

Silje Susanne Alvestad

DOI: <https://doi.org/10.33099/2304-2745/2023-3-79/84-88>

University of Oslo

On the way to detecting the language of disinformation: lessons learned from the “Fakespeak” project

Introduction. The Fakespeak project is an interdisciplinary research project involving linguists from the University of Oslo and computer scientists from SINTEF Digital in Oslo, Norway. Funded by the Norwegian Research Council as part of the Public Safety and Risks program, the project started in 2020 and will continue until the end of 2025. The purpose of the research project is twofold:

firstly, work continues on identifying the language and style of fake news “Fakespeak” (an allusion to the concepts of “Newspeak” and “Doublethink” from Orwell's novel "1984") in Russian, Norwegian and English;

secondly, it investigates whether adding linguistic features of fake news to existing fake news detection tools can make such tools more efficient.

The project also involves Faktisk.no, the first and so far only fact-checking service in Norway, the Norwegian Broadcasting Company (NRK) and the Norwegian News Agency (NTB), which is “Norway's largest provider of content in the form of text, images, video and graphics for Norwegian mass media”. One of the project's goals is to help stakeholders identify potentially harmful fake news more efficiently, accurately, and in a timely manner than it is currently possible. For this purpose, seminars were organized for the knowledge sharing between representatives of external cooperation partners.

The article summarizes the results of the “Fakespeak” project (there are two years left until its completion). Attention is focused on the prerequisites of the project, challenges during its implementation, as well as on possible ways of further development of the project.

Political background. Fake news that are clearly defined at the beginning of the project as information intended to mislead and at the same time the author knows that this information is false [1], is not a new phenomenon. However, the rapid development of social networks allows news from sources of various reputations to spread without filtering at lightning speed and be read by millions of people in a very short time. Open democracies are vulnerable, and fake news and other forms of disinformation can seriously damage them. For example, after examining the vast amount of available evidence, Jamison [2]

concluded that Russian interference most likely swayed the results of the 2016 US presidential election in favor of Donald Trump. The subtitle of her monograph is telling: “*How Russian hackers and trolls helped elect the president.*” Former CIA and NSA director Michael Hayden called the Russian attacks “the most successful covert influence operation in history.” Fake news were part of this attack. It is also worrying that Jamison writes that the US is ill-prepared to deal with such challenges. Moreover, Vladyslav Surkov, the “Kremlin Goebbels”, boasted that Russia was playing with the minds of the West, and already in 2014, Petro Pomerantsev published his book entitled “*Nothing is False and Everything is Possible. The surreal heart of the new Russia*”. In this book, Pomerantsev, in particular, illustrates one of the possible consequences of large-scale and long-term disinformation operations: a kind of end-state in which people are so disillusioned that they consider everything fake, no longer is care about what true and what is not. As the researchers note, almost at the time of writing this article, such a scenario is a serious threat to democracy, national and international security and needs to be mitigated.

Sometimes, the press media and mass media are referred to as the fourth estate, alluding to the separation of powers in government and reflecting their important role in society. However, in 2016, an expert panel convened by the BBC declared “the breakdown of trusted sources of information to be one of the most pressing societal problems of the 21st century”, and also in 2016, the Oxford Dictionary declared “post-truth” the word of the year [3]. Thus, truth and trust – the central values of open democracies – are under threat. It is against this political background that the Fakespeak project was developed, and in early 2019 its idea was that improved fact-checking techniques could help the public be critical of the information they are exposed to and restore trust in the mainstream media.

State of the science on the language of fake news in 2019. There is a growing body of research on the phenomenon of “fake news” with research being conducted in several fields. For example, within media science, important

questions concern the sources, content, and target audiences of fake news. In psychology, the key questions are why readers (listeners) tend to believe fake news, why they share stories that evoke emotion and excitement [4], and why some audiences are immune to the truth in some cases. The lion's share of fake news research was being conducted and continues to be conducted by computer scientists, with the most important research question being how fake news can be detected automatically.

Some research conducted by computer scientists combines computer science methods with some knowledge of linguistics, as for example outlined in [5, 6]. However, linguistics plays only a minor role in these studies, and the projects themselves almost never include linguist participants. Obviously, computer scientists are very useful for timely detection of fake news, but linguistics will help advance this work: As noted in a report by the Reuters Institute, an automated fact-checker in 2018 could only identify simple declarative statements such as “Donald Trump President of the United States”. Automated fact-checking has not yet identified:

implied statements that may be false even if the direct statement is true;

statements embedded in complex sentences in which case the embedded statement may be false even if the complex sentence is true;

cross-references such as anaphora.

