

What to Annotate: Retrieving Lexical Markers of Conspiracy Discourse from an Italian-English Corpus of Telegram Data

Costanza Marini, Elisabetta Jezek

University of Pavia, Department of Humanities
Corso Strada Nuova, 65, 27100 Pavia, Italy
costanza.marini@unipv.it, elisabetta.jezek@unipv.it

Abstract

In this age of social media, Conspiracy Theories (CTs) have become an issue that can no longer be ignored. After providing an overview of CT literature and corpus studies, we describe the creation of a 40,000-token English-Italian bilingual corpus of conspiracy-oriented Telegram comments – the *Complotto* corpus – and the linguistic analysis we performed using the *Sketch Engine* online platform (Kilgarriff et al., 2010) on our annotated data to identify statistically relevant linguistic markers of CT discourse. Thanks to the platform’s *keywords* and key *terms* extraction functions, we were able to assess the statistical significance of the following lexical and semantic phenomena, both cross-linguistically and cross-CT, namely: (1) evidentiality and epistemic modality markers; (2) *debunking* vocabulary referring to another version of the truth lying behind the official one; (3) the conceptual metaphor INSTITUTIONS ARE ABUSERS. All these features qualify as markers of CT discourse and have the potential to be effectively used for future semantic annotation tasks to develop automatic systems for CT identification.

Keywords: conspiracy theories, corpus annotation, linguistic analysis

1. Introduction

Conspiracy Theories (CTs) are “allegation[s] of conspiracy that may or may not be true” (Douglas et al, 2019, p. 4), but whose proliferation in this age of social media poses a threat to society, in that they are contributing to distorting our perception of reality.

The goal of this study is to verify whether CTs are indeed characterised - as certain studies seem to suggest, albeit from a mostly monolingual perspective focusing on single CTs - by common discourse features (especially lexical and semantic), that may be exploited in future annotation tasks to develop automatic systems for CT detection.

In this contribution, after an interdisciplinary overview of CT literature, we describe the collection and annotation of our dataset and the creation of the *Complotto* corpus, an English-Italian bilingual corpus of conspiracy-oriented Telegram comments. The corpus counts 658 comments (317 Italian and 341 English), for a total of 40,045 tokens. For both English and Italian, comments were taken from three language-specific Telegram channels focusing on the same three CTs: Flat Earth, the vaccine conspiracy, and the climate change hoax.

The linguistic analysis we performed using the *Sketch Engine* online platform (Kilgarriff et al., 2010) on our annotated data allowed us to identify statistically relevant linguistic markers of CT discourse. In particular, thanks to the platform’s *keywords* and key *terms* extraction functions, we were able to determine the statistical significance of the following phenomena, both cross-linguistically and cross-CT, namely: (1) evidentiality and epistemic modality markers; (2) *debunking* vocabulary referring to another version of the truth lying behind the official one; (3) the

conceptual metaphor INSTITUTIONS ARE ABUSERS.

2. Conspiracy Theories

While the term *conspiracy* refers to an actual plot orchestrated at somebody’s expense, a CT refers to “an allegation of conspiracy that may or may not be true” (Douglas et al., 2019, p. 4). An example of conspiracy gone awry is the infamous Gunpowder Plot that took place on the 5th of November 1605, when a group of English Catholics, amongst which Guy Fawkes, attempted to blow up the House of Lords and to kill King James I. On the other hand, countless individuals and organisations have been accused of the assassination of U.S. President J.F. Kennedy in 1963, but all these CTs remain unproven.

Depending on the CT, the group of scheming individuals allegedly behind the plot might be identified with the same “epistemological authorities” (Uscinski, 2020) often tasked with making hard and unpopular decisions – i.e., scientists, governments, and other societal institutions.

As pointed out by Mancosu and Vassallo (2022, p. 2), CTs are often linked to a degeneration of public debate, especially in these times of social media, as in the case of the supposed electoral fraud to the detriment of Donald Trump which led to the assault on the U.S. Capitol Hill in January 2020.

