

How “Loco” is the LOCO Corpus? Annotating the Language of Conspiracy Theories

Ludovic Mompelat[†], Zuoyu Tian[†], Amanda Kessler[†], Matthew Luetttgen[†],
Aaryana Rajanala[†], Sandra Kübler[†], Michelle Seelig[‡]

[†] Indiana University, [‡] University of Miami

{lmompela, zuoytian, amckess, mluettg, aarajana, skuebler}@iu.edu, mseelig@miami.edu

Abstract

Conspiracy theories have found a new channel on the internet and spread by bringing together like-minded people, thus functioning as an echo chamber. The new 88-million word corpus *Language of Conspiracy* (LOCO) was created with the intention to provide a text collection to study how the language of conspiracy differs from mainstream language. We use this corpus to develop a robust annotation scheme that will allow us to distinguish between documents containing conspiracy language and documents that do not contain any conspiracy content or that propagate conspiracy theories via misinformation (which we explicitly disregard in our work). We find that focusing on indicators of a *belief* in a conspiracy combined with textual cues of conspiracy language allows us to reach a substantial agreement (based on Fleiss’ kappa and Krippendorff’s alpha). We also find that the automatic retrieval methods used to collect the corpus work well in finding mainstream documents, but include some documents in the conspiracy category that would not belong there based on our definition.

Keywords: corpus, conspiracy theory, annotation scheme

1. Introduction

Conspiracy theories have found a new channel on the internet and spread by bringing together like-minded people, thus functioning as an echo chamber that accelerates the spread of conspiracy theories and contributes to the further polarization of extremes (e.g., (Papacharissi, 2016)). In recent years, researchers have thus become interested in the mechanisms of how conspiracy theories are spread, and which types of people are susceptible to them (Barkum, 2013; Douglas and Sutton, 2018; Samory and Mitra, 2018, a.o.).

Miani et al. (2021) created LOCO, the Language of Conspiracy Corpus. They collected the large-scale corpus from automatically retrieved texts using a seeding approach, one subcorpus focusing on conspiracy theory documents, and the second subcorpus focusing on mainstream documents for each seed. The corpus creators intend this corpus to serve as a basis for investigating the linguistic differences between conspiratorial and mainstream texts.

We use the LOCO corpus as the basis for our work. Ultimately, our goal is to create machine learning approaches that can tell conspiracy content from mainstream content, ideally independent of the individual conspiracy theory. As a first step, we needed to determine how well the retrieval strategies of the LOCO corpus worked, in other words, whether the grouping of documents into conspiracy or mainstream subcorpora was reliable. This led to an annotation project, in which we annotated a considerable number of texts from two different conspiracy theories, and in the process created annotation guidelines. We used documents using the seed “Sandy Hook” as our first set, and documents retrieved using the seed “Coronavirus” to decide whether our annotation guidelines were applicable

across different conspiracy theories. Sandy Hook refers to conspiracy theories centered around the Sandy Hook Elementary School shooting in 2012, including claims that the shooting was staged by the US government, potentially to establish tighter gun control regulations; that nobody died in the event; or that there was a second conspirator, etc. The Coronavirus conspiracy theory revolves around claims that the virus was engineered in China; that the virus was spread by elites to gain influence and increase profit; or that the vaccine is more dangerous than the virus, etc.

The paper is structured as follows: Section 2 explains our research questions in more detail and section 3 introduces related work. Section 4 presents the LOCO corpus, and section 5 describes the first round of annotations of Sandy Hook documents. Section 6 explains the adaptation of the annotation guidelines based on the first annotations, section 7 the experiment on annotating documents from a different seed, and section 8 gives an overview of all the annotations. Section 9 concludes and describes future work.

