgroups or rivals [23]. Moreover, disinformation is considered a direct manifestation of this function, containing deliberately deceptive and propagated information that purposefully weakens public support of a competing body, e.g., as in the U.S. presidential election campaigns [23].

Fake news is one of the information disorders of misinformation and disinformation [3]. It is a notably effective vehicle for disinformation. Fake news is fabricated information that mimics news media content in form [3], i.e., disguised as journalistic articles [27,28]. Fake news is pushed onto newsfeed platforms [27,28] like Twitter. Although its form resembles news, the organizational process or intent of fake news does not. The outlets lack editorial norms and processes of the news media, where accuracy and credibility of information are ensured [3]. Thus, fake news exploits the credibility, timeliness, and the variety of sensitive topics of journalism [27,28]. At its core, it is perniciously parasitic, benefiting from the standard media outlets, yet undermining their credibility [3]. The difficulty in differentiating fake news from traditional news is due to their similarity. They both share an inverted pyramid format, timeliness, negativity and prominence, and subjects such as politics and government. The only two differences are the lack of objectivity, including the opinion of their author(s) and the value of impact because fake news sites focus on trivial stories [29]. Thus, the recipient, journalist, or reader of a news article usually cannot recognize the manipulation [20].

As a concept, fake news avoids any agreed definition [1,30], although it goes back centuries, and it became a buzzword particularly after the 2016 presidential elections in the United States [28,29]. Therefore, it currently has a political flavor [1,3,17], but vaccination, nutrition, and the stock market are also topics of fake news [3]. The research on fake news has focused a lot on the “fakeness,” i.e., the deception as well as the motivation of the deceivers [29]. Furthermore, the focus of the research has been on the role of social media in facilitating the dissemination of fake news. Yet, it has been shown that social media also serves as a networked public space for opposing parties to define, contest, and strategically leverage the “fake news” label to their respective interests, thus challenging democracy with the weaponized discourse of fake news. Moreover, the term “fake news” serves as a discursive device for ordinary citizens to consolidate group identity in everyday political utterances on Twitter [16].

Misinformation has an adverse societal impact because it accelerates propaganda, creates anxiety, induces fear, and sways public opinion [1]. Unfortunately, its speed, geographical reach, depth, and breadth exceed those of information, yet humans seldom detect misinformation [1,17]. The social function of misinformation, i.e., sense-making, determines that ambiguous or potentially threatening contexts, such as crisis, is suitable soil for disseminated false information to breed and flourish, especially as fertilized by the information paucity and anxiety of people [23]. The active role of the audience in fake news is also emphasized. The audience seems to be involved in its co-construction, as the receiver

determines the realness or fakeness of the news. When fake news is regarded as fake (i.e., fiction) by the receiver, the deception process has failed. Yet, when the deception process succeeds, the legitimacy of journalism is taken advantage of. The co-creative role of the audience is accentuated in the social media context due to the information exchange that comprises the negotiation and sharing of meanings. In social media, socialness expedites the construction of fake news by mitigating its penetration into social spheres. Social spheres are strengthened by information exchange, which potentially sidelines the quality of information. Therefore, the legitimizing role of the audience in future studies is called for [28].

Social media is regarded as a powerful global medium that influences people's perceptions of the world and their role in it [31]. Social media has raised critical concerns as context due to its popularity [3] as well as the rapid spread of fake news, thus causing potentially detrimental effects both on individuals and society [17,32]. In particular, the rapid spread of false information on Twitter has caused concern [3,17,18].

On Twitter, the diffusion of falsehood (i.e., retweeting) is significantly farther, faster, deeper, and broader than the truth in all categories of information. This is particularly prevailing in the topic of politics. Two issues were conspicuous concerning fake news on Twitter. Firstly, the novelty that suggests that people share novel information. Secondly, the negative emotions that false stories inspired in replies: fear, disgust, and surprise. The rate of acceleration by robots ("bots") was the same for both true and false news. That implies that the spread of false news is caused by humans [17]. However, social bots (automated accounts impersonating humans) do play a role in magnifying the spread of fake news by orders of magnitude by liking, sharing, and searching for information [3]. The bot population on Twitter has been estimated to range from 9% to 15%.

A recent exploratory study on Finnish Twitter on the scope of politically motivated abusive language focused on the extent to which it is perpetrated by inauthentic accounts, i.e., bots [33]. The results demonstrated that the messaging is mostly with very low levels of both bot and coordinated activity, i.e., it is human-induced. Furthermore, the minority of the identified bots were operating in the Finnish language, while most were using English in an automated or semi-automated manner. The themes that triggered abusive language concerning politics were government corruption and failure, sexism and homophobia, racism and Islamophobia, government handling of COVID-19, and education (in the context of COVID-19). Therefore, it is interesting to explore whether the same or similar themes are likewise related to misinformation on Finnish Twitter in general. A third of Finns tend to trust or totally trust news and information they receive through online social networks and messaging apps, according to the Flash Eurobarometer [19]. Therefore, investigating the spread of misinformation, disinformation, fake news, and propaganda on Finnish Twitter is a relevant research topic, as it has been recognized that the recipient (whether journalist or reader) usually cannot recognize the manipulation [20].

3. Materials and Methods

The methodology of the study is composed of Twitter data collection and analysis by a combination of the computational social science approach [34] and content analysis with Atlas.ti [35]. The description of the methodology follows the recommendations for Twitter data acquisition and analysis by Dann [21]. The computational social science approach is used to create a word cloud—a list of most frequently appearing words and a topic model of the collected tweets. A more detailed word count analysis in Atlas.ti 9 and clustering of keywords and theme words are then conducted to gain a deeper understanding of the discussions.