By definition, CTs attempt to provide alternative explanations to events, therefore spreading the idea that “things are not as they seem, and that the truth behind certain events is hidden from view” (Demata et al., 2022, p.1) providing false evidence to support their claims (Danesi, 2023, p. 13).

As a consequence, CT narratives tend to juxtapose an Insider vs. an Outsider group (Bodner et al., 2020), where the Insider group feels threatened by the Outsider group and has to come up with strategies to counter these threats (Tangherlini, 2018).

In terms of the reasons why people believe in CTs, studies such as Douglas et al. (2017) have shown it satisfies a set of psychological motives, such as the desire for certainty, control, and security. Among the cognitive processes linked to a CT mentality, we find supernatural beliefs, a quasi-religious mentality, feelings of powerlessness and low control in the socio-political domain (Douglas, 2019, p. 7-9).

When it comes to analysing the CT discourse, we can expect all these characteristics, traits, and recurring motives to have their linguistic counterparts.

3. Related Works

Prior studies have focused on compiling CT corpora using both printed documents and social media content, mainly for English. For instance, Uscinski et al. (2011) compiled a corpus of conspiracy letters to the editor of *The New York Times* published from 1897 to 2010. On the other hand, Catenaccio (2022), who also carried out a corpus-driven analysis searching for the linguistic features of CT discourse, compiled a corpus of published books providing alternative accounts of the 9/11 events. Miani et al. (2021) released the *Language of Conspiracy* corpus (LOCO), an 88-million-word corpus of online texts covering a wide range of CTs collected automatically using a seeding approach. The LOCO corpus was used successfully by Mompelat et al. (2022) to design an annotation scheme that was used to develop CT vs. mainstream automatic document retrieval methods. Lastly, Russo et al. (2023) have created a dataset of 25.000 Italian posts extracted from five conspiracy-oriented Telegram channels, that were annotated to perform two computational classification shared tasks: a binary task aimed at determining whether a post is “conspiratorial” or not and a multi-class task aimed at recognizing the specific CT talked about in the post (Covid-19, QAnon, Flat Earth, Pro-Russia).

Previous works aimed at characterising the language of CTs have identified the following indicators: (1) constant reference to insider group vs. outsider group (Holur et al., 2022); (2) a non-standard use of epistemic stance and evidentiality markers (Catenaccio, 2022; Scharloth et al., 2019); (3) a creative *debunking* vocabulary referring to the fact that another version of the truth lies behind the official one (Ebling et al., 2013); (4) an instrumental use of conceptual metaphors to convey conspiratorial content (Danesi, 2023), since conceptual metaphors facilitate processing of non-literal meaning

allowing to view abstract concepts in terms of the properties of more concrete ones (Lakoff and Johnson, 1980).

Since no comprehensive cross-linguistic study has been conducted yet to define a shared annotation scheme for CT language, our work wishes to provide its contribution drawing from the afore-mentioned literature, as well as from existing ISO annotation standards that have not been applied to CT discourse yet.

4. Corpus Design and Annotation

In light of the relevance of social media in the spread of CTs and inspired by Russo et al. (2023), for our study, we decided to focus on Telegram data. If Twitter comments are indeed often short and very contextualised within a thread (Mompelat et al., 2022, p. 12), Telegram comments are usually quite long and exhaustive since they are not posted on a public wall but on the community’s channel, which works as a chatgroup and mainly includes ingroup users who wish to be kept informed.

We first compiled a dataset of Telegram posts from six openly CT-oriented channels fostering conversation on three different CTs: Flat Earth, the vaccine conspiracy, and the climate change hoax.

For each language, we found a different channel dedicated to the above-mentioned CTs, namely: *The Flat Earth Reality, No Vaccination. My Body My Choice* and *Climate Change HOAX* for English, and *Terra Piatta* ‘Flat Earth’, *Vax: le cavie siamo noi?* ‘Vax: are we the guinea pigs?’ and *Scie chimiche e clima* ‘Chemtrails and climate’ for Italian.