2. Research Questions

The goal of this project is to annotate the texts of the LOCO corpus (Miani et al., 2021) for conspiracy theory language. This is a challenge that has not been addressed in this form before (but see the next section). Similar to abusive language detection (e.g., (Lopez Long et al., 2021)), we assume that this type of annotation is non-trivial, since the categories sound intuitive at first but tend to have soft boundaries, which may depend on personal stance and knowledge of the annotator. In order to develop a robust annotation scheme, we need to answer the following research questions:

1. Can we start with a minimal definition of conspiracy theory, and use the difficulties arising from applying this definition in annotation to develop robust annotation guidelines that will lead to high inter-annotator agreement?
2. Do our guidelines cover both conspiracy theory and mainstream texts? Are there differences in the annotation quality between these two types of texts?
3. If the guidelines are developed based on texts from one specific conspiracy theory, are they robust enough so that they can be applied to texts from other conspiracy theories?

Additionally, we will have a look at the quality of the texts in the LOCO corpus. Since the corpus texts were collected automatically without human supervision, it is important to know how reliable the search criteria are that were chosen to create the corpus, and whether the reliability of the retrieval strategies is dependent on the relevant conspiracy theory.

3. Related Work

Before creating guidelines for annotating the language of conspiracy theories, we first need a working definition of what constitutes a conspiracy theory, and what constitutes a conspiracy theory text. Banas and Miller (2013) define conspiracy theories (CTs) as “causal narratives of an event as a covert plan orchestrated by a secret cabal of people or organizations instead of a random or natural happening.” Douglas et al. (2019) define them as “attempts to explain the ultimate causes of significant social and political events and circumstances with claims of secret plots by two or more powerful actors”. Miani et al. (2021) define CTs as follows: “Conspiracy theories are narratives that attempt to explain significant social events as being secretly plotted by powerful and malicious elites at the expense of an unwitting population.”

Samory and Mitra (2018) identify three key elements of previous CT definitions: agent, action, and target. In the work by Douglas et al. (2019), for example, the agent is “two or more powerful actors”, the target is “significant social and political events and circumstances”, and the action is “attempts to explain the ultimate causes ... with claims of secret plots”. We can easily identify these three key elements from a theoretical level, and Samory and Mitra (2018) show that such methods work well for conspiratorial statements in real texts. However, Samory and Mitra (2018) also point out that “conspiracy theories are often collages of many smaller scale theories”, which makes them a difficult phenomenon to study.

Investigating linguistic characteristics of conspiracy theories, Fong et al. (2021) identify lexical cues that represent “psychological themes” relevant to “conspiracy ideation” identification, for example *ingroup vs.*

outgroup language or the “we vs. them” ideology, and *cognitive processes* creating a higher past- and certainty-oriented language that is focused on causal explanations and closure. This distinguishes the language of CT from that of mainstream media, which is more oriented towards the factuality of information. In addition to lexical cues, the authors also identify lexical themes based on power, death, and religion. Introne et al. (2020) use a narrative framework to investigate conspiracy theory texts. They use the following definition: “A conspiracy theory is a narrative explaining an [event or series of events] that involve [deceptive, coordinated actors] working together to achieve [a goal] through [an action or series of actions] that have consequences that intentionally disenfranchise or harm an [individual or population].” They identify six main terms, marked in the square brackets above. Additionally, they distinguish between CTheory (for which annotators need to distinguish actors, actions, consequences, and victims) and CThinking for posts that “implied a conspiracist point of view ... but did not themselves contain identifiable CTheories”. For this category, only one of the six categories needed to be present. Introne et al. found that CTheories are very infrequent in their data, CThinking less so. Additionally, CT posts mostly focused on actors and actions.

The next problem to be addressed concerns how to compile a corpus of CT documents. CT researchers have studied texts with potential CT content on different social media platforms such as Twitter, Facebook, and Reddit (Wood, 2018; Smith and Graham, 2019; Samory and Mitra, 2018). However, Miani et al. (2021) argue that texts from discussion threads are not a good resource for investigating CT narratives and tracking how CT beliefs are transmitted, because in most cases, Twitter comments, etc. are short and very contextualized in a (potentially asynchronous) “conversation”, and it is difficult to interpret such posts independent of the whole thread.