Humans readily recognize both implicit and embedded statements and can readily recognize anaphora. Obviously, language is much more than simple declarative sentences, and therefore the project requires qualified linguists on the team.

Studies of fake news, conducted within media and computer sciences in particular, tend to be content-based and focus on what is true and what is false. One of the problems with this dichotomy is that the news is often neither completely true nor completely false. The political fact-checking service "PolitiFact", for example, operates with the following degrees of credibility of statements [3]:

true; almost true (mostly true); half true; barely true; false; “pants on fire”.

Thus, fake news is not just a question of what is false and what is true, and not about the reliability of their sources: fake news sources sometimes report the story correctly, and serious and authoritative media sometimes report it incorrectly [7]. In the course of the project, it was established that fake news is determined rather by the author's intention to deceive. And the author's intentions are reflected in the language he uses. In

particular, based on the analysis of large samples of natural language, corpus linguists have demonstrated that there are systematic variations in the structure of language depending on the communicative purpose of the author (op. cit.). When telling stories, more past tense verbs and third person pronouns are used. On the other hand, when explaining something, more nouns and prepositions are usually used. When communicating, more questions and exclamations are used. In other words, “the grammar of the text reflects its purpose”. Thus, the language of fake news, namely its structure, rather than its content, may be the key to its detection.

Based on this insight, “Grieve & Woodfield” in 2023 conducted a study of news by Jason Blair [7, 8], which produced very intriguing and promising results. Briefly, the researchers compared and analyzed datasets of fake and genuine articles written by the same author. In particular, in the early 2000s, Jason Blair, a former NYT reporter, was found to have fabricated news from time to time. The NYT began an investigation and, in particular, flagged fabricated texts, resulting in two sets of data:

true news; fabricated stories.

“Grieve & Woodfield” submitted these two data sets for verification to “Register Analysis”, suggesting that given the different communicative purposes of the texts (deceive or inform) in these two sets, true and fabricated texts should be grammatically distinct [7]. They compared the relative frequencies of certain grammatical features in the two sets of texts, and their overall conclusion is that Blair writes in a more formal style in his true stories, while he is more “engaged” in the fictional stories.

The hallmarks of Blair's true stories match those of information-dense writing, while the hallmarks of his false stories resemble those of interactive discourse. Thus, based on Blair's authorship, signs of real news include longer average word length and nominalization (use of nouns in *-tion, -ment, -ness, -ity*), while signs of fake news include increased use of 1st and 3rd person pronouns, as well as a wider use of the present tense and emphatic words such as *really* and *most* (op. cit.: 32).

Against the background of the promising results of the study of Jason Blair's publications, an attempt was made to assemble (compile) data corpora similar to the dataset on Jason Blair's works. However, it quickly became clear that there are very few such corpora even in English and their organization is cumbersome and time-consuming as well as for the “smaller” languages like Russian and Norwegian. Furthermore,

acknowledging the intriguing findings of the Jason Blair study and the fact that, by studying the same journalist writing for the same publication under the same editor, Grieve and Woodfield were able to control for several potentially confounding features such as genre variation, colleagues have raised two types of criticisms of this study. *Firstly*, Jason Blair is an individual journalist. Can the research on Blair's publications be generalized to all other journalists who fabricate news articles? *Secondly*, Blair's motivation for fabricating news articles was financial. In particular, Blair claims in his autobiography that he had a problem with alcohol and needed money to finance his abuse. So, he fabricated the news to increase his profits. Can the results of the study of Blair's publications be generalized to the work of other journalists who could also write both fake and true articles, but with completely different motives for lying? These are timely and adequate questions. Based on research in the Fakespeak project, we can say that the answer to both questions is most likely no. Explanations of this conclusion are given in the next section.

Some preliminary conclusions. Despite the fact that “Jayson Blair” type corpora are few, it was possible to create several other small English corpora of the same type [9]. Researchers at the Fakespeak project conducted a metaphor study based on these single-author datasets of the English language and tentatively found the following: *First*, Blair uses metaphors sparingly, and *second*, when he does use metaphors, they are quite conventional. However, journalists who lie for ideological reasons seem to be more likely to use sports and war metaphors [10]. This means that, contrary to the full first name of our project – “Fakespeak” – *the language of fake news* – there is not one language of fake news, but several. There are many ways in which journalists can lie, and there are many ways in which journalists lie. Therefore, it is not necessary to generalize the example of Jason Blair to other journalists who may have completely different motives for lying and fabricating news articles.