3.1. Methodology Issues Related to the Language Used

The Finnish language is a member of the Finno-Ugric language family and closely related to Estonian. In 1997, Finnish was the native language of 92.7% of Finland's population of 5.15 million people. The Finnish language does not have grammatical gender or

articles. The basic principle of word formation in Finnish is the addition of endings (bound morphemes, suffixes) to stems. Furthermore, Finnish verb forms are built up in the same way. Often, the endings are piled up one behind the other rather mechanically. The adding of endings to a stem is a morphological feature of many European languages, but Finnish is nevertheless different from most others in two respects: (1) English nouns have only one “morphologically marked” case, but Finnish has more case endings than is usual in European languages—about 15 cases (nominative, accusative, genitive, partitive, inessive, elative, illative, adessive, ablative, allative, essive, translative, abessive, comitative, and instructive). Finnish endings normally correspond to the prepositions or postpositions in other languages; (2) Finnish sometimes uses endings, where Indo-European languages generally have independent words. For example, cases in the Finnish noun “auto” (“car” in English) is formed with the endings: auto/ssa, (“in the car”), auto/sta (“out of the car”), auto/on (“into the car”), auto/lla (“by car”), and by attaching the plural ending -I, as in auto/i/ssa (“in the cars”). Finnish possessive suffixes correspond to possessive pronouns, such as -ni (“my”), -si (“your”), -mme (“our”)—for example, auto/ssa/ni (“in my car”), auto/ssa/si (“in your car”), auto/ssa/mme (“in our car”). An example of using the verb stem sano- (“say”) and the endings -n (“I”), -i (past tense), and -han (for emphasis), verbs can be formed: sano/n (“I say”), sano/n/han (“I do say”), sano/i/n (“I said”), and sano/i/n/han (“I did say”). Another set of endings particular to Finnish is that of the enclitic particles, which always occur in the final position after all other endings, used mainly for emphasis: -kin (“too”, “also”) auto/ssa/si/kin (“in your car too”), -han (for emphasis: “you know, don’t you?”), and -ko (English interrogative): On/ko tuo auto? (“Is that a car?”). Moreover, a characteristic feature of Finnish is the wide-ranging use of endings to form new words. For example, kirja (“book”) and its derived forms kirj/e (“letter”), kirja/sto (“library”), kirja/llinen (“literary”), kirja/llis/uus (“literature”), kirjo/itta(a) (“(to) write”), and kirjo/itta/ja (“writer”). When adjectives occur as attributes, they agree in number and case with the headword, i.e., they take the same endings—for example, isossa autossa (“in the big car”) [36].

There are two ways to form new words from existing words and stems in Finnish: (1) derivation and (2) compounding. In derivation, new words (word stems) are made by adding derivative endings or suffixes to the root or another stem, e.g., to adjective kaunis: kaunii- (“beautiful”), add the ending -ta to form the derived verb stem kaunis/ta- (“beautify”), the first infinitive kaunis/ta/a. Similarly, to the verb stem aja- (“drive”), add the ending -o to form the derived noun aj/o (“drive”, “chase”, “hunt”) or the ending -ele- to form the verb stem aj/ele- (“drive around”), the first infinitive aj/el/la. The most common type of compound word is made up of two non-derived nouns, and typical compounds are written without spaces, e.g., autokauppa (“car sales”). The first noun in these compounds is often in the genitive, e.g., auto/n/ikkuna (“car window”) [36].

In text analytics, many research projects rely on software libraries—such as the Natural Language Toolkit (NLTK)—that provide functionalities for language processing, e.g., stemmers and sentence tokenizers [37]. However, as outlined above, Finnish language stemming often leads to word stems that are not valid words, e.g., “kirj” but are also meaningless. Previous studies in classifying Finnish social media text have discovered that lemmatization functions significantly better than stemming for Finnish [38]. However, the most widely used software packages do not support lemmatization in the Finnish language.

3.2. Data Collection

Data collection was made with the Twitter Streaming API collecting tweets from a list of keywords specified as disinformaatio, huhu, misinformaatio, propaganda, uutisankka, valeuutinen (in English: disinformation, rumor, misinformation, propaganda, hoax, fake news). As the boundary, there is the specified filter to only look at tweets that contain the Finnish language. Tweets containing at least one of the keywords were stored in a database. The collecting of tweets was automated using the Tweepy python library [39], which gains access to the API and real-time stream of tweets. The database chosen to collect

Tweets was MongoDB, which is a NoSQL database capable of handling unstructured data. Custom code was implemented using Python to run in a virtual machine [40]. Tweets were collected from 23 June 2020, to 23 March 2021. A total of 43,890 tweets were collected that contained at least one of the defined keywords.

Development has been done with the Twitter API version 1, but the upcoming API version 2 makes it possible to replicate the data collection with new endpoints, which allows the user to search the complete archive [41]. With this method, it is possible to search tweets starting from 2006.

Twitter data are collected in full format into a MongoDB total. The total field count is 145, and for further analysis, only the most relevant fields are used. The fields selected for further analysis are documented in Table 1.

Table 1. Used data fields in the analysis. (Twitter API JSON FORMAT).