Instead of extracting potential CT content from social media resources, other efforts focused on building CT corpora using full documents. For example, Uscinski et al. (2011) compiled a corpus of conspiracy documents using letters to the editor of *The New York Times* from 1897 to 2010. This corpus contains 100 000 documents, out of which 800 were manually annotated as conspiracies. Unfortunately, this corpus is no longer available (p.c. J. Uscinski, 2021). The most recent, large-scale corpus of conspiracy documents was released by (Miani et al., 2021), it covers a wide range of different conspiracy theories and was collected automatically using a seeding approach. This is the corpus we will use for our work, for more details see below.

4. The LOCO Corpus

Miani et al. (2021) created the Language Of Conspiracy Corpus (LOCO) (Miani, A. et al., 2021), which contains 23 937 conspiracy and 72 806 mainstream

Topic	Category	Round	5 ann. agree	4+ ann. agree	Fleiss' kappa	Krippendorff's alpha
Sandy Hook	CT	1	9/20	17/20	0.466	0.469
	mainstream	1	18/20	20/20	-0.020	-0.010
Sandy Hook	CT	2	14/20	17/20	0.696	0.699
	mainstream	2	20/20	20/20	1.0	1.0
Coronavirus	CT		12/20	17/20	0.577	0.575
	mainstream		19/20	20/20	-0.010	0

Table 1: Inter-annotator agreement for documents from two CT seeds, when annotating for CT vs. non-CT.

documents, about 88 million words overall. All texts were retrieved based on a set of seeds, following the strategy used for the WaCky corpus (Baroni et al., 2009). The seeds were collected from a national poll¹, a list of 17 items from Douglas and Sutton (2018), plus an additional "20 seeds corresponding to popular (e.g., Illuminati, genetically modified organisms, Pizzagate) and current (e.g., coronavirus, Bill Gates, 5G) CTs" chosen by Miani et al. (2021).

There are two categories used in the corpus, conspiracy and mainstream documents, which are retrieved via different strategies: To gather conspiracy texts, Miani et al. (2021) used a list of conspiracy theory websites based on scores from mediabiasfactcheck². To retrieve mainstream documents, the authors used Google to search for the seeds and extracted website domains, from which they retrieved the texts. The authors acknowledge that not all conspiracy theory (CT) texts will contain conspiracy content. Mainstream documents may contain CT content, but they reflect the mix of CT and non-CT that the general public is exposed to.

Compared to previous corpora on related areas (conspiracy, rumors, fake news (e.g. (Uscinski et al., 2011; Kwon et al., 2017; Castelo et al., 2019)), the LOCO corpus covers a large set of conspiracy texts and a sizable number of different CTs, plus a detailed set of metadata, including date, website, and measures of social media engagement. To determine the accuracy of the CT and mainstream categories, Miani et al. (2021) randomly sampled 60 documents from the conspiracy and mainstream subcorpora each, and manually annotated them. Their annotation results indicate that 85% of the conspiracy documents and 92% of the mainstream documents are correctly labeled.

It is clear that the LOCO corpus is a valuable resource for exploring the narratives of conspiracy theories and their effect on social media. However, in order to use this corpus for creating machine learning models of CT, we need a better understanding of the quality of the corpus, i.e., the degree to which the automatic grouping into the CT and mainstream subcorpora corresponds to human judgments across the different seeds.

¹<https://www.publicpolicypolling.com/polls/democrats-and-republicans-differ-on-conspiracy-theory-beliefs/>

²<https://mediabiasfactcheck.com/conspiracy/>

5. Annotating Sandy Hook Documents

5.1. Distinguishing Conspiracy Theory Texts from Mainstream

Our first question concerns the problem of defining the target of our annotations. What do we consider a conspiracy theory (CT) document? Where do we draw the line between conspiracy theory and mainstream / non-conspiracy theory? To answer these questions, we conducted a first round of annotations on a sample of 40 documents from the set of documents in the LOCO corpus on Sandy Hook. We chose 20 documents from the conspiracy subcorpus and 20 from the mainstream subcorpus. The annotations were conducted by 2 undergraduate and 3 graduate students, who had read and discussed relevant literature prior to the annotations.