4.1 Annotation Tool

To obtain an actual CT corpus, we decided to trim out non conspiratorial comments by performing an annotation exercise. Two human annotators were asked to classify posts as either [Conspiratorial] or [Non-conspiratorial] using the *Taguette*¹ tool, a user-friendly and open-source online annotation environment. Other annotation tools were considered (e.g., INCEPTION, MAXQDA), but *Taguette* was eventually chosen because it allows to: import several file formats (including HTML); export the whole project, the annotated documents or just the performed annotation depending on the project’s needs; work remotely on the *Taguette* server; annotate by simply selecting the desired span of text and highlighting it using the desired labels.

We decided to import HTML documents in order to obtain an annotation-friendly visualisation of the Telegram data that preserves much of the original layout, thus enabling annotators to understand the chat dynamics. The files were first extracted from the selected Telegram channels,

¹ <https://www.taguette.org/>

slightly simplified using a clean-up tool² and then uploaded onto *Taguette* (Figure 1).

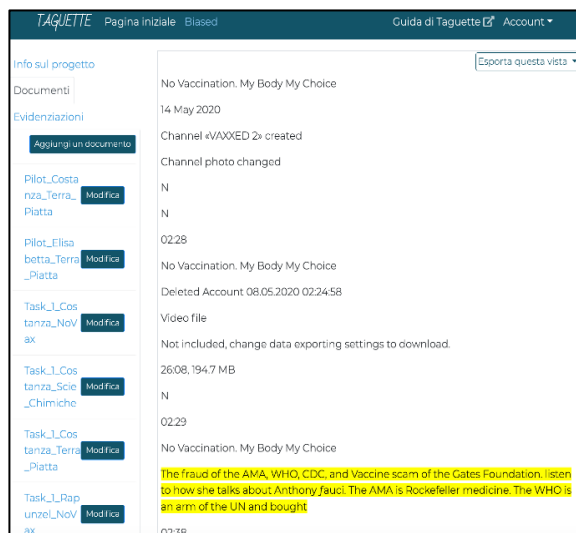


Figure 1: The *Taguette* annotation environment

The menu window on the left allows to access the project's info, the uploaded documents, and the existing annotations, called *evidenziazioni* 'highlights' because they appear as yellow highlights. On the right, you can see the beginning of the document *No vaccination. My Body My Choice*. Only one comment was annotated, namely a conspiratorial one: *The fraud of the AMA, WHO, CDC, and Vaccine scam of the Gates Foundation. listen to how she talks about Anthony Fauci. The AMA is Rockefeller medicine. The WHO is an arm of the UN and bought*.

4.2 Annotation Scheme and Guidelines

Dedicated guidelines were created to provide the two annotators³ with a general description of what CTs are (see § 2) so that they could distinguish [Conspiratorial] comments from [Non-conspiratorial] ones. It was chosen not to provide any linguistic or textual cues, in order not to skew the annotation results.

According to our annotation scheme, [Conspiratorial] comments are defined as "comments in which users (directly or indirectly) express themselves in favour of a CT", whereas [Non-conspiratorial] comments are "comments in which users (1) talk about CTs without expressing their stance or (2) talk about other topics, even unrelated ones".

The guidelines also specified what not to annotate, i.e., usernames, dates, times, Telegram channel names in isolation, footprints of multimedia files that were not included in the download, recurring comments that were not

actually written with a communicative aim but automatically posted in the chatroom. Moreover, what counted as comment was clearly specified. For instance, if a user conveyed content over several separately sent messages, each message was annotated as a separate comment. It was also agreed that the whole comment would be tagged and not smaller spans of text.

Figure 2 shows a portion of the chat conversation from the Telegram channel *The Flat Earth Reality*.