Variable Identification	Type	Description
created_at	String	UTC time when this tweet was created.
Text	String	The actual UTF-8 text of the status update.
extended_tweet.full_text	String	Untruncated text message when longer than 140 characters.
entities.hashtags	array	Represents hashtags that have been parsed out of the tweet text.
user.id_str	String	The string representation of the unique identifier for this user.
user.screen_name	String	The screen name, handle, or alias that this user identifies themselves with. screen_names are unique but subject to change.
user.description	String	The user-defined UTF-8 string describing their account.

3.3. Data Processing

Twitter data processing started with reading an Excel spreadsheet that was extracted from MongoDB to a Python Pandas data frame. Data were formatted to have only one text field for text content. This was done because the Twitter API provides two text fields for text content (see Table 1). The default text field is made with the rule that if the tweet text is longer than 140 characters, then the text is truncated. Another text field provided by the API contains the extended tweet full text up to 280 characters [42]. In the case of truncated text, a custom script was developed to combine the default tweet text field and extended tweet text field to one field. Then the combined text field was processed with the Python regular expression operations listed in Table 2 to have text as general as possible.

Table 2. Text processing steps.

Action	Description
1. Remove links	Remove links to reduce unstructured text by removing links.
2. Make all letters lower case.	All letters are converted into lower case because the analysis is case-sensitive.
3. Remove punctuation, digits, and special markers	Removing all punctuation to reduce unstructured text and numbers does not usually change the meaning of the text. Removing special markers, usually @, is commonly used when a user is mentioned.
4. Remove white spaces	All unnecessary white spaces are removed.

3.4. Word Count Analysis in Atlas.ti 9

The word count analysis to detect the manifestation of disinformaatio, misinformaatio, propaganda, uutisankka, valeuutinen (disinformation misinformation, propaganda, hoax, and fake news) and the clustering of themes related to the keywords were carried out in two phases (Table 3). The first phase was processed using Atlas.ti research analysis software version 9. The process began with importing the data from an Excel spreadsheet that was extracted from MongoDB and preprocessed to include only the most relevant fields. Next, the general terms in the Finnish stoplist and general terms such as “https:\\” were excluded. After that, the list of 47,013 words was exported to Excel, where the two-phased exclusions and inclusion were executed, resulting in 602 words that had a minimum of 50 manifestations in tweets. In the end, these words were clustered into 88 clusters, which varied from 5706 manifestations of propaganda to 50 manifestations of aivopesu (“brain wash”), kaksinaismoraali (“double standards”), and vihreäpropaganda (“Green propaganda”).

Table 3. The process of word count analysis in Atlas.ti and the spreadsheet.

Action	Inclusion/Exclusion Criteria	Quantity of Data
1. Data import from Excel to Atlas.ti	inclusion 16,463	16,463 documents (tweets)
2. General terms to the stoplist	general Finnish stop terms 746 in Atlas.ti and other general words and Twitter display names	47,013 words
3. Wordlist with quantities to Excel spread list	Inclusion criteria minimum 50 tweets per word	468 words
4. Exclusion	Exclusion of 213 + 37 general words and Twitter display names	218 words
5. Inclusion	385 derived or compounded from stem word	602 words
6. Clustering	Clustering of words	88 clusters

4. Results

The results of the study were created using custom-developed Python scripts and Atlas.ti 9 software. The results generated with developed Python scripts include a word cloud of tweets, the top 30 most frequent words in tweets, and a topic model of tweets. Atlas.ti was used to compute a more extensive word count and to export discovered main clusters, which were further cleaned and processed in an Excel spreadsheet. The combination of computational analysis with Python and qualitative data analysis software was performed to gain a deeper understanding of the discussions related to misinformation, disinformation, fake news, and propaganda.

4.1. Word Cloud of Tweets

A word cloud of tweets is used to illustrate and describe the data collected for the study (Figure 1). A word cloud contains the data in raw format without stemming, lemmatization, or the use of other natural language processing techniques.



Figure 1. Word cloud of tweets.

As seen from the word cloud propaganda, valeutinen (fake news), disinformaatio (disinformation), yleutiset (Yleisradio, “Yle”, is the Finnish Broadcasting Company, uutiset means news), Suomen (Finland’s), suomessa (in Finland), mariaohisalo (Maria Ohisalo is the Minister of the Interior of Finland), perussuomalaiset (Finns Party), persut (slang for Finns Party), Russia, venäjän (Russia’s), venäjä (Russia), and the Finnish Prime Minister Sanna Marin spelled in various ways (e.g., marinsanna, marin, marin) are among the biggest terms. From the processed tweet text (according to Table 2) of the word cloud, the top 30 most frequently used words are listed in Table 4.

Table 4. Top 30 most frequent words in tweets.

Word	Freq
Propaganda	4804
Valeutinen	1676
Disinformaatio	1500
Suomen	492
Yleutiset	461
mariaohisalo	421
Leviää	358
Suomessa	354
Yle	354
Russia	334
Dimmu	328
Persut	326
propagandaa	306
Ylen	287
Saa	275
Osa	268
Media	267
astatenhunen	267
Pitää	266
Mm	255
Uutisankka	253
Keronen	249
Amp	248

Table 4. Cont.

Word	Freq
Marinsanna	244
Hsfi	225
Twitterissä	217
Venäjän	216
Hallituksen	206
Journalismi	205
Somessa	197

The top 30 most frequent words in tweets point out that stemming would have worked only for five of the words, i.e., yle (stem), yleuutiset ja ylen, propaganda (stem), and propagandaa. None of the top 30 most frequent words gives any indication of what the conversations were about. Instead, they describe the organization or context related to the conversations, e.g., public broadcasting company, Russia, government, etc.