Our starting definition of CT was the definition by Douglas et al. (2019) (see section 3). However, after our pilot annotation, we found this definition too general for our goal since it does not give any guidance on the distinction between reports of the event, reports of conspiracy theories related to the event, and the propagation of conspiracy theories. Since we are mostly interested in the latter, we decided to incorporate the concept of *conspiracy belief*, as defined by Barkum (2013): "A conspiracy belief is the belief that an organization made up of individuals or groups was or is acting covertly to achieve some malevolent end." We adopted the definition proposed by Seelig et al. (2022), which is based on the definition by Banas and Miller (2013):

- (1) A conspiracy belief is the belief that an organization made up of individuals or groups was or is acting covertly to achieve some malevolent end. It depicts causal narratives of an event as a covert plan orchestrated by a secret cabal of people or organizations instead of a random or natural happening.

The results of this first annotation round are shown in the first two rows of Table 1. We found that the mainstream documents from the LOCO corpus were mostly labeled correctly, and our annotators agreed in most cases: Only 2 documents had 1 annotator disagreeing with the majority. Note that the kappa and alpha values for the mainstream subcorpora show either negative numbers or 0 even though the annotators mostly agreed. The reason for this can be found in the very

How Zionist Politicians Brought On Newtown Killings - Part 2

The first half of this analysis of the Connecticut shootings, MK-ULTRA Links to Sandy Hook Assault, examined how the CIA's mind-control program spread like a metastasizing cancer across the Eastern Seaboard, delivering a nightmarish cocktail of synthetic drugs, sexual abuse and lethal violence. The focus of that essay was on the three major players in the New England region - CIA/FBI agents, the pedophile Catholic clergy and the Irish drug-trafficking mob, while giving only passing mention of the Jewish politicians whose salesmanship was needed for the monumental task of social engineering a proud nation into a herd of sheep.

The major political figure in the Newtown tragedy who has once again evaded personal responsibility for the bloody consequences of his idiotic policies, which include the war in Iraq and arms shipments to Israel, is Joseph Isadore Lieberman. The chairman of the Homeland Security Committee and U.S. senator from Tel Aviv and Stamford is the elephant in the schoolroom that nobody seems to notice.

Soon to retire from the senatorial seat he's kept warm for 22 years, Joe Lieberman has been a contemporary of New England gangland boss Whitey Bulger and his CIA controllers. His Senate term has run exactly parallel to the takeover and transformation of once-puritanical Connecticut into a sleazy hub of underage prostitution, child porn, drug peddling and gambling.

Without the powerful senator's protection and nurturing of unsavory characters and corporate criminals over the decades, the Sandy Hook school massacre probably would never have happened. Here, in Part 2, the role of Jewish politicians in first promoting and later suppressing child prostitution, kiddie porn and drug use is explored, along with the blowback from that policy reversal provoking the school attack and subsequent cover-up.

Figure 1: Example of a clear CT document [LOCO ID: C006b9].

Trial Date Set in Sandy Hook Families' Lawsuit Against Remington – Infinite Unknown

A lawsuit by families of Sandy Hook victims is proceeding against Remington, manufacturer of the AR-15, in the new push to hold gun manufacturers responsible for what is done by people who purchase their products and use them illegally.

The New York Times is pretty excited about it: The legal challenge faces long odds, and a key hearing next week will determine its future.

Question: Do you think it's a bit hypocritical of the system to applaud the Sandy Hook families for suing Remington and decry the fact that people can't sue to hold a company responsible for what people who purchase its products could potentially do to others, but completely ignore the fact that we live in a country where no one is allowed to sue vaccine manufacturers directly for vaccine damage?

Also, can you imagine if every company could be sued for every time someone used their products in the commission of a crime to hurt someone else?

Knife manufacturers sued for stabbings... Car manufacturers and alcohol producers sued for DUI deaths... Companies who sell lighters sued if an arsonist decides to burn someone's house down... Shoelace manufacturers sued for someone being strangled by one... Swimming pool manufacturers being sued if someone drowns in one... Personal responsibility be damned when there's an agenda, and this agenda is pretty obvious. If they can't get the laws passed to gut the Second Amendment, they'll just try to sue gun manufacturers out of existence instead.

