

# Linguistic cues could be key to exposing fake news

September 27 2022, by Mari Lilleslåtten

---



Credit: CC0 Public Domain

After the revelations about the 2016 U.S. presidential election being influenced by Russian-generated "fake news," many people became more critical towards news on social media. "Fake news" was

subsequently coined by several dictionaries and language organizations, such as the Language Council of Norway, as being the 2017 word of the year. In Norway many of us learned that if something appears to be too good to be true, then it often is.

But what about the language itself—can it provide an indication of how true the text you are reading is?

At the University of Oslo, linguists are now working with computer scientists and artificial intelligence researchers at the independent research organization SINTEF in order to expose the language of fake [news](#), what they call Fakespeak.

"We are investigating to see whether or not there are linguistic differences between real and false news texts in Norwegian, English and Russian. Our goal is to improve current fact-checking tools," says Silje Susanne Alvestad.

She is the head of [the Fakespeak project](#) and recognizes that linguistics, which is her field of expertise, can provide important societal benefits by combatting fake news.

"For several years, research in media studies and computer science has been conducted on various aspects of fake news, for example, the way in which it is spread. But in linguistics, there have been gaps with regards to this phenomenon," says Alvestad.

## **Informal style and verbs in the present tense can be important signs**

Admittedly, there are some linguists who have tackled fake news articles in the past. In 2003, the *New York Times* journalist Jayson Blair was

caught [fabricating a number of news articles](#). Jack Grieve and his colleagues at the University of Birmingham have gathered these false texts into what linguists call a corpus, comparing them with a selection of real news stories written by Blair.

"The researchers assumed that since Jayson Blair had different motives for writing these two types of articles—seeking to provide information in his genuine texts and intending to mislead people with his fake ones—the style and the linguistic features would also be different," says Alvestad.

And sure enough, the texts were different in style. "The untrue ones had an informal style, while the genuine texts were similar to other texts containing a high density of information."

The British researchers discovered several linguistic differences:

- Real texts: more frequent use of nouns and words that modify nouns. The words were longer on average.
- Fake texts: more frequent use of verbs, especially in the present tense. Also, more use of pronouns, adjectives and small words used for emphasizing the meaning (emphatic words).

Alvestad and her linguist colleagues, Nele Põldvere and Elizaveta Kibisova, are building on these findings as they are now investigating the linguistic characteristics of fake news in Norwegian, English and Russian.

## **How metaphors are used could be an important sign**

A metaphor is an expression taken from one domain and applied to another. For example, one can use a metaphor from war in the field of health when talking about how to "attack a virus."

The UiO researchers led by Nele Pöldvere have taken a closer look at Blair's use of metaphors.

"He uses fewer metaphors in his fake news articles than when he writes the truth. One possible explanation for this is that we most often use metaphors when we retell stories about something we have actually experienced ourselves," says Alvestad.

In addition, Blair uses linguistic elements that describe or try to promote positive emotions.

"Previous research has shown that when you deliberately want to mislead people, you usually try to elicit strong negative emotions. However, the opposite was true with Jayson Blair. When he writes false articles, he uses words, phrases and wording that create positive emotions."

Alvestad points out that this could be due to the topic: Several of Blair's texts were fake stories about heroic American soldiers during the Iraq war.

"Blair wanted to present the Iraq war in a positive light."

## **A challenge to find enough fake news in Norwegian**

When researchers compare true and fake texts written by the same person, as they are doing with Blair's texts, valuable data emerges. They safeguard against several potential sources of error, such as differences in personal writing style and differences in genre. At the same time, it can be difficult to generalize on the basis of findings sourced from on a single individual.

"Jack Grieve and his colleagues conducted several smaller studies similar to the Jayson Blair study and they concluded that people lie in different

ways," Alvestad points out.

One author often fails to produce enough text. While Jayson Blair's texts reach a total of 80 pages, machine-learning specialists prefer to work with collections of text that are much larger than that. The researchers have therefore chosen to combine sets of text written by one author with texts written by different authors, which they collect from fact-checking services.

Alvestad and her colleagues have made good progress in analyzing the language of fake news in English, while both Norwegian and Russian are presenting some methodological challenges.

"While English is the most commonly used language online and has been the subject of most research, it is difficult to find enough material in Norwegian. Norway features at the top of the list of studies about trust in the media, so this is hardly surprising."

Nevertheless, the researchers have some examples of fake news from individual authors who have also written real articles that they can make comparisons with, and they are collaborating with the Norwegian fact-checking service Faktisk.no in order to collect a larger set of texts.

"The latter takes a lot of time, because none of the fact-checking services we have been in contact with have archives they can share with us. We therefore have to find our way back to the original [text](#) which has often been amended or removed after the actual facts have been checked. We of course want to examine these articles as they were before their facts were checked," says Alvestad.

## **Difficult to verify sources in Russian fake news items**

It is well documented that falsehoods abound in the Russian media. Still,

it is challenging for Alvestad and her colleagues to find Russian texts that they can use as research material.

"For example, it would be interesting to investigate the impact of Russian information prior to the invasion of Ukraine," says Alvestad.

However, such a study presents a number of challenges.

"First of all, it became difficult early on for journalists in Russia to write something that deviated from the authorities' version of reality. Consequently, the texts look more like press releases than news articles and they often lack the authors' names. We want to include the authors' names and sources so that we can also find texts with which we can compare the misleading texts."

Furthermore, fact-checking services in Russia are somewhat different than they are in countries like Norway.

"In Russia it is forbidden to spread fake news on certain topics, but their definition of fake news is not quite the same as ours."

In order to find good material in Russian, the researchers are now looking at fact-checking services and media based outside Russia, such as the Ukrainian stopfake.org.

## **Better tools for uncovering fake news**

[The social media platform Facebook currently uses artificial intelligence to warn about potential disinformation.](#) If things go the way the researchers in the Fakespeak project are hoping, such tools could be improved.

"This is how we work: first, the linguists working on the project analyze

the texts. Then they hand over the results to the computer scientists, who incorporate the linguistic characteristics into the existing tools. The aim of this is to ensure that [fake news](#) can be detected faster than what is possible at the moment."

## **Will your results be relevant for other languages?**

"If the Fakespeak project discovers that there are common features in the three languages that we are investigating, this would be an interesting finding. However, these are just the Indo-European languages—there are many other language families. We will need many more studies to be able to say something about whether these traits are universal."

Alvestad reports that there is great interest in the Fakespeak research both inside and outside the world of academia and that she often receives enquiries from researchers who are interested in collaborating. She points to the value of researchers collaborating closely across both disciplines and institutions in a way that is generating new knowledge.

"We are actually an example of an interdisciplinary project that is making humanities research highly useful to society," she concludes.

Provided by University of Oslo

Citation: Linguistic cues could be key to exposing fake news (2022, September 27) retrieved 13 May 2024 from <https://phys.org/news/2022-09-linguistic-cues-key-exposing-fake.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--