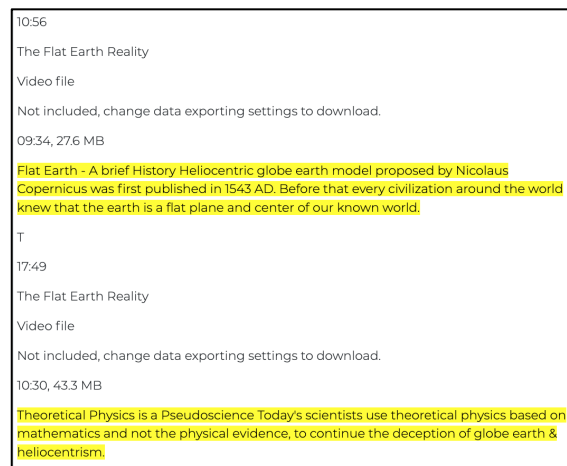


Figure 2: Annotating *The Flat Earth Reality* data

As you can see, the name of the channel, as well as dates, and multimedia files fingerprints were not tagged, while two comments were annotated as [Conspiratorial] because they both clearly show the author's belief that the Earth is flat and that the truth is being covered up (e.g., *scientists use theoretical physics based on mathematics and not the physical evidence, to continue the deception of globe earth & heliocentrism*).

4.3 Annotation Results

After a successful pilot test, the two annotators were asked to annotate at least 111 [Conspiratorial] comments per Telegram channel, in order to reach 1000 comments if the corpus was expanded to a third language⁴ following the same design, i.e., adding three other Telegram channels. By the end of the annotation, 1025 comments were annotated, of which 658 were identified as [Conspiratorial] and 304 as [Non-conspiratorial] by both annotators. According to Landis & Koch's (1977) interpretation of Cohen's κ , the two annotators reached a perfect level of Inter Annotator Agreement (IAA) in all documents, as summarised in Table 1 below.

² <https://www.htmlwasher.com/>

³ The two annotators are the first author of the paper and an archival specialist with prior experience in annotation.

⁴ The creation of a *German Complotto* corpus is currently underway.

Telegram channel	1st comment	Cohen's κ
<i>Terra Piatta</i>	Nov 2016	0.86
<i>Vax: le cavie siamo noi?</i>	July 2017	0.81
<i>Scie chimiche e clima</i>	Nov 2022	0.86
<i>The Flat Earth Reality</i>	Oct 2020	0.83
<i>No Vaccination.</i>	May 2020	0.84
<i>My Body My Choice</i>	Aug 2021	0.87

Table 1: IAA when annotating for [Conspiratorial] vs. [Non conspiratorial] comments.

Only the 658 comments that both annotators annotated as [Conspiratorial] – 317 Italian and 341 English – were included in the final corpus, which was then uploaded onto the *Sketch Engine* (Kilgariff et al., 2010) online platform to identify relevant indicators of CT discourse.

5. Corpus Methods and Results

The *Complotto* corpus counts 40,045 tokens but was uploaded onto the *Sketch Engine* platform as two separate corpora - the *Italian Complotto* and the *English Complotto* - of three documents each, in order to retrieve relevant language-specific and CT-specific features.

The *Italian Complotto* corpus counts a total of 22,252 tokens, of which 7,453 tokens in the *No-Vax* subcorpus (short for *Vax: le cavie siamo noi?*), 12,331 in the *Scie Chimiche* subcorpus and 2,468 in the *Terra Piatta* subcorpus. On the other hand, the *English Complotto* corpus counts 17,793 tokens, of which 4,612 in the *Climate Change* subcorpus, 8,434 in the *Flat Earth* subcorpus and 4,747 in the *My Body My Choice* subcorpus.

The two *Sketch Engine* functions that were used so far for the linguistic analysis are terminology extraction-related and are called *keywords* and *terms* (i.e., key multi-word expressions). They are able to extract keywords and terms by comparing the observed frequency data of a focus corpus and those of a larger reference corpus. Only words and multi-word expressions that appear with a statistically significant higher frequency in the focus corpus than in the reference one obtain *key status*. In our case, the reference corpus used for the *Italian Complotto* corpus was the *Italian Web 2020*, a 12-billion-word corpus made of Italian texts collected from the web, while the English reference corpus for the *English Complotto* was the *English Web 2021*, a 52-billion-word corpus of English web-crawled texts. As an advanced setting, we specified that keywords and terms should not be either too rare or too common and appear at least 4 times in the focus corpus to be considered for key status.