4.2. Topic Model of Tweets

Topic Modeling and Latent Dirichlet Allocation (LDA) was next performed to gain an overview of topics and the similarity between topics. The topic model was trained with a total of 15 topics. The top 10 words are listed in Table 5. Data collected using the "rumor" keyword were excluded from the topic model as it referred mostly to unrelated discussions in foreign languages.

Table 5. Top 10 words in 15 topics.

Topic 1	Topic 4	Topic 7	Topic 10	Topic 13
propaganda	propaganda	propaganda	propaganda	propaganda
mariaohisalo	disinformaatio	valeuutinen	disinformaatio	valeuutinen
somessa	valeuutinen	leviää	yle	hai
jaa	marinsanna	persut	mariaohisalo	govt
valeuutisia	hai	mariaohisalo	kansainvälistä	disinformaatio
linkkejä	lapset	disinformaatio	ritken	sanoin
sitaatteja	govt	liittyen	ajoista	mattimuukkonen
voida	suomen	opparviainen	vietetään	paikassa
alkuperää	pitäisi	sisäministeri	synkistä	vaiennus
todistaa	yleisradio	sosiaalisi	kansanmurhien	lakia
Topic 2	Topic 5	Topic 8	Topic 11	Topic 14
propaganda	propaganda	propaganda	propaganda	propaganda
valeuutinen	disinformaatio	disinformaatio	valeuutinen	valeuutinen
disinformaatio	valeuutinen	vaarallista	russia	leviää
mariaohisalo	tuli	päivää	suomen	persut
leviää	vihapuhe	levinnyt	propagandaa	kuvaa
liittyen	mm	syystä	leviää	govt
sosiaalisi	pari	pari	disinformaatio	hai
sisäministeri	yle	valheellinen	mariaohisalo	disinformaatio
opparviainen	suomen	liandersson	the	somealustoilla
propagandaa	kuntavaalit	kuvamanipulaa	opparviainen	kannattajia
Topic 3	Topic 6	Topic 9	Topic 12	Topic 15
propaganda	propaganda	propaganda	propaganda	propaganda
valeuutinen	disinformaatio	valeuutinen	astatenhunen	valeuutinen
disinformaatio	valeuutinen	lakia	susi	kysymyksiä
leviää	yleuutiset	sensuuripykälä	disinformaatio	mariaohisalo
yleuutiset	russia	mattimuukkonen	valeuutinen	sanna
media	toimii	vaiennus	suomen	leviää
ylen	korona	mieltä	syy	keronen
mariaohisalo	lapset	sanoin	vihreät	nuorten
opparviainen	venäjän	paikassa	mm	hallituksen
liittyen	suomessa	rikotaan	amp	suulla

Some of the discussions can be identified based on the topic model centered around people, e.g., Minister of Interior (topics 1–3), political parties, e.g., Finns Part (topic 14), elections (topic 5), Russian children’s rights in Finland (topic 6), censorship (topic 9), and wolves and the Green League. However, for many of the topics, a more detailed investigation of the tweets would be needed to determine the connections between the words on the topic model.

Figure 2 presents the results of the topic model that was manually modified by combining synonymous terms. For instance, the Minister of the Interior was referred to both by title and by name (Twitter account) in the original data, and several indicated social media, such as Facebook, social media, and social media platforms, which were categorized under one term. Furthermore, to maintain anonymity, some terms that referred to individual persons’ accounts were replaced by a term indicating their profile on Twitter, e.g., Politician * and Student **.

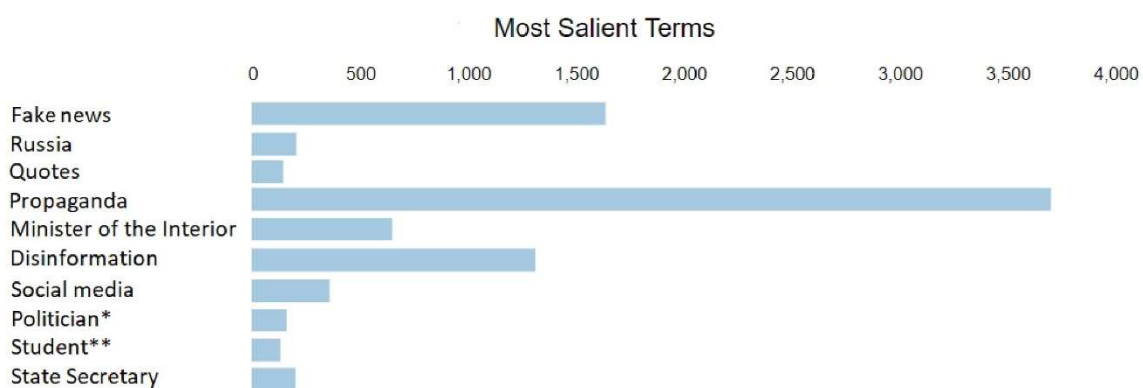


Figure 2. Topic model of Top-10 most salient terms in tweets. * Anonymized Twitter account of a politician, ** anonymized Twitter account of a student.

Obvious terms, such as fake news, propaganda, and disinformation, were found among the most salient terms in the topic model (Figure 2). However, misinformation was not included in the Top-10 or Top-30 most salient terms. National politics was a common topic, with frequent references to the Minister of the Interior, state secretary of the Ministry of the Interior, and politically active persons. Russia, as a country, emerged among the most salient terms as well. Surprisingly, an individual student was also among the most salient terms. Overall, the automated topic model functioned rather poorly for the Finnish language, which necessitated the manual combination of terms and cleaning the automatically generated figures.