Since there are not many individual author corpora, it was necessary to start the project in two dimensions, firstly, from the point of view of data sets for research and, in parallel with this, the definition of “fake news”. In particular, based on links from fact-checking services such as “PolitiFact” (for English in the USA), “Faktisk” (for Norwegian) and “provereno.media” (for Russian), a collection of text corpora was started consisting of several authors. As a result, texts

written by several different authors representing different genres, such as news articles and blog posts, have been collected in the same data set. However, a certain level of objectivity and quality can be guaranteed, since all articles are checked by professional fact-checkers [9, 11]. In particular, for these data sets, one cannot be sure of the author's intention to mislead. (Recall that it was a defining feature of fake news according to the original clear definition). Therefore, these multi-author datasets are most likely to contain instances of misinformation that may be unintentional, in addition to misinformation that is believed to be created with intent to mislead.

It was made a specific preliminary observation. In particular, preliminary observations suggest that adverbs and other constructions (e.g., “*that-clauses*”) that express epistemic certainty are overrepresented in fake news, at least in English and Russian. Regarding the Norwegian language, there is still too little data available to say anything useful [12]. Examples of such constructions are adverbs such as *of course, evidently, obviously, clearly, actually, in fact, definitely*, etc., as well as sentences with *that-clauses* such as *I am absolutely certain that*. Thus, one gets the impression that the less confident the author is in the truth of the statement, the more likely they are to use expressions that convey confidence in it.

Prospects for the future. The Language Council of Norway announced “falske nyheter” – “fake news” – as the word of the year for 2017. The idea of the “Fakespeak” project arose at the end of 2018 – the beginning of 2019. At that time, only works [13] about the language of fake news were known, and later work [14] appeared. Since then, interest in fake news and similar phenomena (such as propaganda, conspiracy theories, pseudoscience, etc.) in linguistics has almost exploded. One example of this is the fact that the “Linguistics Vanguard” special collection on the language of fake news has received almost 30 articles covering languages from four continents and representing a wide range of linguistic approaches. Such huge interest reflects the fact that since the launch of the project in 2020, the threat posed by fake news and other types of disinformation unfortunately has not been decreased rather than opposite. Especially with the COVID-19 pandemic, Russia's full-scale invasion of Ukraine, and the recent war between Israel and Hamas, this issue has become particularly prominent.

With the advent of large language models (LLMs), the problem of fake news and other types of disinformation has become even more

urgent. Some artificial intelligence experts estimate that by 2026, almost 90% of the content on the Internet will be generated synthetically. Creating malicious content will become increasingly cheaper and easier. Something is already known about the language of fake news and disinformation – when fake news is written by people. But it is necessary to be able to mention something about the language created by artificial intelligence (artificial language), that is the language of large language models in general, and the language of disinformation created by artificial intelligence in particular. However, it should be noted that there are questions to be answered by future linguistic research: “how can you create linguistic knowledge that relates to”:

- several artificial languages, not just one;
- several artificial languages for a long time, and not only until the next update;
- artificial language, about which nothing is known, since they can be created and prepared by enemy (state) entities.