For the purposes of this study, we excluded CT-specific lexis from the analysis, such as the keywords *chemtrails* for the *Climate Change* subcorpus or *Flat Earth* for the *Flat Earth* one, because we are looking at common markers of CT discourse.

5.1 Italian Results

For the *Italian Complotto* corpus, *Sketch Engine*'s wordlist function was useful to confirm the insider vs. outsider group dynamics of the corpus (Holur et al. 2022), since the pronouns *noi* 'we' and *ci* 'us' appear as the most frequent personal pronouns overall.

Among the top five keywords of the Italian *No-Vax* subcorpus, we unsurprisingly found the lemmas *vaccino* 'vaccine', *vaccinare* 'to vaccinate', and *vaccinazione* 'vaccination', which were excluded from the analysis since they are CT-related. However, when looking at their observed modifiers, we found the expression *vaccinazione coercitiva* 'enforced vaccination', which contributes to indirectly conveying an interesting and novel conceptual metaphor, i.e., INSTITUTIONS ARE ABUSERS (specifically governments), which is in line with Lakoff and Johnson (1980)'s framework where an abstract target domain (as, in our case, INSTITUTIONS) is viewed in terms of a more concrete source domain (ABUSERS).

Among the top 50 keywords in the subcorpus we found three epistemically charged lemmas – *naturalmente* 'naturally', *assolutamente* 'absolutely' and *probabilmente* 'probably' – and seven lemmas clearly linked to the evidential necessity of providing a source for one's claims, i.e., *filmato* 'video', *news* 'news', *dichiarazione* 'declaration', *documentazione* 'documentation', *notizia* 'piece of news', *indicazione* 'direction', and *fonte* 'source'. These findings are in line with Scharloth et al. (2019) and Catenaccio (2022) and strongly suggest that the lexicon of the semantic field of EVIDENCE should be included among potential indicators of CT discourse. To introduce said evidence, we often find the presentative discourse marker *ecco* 'here' (e.g., *Ecco qui le prove schiaccianti... che non lasciano dubbi* 'Here the undeniable proof... which does not leave room for doubt'), which is also among the top 50 keywords of the subcorpus.

As a presentative discourse marker, *ecco* 'here' can be considered an indicator of dialogue acts (Bunt et al., 2010) characterised by an information-providing communicative function (ISO 24617-2), which seem particularly frequent in the whole corpus.

The only key multi-word term found in the subcorpus is *libertà di scelta* 'freedom of choice', which sums up the stance of chat members on the topic of vaccination and helps characterise them as the allegedly oppressed and threatened ingroup (Holur et al., 2022).

On the other hand, the most relevant *debunking* lemmas among the top 50 keywords in the *Scie Chimiche* subcorpus are the lemmas *mentire*, 'to lie', *menzogna* 'lie', and *manipolazione* 'manipulation', which all hint at the fact that another version of the truth allegedly exists but is hidden by official institutions (Ebling et al., 2013), in line with the conceptual metaphor

INSTITUTIONS ARE ABUSERS, specifically LIARS. No interesting key multi-word terms were found.

Finally, also among the top 50 keywords of the Italian *Terra Piatta* subcorpus, we find the lemmas *foto* 'photo', *test* 'test', *video* 'video', *prova* 'proof' – all belonging to the semantic field of EVIDENCE – as well as the lemma *fotomontaggio* 'fake photo' (actually among the top five), which implies that the evidence provided by scientific institutions such as NASA is actually unreliable and that, once again, INSTITUTIONS ARE ABUSERS, specifically LIARS.