4.3. Word Count Analysis and Clustering

The results of the word count analysis carried out in Atlas.ti and clustered in the Excel spreadsheet are presented in Tables 6 and 7. In Table 6, the main 22 clusters of 88 are presented. First, the five keywords are presented, together with the cluster of related words, such as lie, waddle, truth, and fact. Second, the 16 main theme word clusters, which include more than 200 manifestations in tweets, are presented.

The “Media” and “Social media” clusters included mention of various media and social media producers: newspapers and television channels, certain journalists, as well as general terms regarding news, such as news in Finnish “uutinen”. The “Finland” cluster included various Finland-related issues, such as words referring to the country Finland, and also Finns and Finnish language. Similarly, the “Country and world” cluster contains words that refer to the country (Fin. maa) and the world (Fin. maailma). The “Animals” cluster comprises various animals. “Children and young people” cluster consists of words referring to child or children (Fin. lapsi, lapset) or young people (Fin. nuoret). The “Police” and “Researcher and Research” clusters include words referring to those occupations. “Opposition” (Fin. vastaisuus) and “Hate speech” included words referring to those terms.

Table 6. Result of word count analysis.

Main Clusters in English	Quantity of Word Manifestations
Keyword clusters	
1. Propaganda (Fin. propaganda)	5706
2. Fake news (Fin. valeuutinen or fake news)	2119
3. Disinformation (Fin. disinformaatio)	1991
4. Hoax (Fin. uutisankka)	256
5. Misinformation (Fin. misinformaatio)	198
6. Lie, Waddle, Truth, Fact (Fin. vale, valhe, huuhaa, totuus, fakta)	1274
Theme word clusters	
7. Media	5692
8. Politics	3821
9. Foreign countries	2769
10. Politician	2101
11. Finland	1601
12. Health (corona, vaccination, virus)	1538
13. Social media	1416
14. Animals	520
15. Children and young people	507
16. Movements (Qanon, Isis, Elokapina)	453
17. Country and World	340
18. School	290
19. Police	253
20. Researcher and Research	247
21. Opposition (Fin. vastaisuus)	228
22. Hate Speech	201

Table 7. Results in detail of three major word clusters.

The Content of Three Main Clusters in English	Quantity of Manifestations
Clusters	
Politics	3821
Finns Party (Fin. persu, persujen, persut, perussuomalaiset, perussuomalaisten, perussuomalaisiin, ps, ps:n)	848
The Greens and left-wing greens (Fin. vihreät, vihreat, vihreiden, vihreille, vihreitä, vihreistä, vihreitä, vihreä, vihreän, vihervasemmisto, vihervasemmiston, vihervasemmisto'lainen)	564
Right-wing (Fin. äärioikeisto, äärioikeistolainen, äärioikeiston, äärioikeis, äärioikeistolaista, äärioikeistollisten)	384
Government (Fin. hallituksen, hallitus)	350
Municipal election, election (Fin. kuntavaaliehdokkaat, kuntavaalien, kuntavaalit, kuntavaalit2021, vaalit, vaaleihin, vaaleissa, vaaleja, vaalien)	297
Left-wing (Fin. vasemmisto, vasemmistolainen, vasemmalla, vasemmistolaisuus, vasemmiston, vasemmistopopulismista)	259
Communism (Fin. kommunismi, kommunisti, kommunistien, kommunistinen, kommunistisen, kommunistista, kommunistit)	134
Politicians	
The prime minister (Fin. marin, marinin, marinia, marinista, marinsanna, sanna, pääministeri, pääministeriltä)	898
Minister of the Interior (Fin. mariaohisalo, ohisalo, ohisalon, ohisalosta, sisäministe, sisäministeri)	705

Table 7. Cont.

The Content of Three Main Clusters in English	Quantity of Manifestations
Foreign countries	2769
Russia (Fin. venäjä, venäjän, venäjällä, venäjä'n, venäläinen, venäläispropagandistien, russia, russian, russians)	867
Trump (Fin. trump, trumpin, trumpia, trumpia)	409
EU (Fin. eu, eu:n, eu'n, euroopan)	326
USA (Fin. usa, usa:n, usan, usa:ssa, usassa, usavaalit, yhdysvallat, yhdysvalloissa, yhdysvaltain)	237
China (Fin. kiina, kiinaa, kiinan, kiinassa, ürümqi)	225
Sweden (Fin. ruotsin, sek)	104
Putin (Fin. putin, putinin)	98
Soviet Union (Fin. neuvostoliitto, neuvostoliiton)	63

In Table 7, three of the main clusters are examined in more detail. The results comprise the most manifested words in the clusters “Politics,” “Politicians,” and “Foreign countries”. The cluster of “Politics” included general terms of politics, such as politics (“politiikka” in Finnish) and government. Furthermore, there were words referring to two parties (Finns Party and The Greens), and the political spectrums of right-wing and left-wing, and even the political ideology of communism. The elections were also presented among the most manifested words.

The two most manifested in the Politicians clusters were the Prime Minister and Minister of the Interior with various terms. Their manifestations were 75% of the total number in the Politician cluster. In the foreign countries cluster, Russia was the most manifested. However, EU, USA, China, Sweden, and even the Soviet Union were also manifested in the tweets. Presidents Putin and Trump were manifested highly in Finnish tweets regarding misinformation.