REFERENCES

1. Horne, B. D. and S. Adalı. 2017. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. Available at <https://arxiv.org/abs/1703.09398>, accessed December 1, 2023.
2. Jamieson, K. H. 2018. *Cyberwar. How Russian hackers and trolls helped elect a president. What we don't, can't, and do know*. Oxford: Oxford University Press.
3. Choy, M. and M. Chong. 2018. Seeing through misinformation: A framework for identifying fake online news. Available at <https://arxiv.org/pdf/1804.03508.pdf>, accessed December 21, 2023.
4. Rimé, B. 2009. Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion review* 1(1): 60–85.
5. Conroy, N. J., V. L. Rubin, and Y. Chen. 2015. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4.
6. Pérez-Rosas, V., B. Kleinberg, A. Lefevre, and R. Mihalcea. 2018. Automatic detection of fake news. *Proceedings of the 27th International Conference on Computational Linguistics*, 3391–3401. Santa Fe, New Mexico, USA, August 20-26, 2018. Available at <http://aclweb.org/anthology/C18-1287>, accessed December 21, 2023.
7. Grieve, J. 2019. Linguistics approaches to the detection and obfuscation of disinformation. A multi- and inter-disciplinary approach to disinformation research and policy. Presentation held at St. Anthony's College Oxford, March 11, 2019.
8. Grieve, J. and Woodfield, H. 2023. *The Language of Fake News*. Cambridge Elements in Forensic Linguistics. Cambridge: Cambridge University Press.
9. Pöldvere, N., Kibisova, E. and Alvestad, S. S. 2023. Investigating the language of fake news across cultures. In Maci, S. M., Demata, M., McGlashan, M. and Seargeant, S. (eds.) *The Routledge Handbook of Discourse and Disinformation*, p. 153-165. Routledge.
10. Trnavac, R. and Pöldvere, N. In press. Investigating Appraisal and the language of evaluation in fake news corpora. *Corpus Pragmatics*.
11. Pöldvere, N., Uddin, Z. and Thomas, A. 2023. The PolitiFact-Oslo Corpus: A new dataset for fake news analysis and detection. *Information*, 14, article 627. <https://doi.org/10.3390/info14120627>, accessed December 21, 2023.
12. Pöldvere, N., Kibisova, E., Alvestad, S. S. and Trnavac, R. 2023. Fake news around the world: A corpus-based analysis of stance in fake news in English, Norwegian and Russian. Presentation held at BAAL2023, The language of fake news symposium, University of York, August 24, 2023.
13. Grieve, J. 2018. The language of fake news. Text available at <https://www.birmingham.ac.uk/news/thebirminghambrief/items/2018/09/the-language-of-fake-news.aspx>, accessed December 21, 2023.
14. Asr, F. T. and Taboada, M. 2019. Big Data and quality data for fake news and misinformation detection. *Big Data & Society* 6(1). <https://doi.org/10.1177/2053951719843310>.

The article has been submitted to the editorial board 17.11.2023

На шляху до виявлення мови дезінформації: досвід проєкту “Fakespeak”

Анотація

Проєкт “Fakespeak” – це міждисциплінарний дослідницький проєкт, у якому беруть участь лінгвісти з Університету Осло та комп'ютерні науковці з SINTEF Digital в Осло, Норвегія. Фінансований Норвезькою дослідницькою радою в рамках програми “Суспільна безпека та ризику” проєкт розпочався у 2020 році і триватиме до кінця 2025 року. Мета дослідницького проєкту є подвійною:

по-перше, триває робота над виявленням мови та стилю фейкових новин “Fakespeak” (алюзія на поняття “Newspeak” і “Doublethink” з роману Оруела “1984” – російською, норвезькою та англійською мовами);

по-друге, досліджується питання, чи може додавання лінгвістичних особливостей фейкових новин до існуючих інструментів виявлення фейкових новин зробити ці інструменти більш ефективними.

У проєкті також беруть участь Faktisk.no, перший і поки єдиний сервіс фактчекінгу в Норвегії, Норвезька телерадіокомпанія (NRK) і Норвезьке агентство новин (NTB), яке є “найбільшим у Норвегії постачальником контенту у вигляді тексту, зображень, відео та графіки для норвезьких ЗМІ”. Одна з цілей проєкту – допомогти зацікавленим сторонам виявляти потенційно шкідливі фейкові новини ефективніше, точніше і своєчасно, ніж це можливо зараз. З цією метою організовано семінари для обміну знаннями з представниками зовнішніх партнерів по співпраці.

У статті підведено підсумки проєкту “Fakespeak” (до його завершення залишилися два роки). Увагу зосереджено на передумовах виникнення проєкту, викликах під час його виконання, а також на можливих шляхах подальшого розвитку проєкту.

Існують питання, на які мають відповіді майбутні лінгвістичні дослідження: “як можна створити лінгвістичні знання, які стосуються”:

кількох штучних мов (ШМ), а не лише однієї;

кількох ШМ протягом тривалого часу, а не лише до наступного оновлення;

ШМ, про які нічого відомо, оскільки вони можуть бути створені та підготовлені ворожими (державними) суб'єктами.

Ключові слова: фейкові новини; лінгвістичні дослідження; штучна мова; комп'ютерні науковці.