5.2 English Results

Among the top 50 keywords of the *Climate Change* subcorpus, we can easily spot *debunking* expressions (Ebling et al., 2013) hinting at another version of the truth, such as *whistleblower* and *reportedly*. The latter is both an evidential and an epistemic marker used to signal that who writes is not the source of the information and that the information itself is not necessarily trustworthy (e.g., *Watch Argentinian Engineer Juan Baigorri Velar Reportedly Invented a Functional Rainmaker, But It Is Lost to History*). Similarly, the presence of the adjective *so-called*, which is always to be found within the key multi-word expression *so-called expert*, is aimed at discrediting the scientific community. The keyword *proof* fits among the lemmas making up the semantic field of EVIDENCE, together with the multi-word expression *full interview*. The presence of *poison* among the top keywords, and of *spray pollution* among the top terms, are both clear indicators of distrust in the institutions.

As for the *Flat Earth* subcorpus, the most interesting *debunking* keywords are *deception*, *hoax*, *debunk*, *scientism*, *conspiracy* and *fake*, which are all hinting at another alleged version of the truth (Ebling et al., 2013), as well as at the institutional and scientific responsibility behind said deception. This supports our proposed conceptual metaphor INSTITUTIONS ARE ABUSERS, specifically DECEIVERS. Last but not least, among the top fifty keywords of the *My Body My Choice* subcorpus, we can find the words *false*, *fully*, and *completely* - which are all "epistemically charged" expressions, to quote Catenaccio (2022, p. 31).

The next step in our analysis will be focused on corpus wordlists. From a first analysis of the *English Complotto* wordlist, we noticed several grammatical words that do not appear among the top 50 corpus keywords but have a considerably high rank in the *English Complotto* corpus and a much lower one in the *English Web 2021*. This is the case, for instance, of *here* (42nd most frequent word in the *English Complotto*, 132nd in the *English Web*) and *how* (54th in the *English Complotto*, 91st in the *English Web*). The following examples wish to provide insight on their use within the *English Complotto* corpus:

(1) Here's *why pilots can't prove curvature by demonstrating any change in level when flying between two locations, no matter how far apart.*

(2) *How can we be certain?*

(3) *Most people do not realise how integral artistic rendition is part of NASA's deception.*

Once again, the presentative discourse marker *here* in (1), and the conjunction *how* in (3) can be seen as indicators of information-providing communicative functions, while the use of *how* as an interrogative adverb in (2) points towards an information-seeking communicative function (Bunt et al., 2010).

6. Conclusion

According to the relevant literature, Conspiracy Theories (CTs) can be defined as "allegation[s] of conspiracy that may or may not be true" (Douglas et al, 2019, p. 4). However, their proliferation on social media is an undeniable threat to democratic societies.

In this study, after offering an interdisciplinary review of CT literature (§ 2) and related corpus works (§ 3), we have provided a detailed description of the design and annotation task that led to the creation of the *Complotto* corpus, a 40,000-token English-Italian corpus of conspiracy-oriented Telegram comments (§ 4). The *Complotto* corpus is a collection of 658 comments (317 Italian and 341 English) that were taken from three English- and three Italian-speaking Telegram channels focusing on the same three CTs: Flat Earth, the vaccine conspiracy, and the climate change hoax. In section 5, our corpus methods and results are explained.

The linguistic analysis we performed using the *Sketch Engine* online platform (Kilgariff et al., 2010) on our annotated data allowed us to assess the statistical relevance (both cross-linguistically and cross-CT) of the following phenomena thanks to the platform's *keywords* and *key terms* extraction functions, namely: (1) evidentiality and epistemic modality markers; (2) debunking vocabulary conveying the idea that another version of the truth lies behind the official one; (3) the conceptual metaphor INSTITUTIONS ARE ABUSERS.

All these features qualify as markers of CT discourse and have the potential to be effectively used as tags for future fine-grained semantic annotation tasks to develop systems of automatic CT identification. On-going work is focused on adding a third language to the corpus, namely German.

7. Acknowledgments

We thank the second annotator that contributed to the creation of the *Complotto* corpus and the two reviewers for their useful comments.