5. Discussion

The rapid spread of false information on Twitter has caused concern. Along with the social media applications, the various methods to study and analyze social media data, such as Twitter data, are evolving, too. Therefore, the functional analysis methods enable the investigation of discussions on misinformation and its related concepts, such as disinformation, fake news, and propaganda themes linked with those concepts in Twitter. By answering the research question: “What are the functional methods to analyze Finnish Twitter data to discover out discussion themes that are linked with misinformation and concepts related to it?”, our article makes a two-fold contribution. First, it contributes to the applicability of three analysis methods on Finnish Twitter data. Secondly, it contributes to the discovered main themes that are connected to the misinformation-related concepts in Finnish Twitter discussions.

Regarding the contribution of the applicability of three analysis methods to Twitter data, the article focused on describing and comparing methods to detect and analyze the text content of tweets in Twitter conversations in Finnish that includes the concepts of misinformation, disinformation, fake news, hoax, and propaganda, and the themes linked to them. Limiting the investigation to the text content of tweets was done to uncover the inherent challenges of text mining Finnish tweet contents and to understand what kind of insights could be drawn from only the text. The grammatical characteristics of the Finnish language, particularly regarding the word formation with endings to word stems, as well as word formation via derivation and compounding, challenges conventional data analysis methods as well as machine learning.

Our results were obtained using three methods: (1) word cloud analysis, (2) topic model of tweets, and (3) word count analysis and clustering. Of these methods, the result of word cloud clustering is the most non-specific. Word cloud clustering does highlight the most utilized words visually and thus can easily present the relations in the manifestations

of various words. However, it regards variant forms of words as separate, therefore it does not lemmatize the same stem variants. Thus, for Finnish text, word cloud clustering is not the most ideal for producing 100% correct results. Rather, it is a good method to look for word stems for further analysis. Compared to word cloud analysis, the results of the second analysis method—the topic model of tweets—gives a more distinct view of the most utilized words in Twitter data. However, the topic model does not consider the stem variants or lemmatize the same stem variants. Therefore, the accuracy of the results with a language like Finnish is not precise. As with word cloud clustering, topic modeling is feasible for an overview of the themes of Twitter discussion and detecting the word stems of the most used words. The third method—word count analysis followed by clustering in a spreadsheet—does produce very detailed results for word count, followed by clustering the words to themes. Yet, manual stem variant detection and lemmatizing is a labour-intensive phase in the research. However, it does bring value to the research if accuracy and inclusiveness of as many words as possible are required. Another possible method could be that the main word stems are discovered with word cloud analysis, or better yet, with the topic model of tweets, and then the documents are coded automatically in Atlas.ti with the word stems. However, with the Finnish language, there is still the risk that irregularly declinable words or conjugated verbs will not be included in the code list. In addition, natural language processing in Finnish is very limited. Finnish lemmatization could be carried out with Python with a very few libraries, such as Voikko and FINNPOS, yet their documentation is almost non-existent [43]. For further research, the above-mentioned coding in Atlas.ti could be suggested. Another avenue would be to consider possible machine learning methods to automate the labor-intensive phase of thematic analysis. However, compared to the word cloud analysis and the topic model of tweets, the combination of word count analysis and clustering does illustrate with more detail which political themes (for example) Twitter users combine with misinformation-related concepts. Nevertheless, word count analysis and clustering do not reveal which themes are linked with each other concerning misinformation, for example, which themes are linked with the Prime Minister. This would require coding the data with word stems and using manually coded data in combination with topic modeling.

Regarding the contribution to the discovered main themes connected to the misinformation-related concepts in Finnish Twitter discussions, our results revealed that "propaganda" is the most utilized misinformation-related word. Conspicuously, this result was detectable with all three analysis methods. In addition, fake news (Fin. *valeuutinen*) and disinformation (Fin. *disinformaatio*) were particularly represented misinformation-related words. Besides the keywords, the most used words in the Twitter data were generally neutral, i.e., their information value is not high without knowing the context. These are media-related words, such as news. Yet, they emphasize the communication style typical for microblogging on Twitter: tweeting, re-tweeting, and commenting on news and stories. Furthermore, other general words, such as politics, country, and world, are neutral words lacking context. Yet, they portray the landscape of the discussion content on Twitter. Furthermore, our results, particularly with the cluster analysis, revealed that politics and politicians are the main topics of discussion with regards to misinformation, disinformation, fake news, hoax, and propaganda in Finnish Twitter. Thus, this confirms the prior studies on fake news [29], as well as studies on false news spread on English Twitter [17].

In addition to the textual correlation, the study of disinformation has an emotional correlation. However, how and why disinformation flows—for example, regarding political polarization and immigration, particularly on social media such as Twitter—are questions that research regarding methods to analyze the text content of tweets in Twitter does not give answers. Therefore, we suggest further research of a descriptive or explanatory type with qualitative analysis methods on Twitter interactions and Twitter users. Patterns of how disinformation flows from one user to another user can be revealed, e.g., by analysis of information cascades on Twitter [44]. However, the rationale for why people

disseminate misinformation inadvertently or disinformation intentionally on social media is discoverable from human motivation.

A limitation of this study is that it only investigated the text content of tweets. Moreover, the data was limited to Twitter data in the Finnish language. Furthermore, the study was limited to the functionality of three methods for analyzing Twitter data. Naturally, there could also be other types of analysis methods for different types of Twitter data besides the text content of tweets that could be functional especially for tweets in the Finnish language or other languages.