Figure 2: Example of a document difficult to label [LOCO ID: C06962].

high expected values. Neither metric is useful for data with very high agreement and small sample size (Zhao et al., 2013). Given the results in Table 1, we decided to trust the retrieval strategy used for mainstream documents, with which the annotators agreed in most cases. For the CT documents, however, inter-annotator agreement was low, only for 9 out of the 20 documents did all 5 annotators agree, and Fleiss' kappa reached 0.466. Figure 1 shows a clear case of CT.

When discussing the documents on which the annotators did not agree, we found that in some cases, a conspiracy theory may be perpetuated, but the text itself did not show any evidence of the writer's belief in the CT. Other examples were unclear. One example for such a difficult decision is shown in Figure 2.

This article was particularly difficult to label as CT or non-CT: While the author is clearly opposed to the lawsuit against gun manufacturers, and while the document contains leading questions (e.g., "Also, can you imagine if every company could be sued for every time someone used their products in the commission of a crime to hurt someone else?") and mentions an "agenda", there is no indication of a belief in a conspiracy. After the discussion, all annotators agreed that this text should be classified as non-CT.

Many of those documents contain statements that were verifiably incorrect or misleading and that would indicate covert activities with malevolent intentions. An example of such a document is shown in Figure 3. In this example, the claim that the property records show

Over 30 Sandy Hook Homes "Gifted" In 2009
 Newtown property records suggest that on December 25, 2009 a total of 35 properties located on and around Yogananda Street in Sandy Hook were transferred at zero value to new owners. The transactions include the house belonging to mysterious figure Chris Manfredonia, who was apprehended by police on Sandy Hook School grounds on the morning of December 14, 2012. "It's not just Yogananda Street that was given away on Christmas of '09," the researcher argues. Yogananda addresses 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 21, and 23 all bear identical transactions to the ones exhibited here; 24 is owned by the town, while 18 is a normal transaction. On Charter ridge, 45, 47, 63, 71, and 72 appear normal and 61, 62, 64, 65, 66, 67, 68, 69, 70, and 73 are all December 25, 2009 transactions. All of these properties surround the Lanza home.

Figure 3: Example of a misinformation/fake news document [LOCO ID: C060d0].

Democrats Call For A Complete Ban on All Cryptocurrencies
 Brad Sherman told a subcommittee for the House of Representatives Financial Services. Democrats are calling for a blanket ban on all forms of Cryptocurrencies including Bitcoin, claiming that digital money warrants heavy regulation by lawmakers. Congressman Brad Sherman told a subcommittee for the House of Representatives Financial Services that the American public should not be allowed to purchase any form of digital currency. "We should prohibit U.S. persons from buying or mining cryptocurrencies," the California Democrat said. According to Coindesk: He added that, beyond cryptocurrencies being potentially used as a form of money in the future, it can currently be used by tax evaders and rogue states seeking to bypass U.S. sanctions.

Figure 4: Example of an unrelated article [LOCO ID: C05e2a].

that 35 properties were transferred at zero value to new owners on the same day is technically correct but ignores that this was due to missing information in the computer system.

We decided that we would focus on the *language of conspiracy theories*, in the sense that a belief in the CT was shown in the language of the text. We consider fact checking a separate, but clearly related problem. When focusing on the language of CT, this document can be considered non-CT, even though the misinformation indicates a CT. This decision was made in order to keep the annotations feasible given time and budget constraints. A combination of fact checking and information about conspiracy beliefs anchored in the language will have to be addressed in the future.

5.2. Relatedness

From a cursory look at the CT documents, it became clear that solely distinguishing between conspiracy and non-conspiracy was not sufficient since we found that some documents, which were collected for a specific seed, may mention that seed, but were otherwise unrelated to the CT. For this reason, we added a *Relatedness* category, with three different labels: closely related, broadly related, and not related.

Figure 4 shows an extreme example. This is a text on cryptocurrencies, but the corpus groups it under the seed Sandy Hook. There is no mention of Sandy Hook in the whole document, and it is unclear how it was retrieved. We consider this document not related to the Sandy Hook CT.