8. Bibliographical References

- Bodner, J., Welch, W., and Brodie, I. (2020). *COVID-19 conspiracy theories: QAnon, 5G, the New World Order and other viral ideas*. McFarland, Jefferson.
- Bunt, H., Alexandersson, J., Carletta, J., Chloe, J.W., Fang, A.C., Hasida, K., Lee, K., Petukhova, V., Popescu-Belis, A., Romary, L., Soria, C., and Traum, D. (2010). Towards an ISO standard for dialogue act annotation. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010)*, La Valletta, Malta, May. European Language Resource Association (ELRA).
- Catenaccio, P. (2022). A corpus-driven exploration of conspiracy theorizing as a discourse type. In M. Demata, V. Zorzi, & A. Zottola (Eds.), *Conspiracy Theory Discourses*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, pp. 25–47.
- Danesi, M. (2023). *Politics, lies and conspiracy theories: A cognitive linguistic perspective*. Routledge, Oxon/ New York.
- Demata, M., Zorzi, V., and Zottola, A. (2022). Conspiracy theory discourses – Critical inquiries into the language of anti-science, post-truthism, mis/disinformation and alternative media. In M. Demata, V. Zorzi, & A. Zottola (Eds.), *Conspiracy Theory Discourses*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, pp. 1–22.
- Douglas, K. M., and Leite, A. C. (2017). Suspicion in the workplace: Organizational conspiracy theories and work-related outcomes. *British Journal of Psychology*.
- Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., and Deravi, F. (2019). Understanding conspiracy theories. *Advances in Political Psychology*, 40(1):3–35.
- Ebling, S., Scharloth, J., Dussa, T., and Bubenhofer, N. (2013). Gibt es eine Sprache des politischen Extremismus? In *Die da oben. Sprache, Politik, Partizipation*. Bremen: Hempen Verlag, pp. 43–69.
- ISO. (2020). *Language resource management-Semantic annotation framework (SemAF) - Part 2 - Dialogue acts*. Standard, Geneva.
- Kilgarriff, A., Vitek, B., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, Rychlý, P., and Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1):7–36.
- Lakoff, G., and Johnson. M. (1980). *Metaphors We Live By*. University of Chicago Press, Chicago.
- Landis, J.R., and Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Mancosu, M., and Vassallo, S. (2022). The life cycle of conspiracy theories: evidence from a long-term panel survey on conspiracy beliefs in Italy. *Italian Political Science Review/ Rivista Italiana di Scienza Politica*, 52:1–17.
- Miani, A., Hills, T., and Bangerter, A. (2021). LOCO: The 88-million-word language of conspiracy corpus. *Behavior Research Methods*.
- Mompelat, L., Tian, Z., Kessler, A., Luetgen, M., Rajanala, A., Kübler, S., and Seelig, M. (2022). How “Loco” Is the LOCO Corpus? Annotating the Language of Conspiracy Theories. In *Proceedings of the 16th Linguistic Annotation Workshop (LAW-XVI) within LREC2022*, pages 111–119, Marseille, France. European Language Resources Association.
- Russo, G., Stoehr, N., and Horta Ribeiro, M. (2023). ACTI at EVALITA 2023: Automatic Conspiracy Theory Identification Task Overview. In *EVALITA 2023: 8th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*, Parma, Italy, September. CEUR.
- Scharloth, J., Obert, J., and Keilholz, F. (2019). Epistemische Positionierung in verschwörungstheoretischen Texten. Korpuspragmatische Untersuchung von epistemischer Modalität und Evidentialität am Beispiel der Holocaustleugnung. *Zeitschrift für Diskursforschung*, 4:159–198.
- Tangherlini, T.R. (2018). Toward a generative model of legend: Pizzas, bridges, vaccines, and witches. *Humanities*, 7(1):1.
- Uscinski, J. E., Parent, J., and Torres, B. (2011). Conspiracy theories are for losers. In *APSA 2011 Annual Meeting*.
- Uscinski, J. E. (2020). *Conspiracy theories. A Primer*. Rowman & Littlefield, London.