6. Conclusions

With the three methods we utilized in the analysis of Finnish Twitter data—(1) word cloud analysis, (2) topic model of tweets, and (3) word count analysis and clustering—we detected topics that are related to misinformation-related concepts: disinformation, fake news, and propaganda. We detected the advantages and disadvantages of each method and different levels of the result accuracy of each method. Furthermore, we noticed that each method has critical limitations, especially all the automated analysis methods for Finnish language processing; yet when combined, they bring value to the analysis. However, further research is needed on the applicability of analysis methods for Twitter data in Finnish due to the particularities in word formation.

Politics, both internal and external, are present in the Twitter discussions in connection with misinformation concepts. Thus, our results support the previous findings of studies on fake news [29], as well as studies on the spreading of false news on English Twitter [17].

Author Contributions: Conceptualization, J.J. and A.H.S.; methodology, A.H.S. and A.P.; software, A.P.; validation, J.J. and A.H.S.; formal analysis, A.P.; investigation, A.H.S. and J.J.; resources, J.J.; data curation, A.P. and T.H.; writing—original draft preparation, J.J., A.H.S., and A.P.; writing—review and editing, A.H.S. and J.J.; visualization, A.P. and T.H.; supervision, J.J.; project administration, J.J.; funding acquisition, J.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study is available at: <https://github.com/hank-uas/TwitterAnalysis> (accessed on 10 June 2021). Owing to Twitter's policy we are restricted to sharing tweet-ids and users can rehydrate this dataset using hydrator: <https://github.com/DocNow/hydrator> (accessed on 10 June 2021).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Choudrie, J.; Banerjee, S.; Kotecha, K.; Walambe, R.; Karende, H.; Ameta, J. Machine learning techniques and older adults processing of online information and misinformation: A covid 19 study. *Comput. Human Behav.* **2021**, *119*, 106716. [CrossRef]
2. EU. Final Report of the High Level Expert Group on Fake News and Online Disinformation. 2018. Available online: <https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation> (accessed on 10 June 2021).
3. Lazer, D.M.J.; Baum, M.A.; Benkler, Y.; Berinsky, A.J.; Greenhill, K.M.; Menczer, F.; Metzger, M.J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. The science of fake news. *Science* **2018**, *359*, 1094–1096. [CrossRef]
4. Shu, K.; Bhattacharjee, A.; Alatawi, F.; Nazer, T.; Ding, K.; Karami, M.; Liu, H. Combating Disinformation in A Social Media Age. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, 1–39. [CrossRef]
5. Spradling, M.; Straub, J.; Strong, J. Protection from 'Fake News': The Need for Descriptive Factual Labeling for Online Content. *Futur. Internet* **2021**, *13*, 142. [CrossRef]
6. Helmstetter, S.; Paulheim, H. Collecting a Large Scale Dataset for Classifying Fake News Tweets Using Weak Supervision. *Futur. Internet* **2021**, *13*, 114. [CrossRef]
7. Carchiolo, V.; Longheu, A.; Malgeri, M.; Mangioni, G.; Previti, M. Mutual Influence of Users Credibility and News Spreading in Online Social Networks. *Future Internet* **2021**, *13*, 107. [CrossRef]
8. Haselton, M.G.; Nettle, D. The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personal. Soc. Psychol. Rev.* **2006**, *10*, 47–66. [CrossRef] [PubMed]

9. Binder, J.R.; Desai, R.H.; Graves, W.W.; Conant, L. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* **2009**, *19*, 2767–2796. [CrossRef] [PubMed]
10. Fedorenko, E.; Scott, T.L.; Brunner, P.; Coon, W.G.; Pritchett, B.; Schalk, G.; Kanwisher, N. Neural correlate of the construction of sentence meaning. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E6256–E6262. [CrossRef]
11. Freelon, D.; Wells, C. Disinformation as political communication. *Polit. Commun.* **2020**, *37*, 145–156. [CrossRef]
12. Bradshaw, S.; Howard, P.N.; Kollanyi, B.; Neudert, L.M. Sourcing and automation of political news and information over social media in the United States, 2016–2018. *Polit. Commun.* **2020**, *37*, 173–193. [CrossRef]
13. Bello-Organ, G.; Jung, J.J.; Camacho, D. Social big data: Recent achievements and new challenges. *Inf. Fusion* **2016**, *28*, 45–59. [CrossRef]
14. Lytras, M.D.; Visvizi, A.; Jussila, J. Social media mining for smart cities and smart villages research. *Soft Comput.* **2020**, *24*, 10983–10987. [CrossRef]
15. Vatrupu, R.; Mukkamala, R.R.; Hussain, A.; Flesch, B. Social set analysis: A set theoretical approach to big data analytics. *IEEE Access* **2016**, *4*, 2542–2571. [CrossRef]
16. Li, J.; Su, M.H. Real Talk about Fake News: Identity Language and Disconnected Networks of the US Public’s “Fake News” Discourse on Twitter. *Soc. Media Soc.* **2020**, *6*. [CrossRef]
17. Vosoughi, S.; Roy, D.; Aral, S. The spread of true and false news online. *Soc. Sci.* **2018**, *359*, 1146–1151. [CrossRef] [PubMed]
18. Varol, O.; Ferrara, E.; Davis, C.A.; Menczer, F.; Flammini, A. Online Human-Bot Interactions: Detection, Estimation, and Characterization. In Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017), Montréal, QC, Canada, 15–18 May 2017; pp. 280–289.
19. EU Institutions Data Flash Eurobarometer 464: Fake News and Disinformation Online. Available online: https://data.europa.eu/data/datasets/s2183_464_eng?locale=en (accessed on 10 June 2021).
20. Zakharchenko, A.; Peráček, T.; Fedushko, S.; Syerov, Y.; Trach, O. When Fact-Checking and ‘BBC Standards’ Are Helpless: ‘Fake Newsworthy Event’ Manipulation and the Reaction of the ‘High-Quality Media’ on It. *Sustainability* **2021**, *13*, 573. [CrossRef]
21. Dann, S. Twitter data acquisition and analysis: Methodology and best practice. In *Maximizing Commerce and Marketing Strategies through Micro-Blogging*; IGI Global: Hershey, PA, USA, 2015; pp. 280–296.
22. UN. UN Tackles ‘Infodemic’ of Misinformation and Cybercrime in COVID-19 Crisis | United Nations. Available online: <https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-cybercrime-covid-19> (accessed on 10 June 2021).
23. Zeng, J.; Chan, C.H. A cross-national diagnosis of infodemics: Comparing the topical and temporal features of misinformation around COVID-19 in China, India, the US, Germany and France. *Online Inf. Rev.* **2021**. [CrossRef]
24. Fetzer, J.H. Disinformation: The use of false information. *Minds Mach.* **2004**, *14*, 231–240. [CrossRef]
25. Pal, A.; Banerjee, S. Handbook of Research on Deception, Fake News, and Misinformation Online. In *Advances in Media, Entertainment, and the Arts*; Chilwa, I.E., Samoilenko, S.A., Eds.; IGI Global: Hershey, PA, USA, 2019; ISBN 9781522585350.
26. Fetzer, J.H. Information: Does it have to Be True? *Minds Mach.* **2004**, *14*, 223–229. [CrossRef]
27. Bastick, Z. Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Comput. Human Behav.* **2021**, *116*, 106633. [CrossRef]
28. Tandoc, E.C.; Lim, Z.W.; Ling, R. Defining “Fake News”: A typology of scholarly definitions. *Digit. J.* **2018**, *6*, 137–153. [CrossRef]
29. Tandoc, E.C., Jr.; Thomas, R.J.; Bishop, L. What Is (Fake) News? Analyzing News Values (and More) in Fake Stories. *Media Commun.* **2021**, *9*, 110–119. [CrossRef]
30. UK Parliament Disinformation and ‘fake news’: Interim Report. Available online: <https://publications.parliament.uk/pa/cm201719/cmselect/cmcomeds/363/363.pdf> (accessed on 28 May 2021).
31. Visvizi, A.; Jussila, J.; Lytras, M.D.; Ijäs, M. Tweeting and mining OECD-related microcontent in the post-truth era: A cloud-based app. *Comput. Human Behav.* **2020**, *107*, 105958. [CrossRef]
32. Pan, W.; Fang, J. An Examination of Factors Contributing to the Acceptance of Online Health Misinformation. *Front. Psychol.* **2020**, *12*, 524. [CrossRef]
33. Van Sant, K.; Fredheim, R.; Bergmanis-Korats, G. Abuse of Power: Coordinated Online Harassment of Finnish Government Ministers. Riga: NATO Strategic Communications Centre of Excellence. Available online: https://stratcomcoe.org/pdfs/?file=/cuploads/pfiles/abuse_of_power_online_harassment_of_fin_ministers_16-03-2021.pdf?zoom=page-fit (accessed on 10 June 2021).
34. Mejova, Y.; Weber, I.; Macy, M.W. *Twitter: A Digital Socioscope*; Cambridge University Press: New York, NY, USA, 2015.
35. Friese, S. *Qualitative Data Analysis with ATLAS.ti*; SAGE: Los Angeles, SC, USA, 2019.
36. Karlsson, F. *Finnish: An Essential Grammar*; Taylor & Francis e-Library: Abingdon, UK, 2002; ISBN 0-203-18753-9.
37. Bird, S.; Klein, E.; Loper, E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*; O’Reilly Media, Inc.: Newton, MA, USA, 2009.
38. Korenius, T.; Laurikkala, J.; Järvelin, K.; Juhola, M. Stemming and lemmatization in the clustering of finnish text documents. In Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management, Washington, DC, USA, 8–13 November 2004; pp. 625–633.
39. Roesslein, J. Tweepy. Available online: <https://docs.tweepy.org/en/stable/> (accessed on 3 May 2021).
40. Partanen, A. TweetCollector. Available online: <https://github.com/hamk-uas/TweetCollector> (accessed on 11 June 2021).

-
41. Twitter Twitter API v2: Early Access. Available online: <https://developer.twitter.com/en/docs/twitter-api/early-access> (accessed on 11 June 2021).
 42. Rosen, A. Tweeting Made Easier. Available online: https://blog.twitter.com/en_us/topics/product/2017/tweetingmadeeasier.html (accessed on 7 May 2021).
 43. Antupis Finnish Lemmatization with Python. Available online: <https://antupis.github.io/lemmatization/finnish/2019/06/12/Lemmatizing-finnish-text.html> (accessed on 28 May 2021).
 44. Bakshy, E.; Hofman, J.M.; Mason, W.A.; Watts, D.J. Everyone's an influencer: Quantifying influence on twitter. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, Hong Kong, China, 9–12 February 2011; pp. 65–74.