The first two rows in Table 2 show the results of the first round of annotations. We see a similar picture to the annotations of CT vs. non-CT for the mainstream

documents, Fleiss' kappa reaches 0.512. For the CT documents, the results are higher, with Fleiss' kappa reaching 0.655. A closer look at the documents where annotators disagreed shows that disagreements concern the hard boundaries between the labels. Is one cursory mention of Sandy Hook enough to make a document closely related? Does Sandy Hook need to be the only topic in a document for it to count as closely related? For the future, we will investigate a continuous scale for this type of annotation.

6. The New Annotation Scheme

After the first round of annotations and the discussion of the documents that had conflicting annotations, we updated the definition in (1) to the one in (2).

- (2) A conspiracy belief is the belief that an organization made up of individuals or groups was or is acting covertly to achieve some malevolent end. It depicts causal narratives of an event as a covert plan orchestrated by a secret cabal of people or organizations instead of a random or natural happening. A document is considered CT if and only if such a belief is manifested in the text via specific expressions. We explicitly exclude fact checking beyond obvious inconsistencies with information present in mainstream coverage of the event underlying the CT.

Given the retrieval strategies used in the creation of the LOCO corpus (see section 4), there are obvious differences since most of the mainstream documents are retrieved from news outlets while the CT documents tend to come from less official outlets. Thus, docu-

Topic	Category	Round	5 ann. agree	4+ ann. agree	Fleiss' kappa	Krippendorff's alpha
Sandy Hook	CT	1	11/20	16/20	0.655	0.657
Sandy Hook	Mainstream	1	8/20	13/20	0.512	0.508
Sandy Hook	CT	2	16/20	18/20	0.776	0.778
Sandy Hook	Mainstream	2	16/20	18/20	0.819	0.820
Coronavirus	CT		17/20	18/20	0.751	0.753
Coronavirus	Mainstream		13/20	19/20	0.517	0.518

Table 2: Inter-annotator agreement on relatedness for documents from two CT seeds

cue	example
contradictory	FBI says No One Killed at Sandy Hook [LOCO ID: C005a9] Watch Infowars explore why people believe the Sandy Hook shooting to be a hoax. [LOCO ID: C042fa] We at Prepare for Change (PFC) bring you information that is not offered by the mainstream news, and therefore may seem controversial. [LOCO ID: C0443c]
sensational	Americans Under Surveillance [LOCO ID: C00650] If you want more evidence of a government seeking control, look no further than the IRS scandal where the Obama administration was using the IRS to stop conservatives and religious groups from organizing opposition. [LOCO ID: C00650] MK-ULTRA is obsolete when private medical insurance plans are covering the costs of date-rape capsules [LOCO ID: C006b9]
other CT	Internet sleuths immediately took to the web to stitch together clues indicating the shooting could be a carefully-scripted false flag event, similar to the 9/11 terror attacks, the central tenet being that the event would be used to galvanize future support for gun control legislation [LOCO ID: C005a9]
all caps	RED ALERT: Google Censorship Is Destroying the Truth Movement [LOCO ID: C00775] They reported Fake numbers that they made up & don't even exist. WE WILL WIN AGAIN! [LOCO ID: C00775]
named entities	The Obama White House [LOCO ID: C00a2d] 'Sleepy Joe' makes another gaffe on his campaign trail [LOCO ID: C0690d]
punctuation	Somebody is going to jail over this un-constitutional crime!!! [LOCO ID: C00a2d]
pronouns	I am aware of books by former insiders that describe the CIA's alliance with members of the media. When I was a member of the congressional staff, I was warned of the Washington Post's collaboration with the CIA. [LOCO ID : C0487a]
questions	Lauren Rousseau's Car Riddled With Bullet Holes In Sandy Hook Parking Lot? [...] how is it possible for a bullet hole to penetrate the side of her car at the trajectory shown above? Was there no car beside her? This is just one of the many mysteries about the official story. More research coming in different articles, stay tuned.[LOCO ID: C06689]
paraphrases	Recently released FBI crime statistics curiously show that no murders occurred in Newtown, Connecticut, in 2012, despite reports that numerous schoolchildren and faculty members were slaughtered during a shooting rampage in December of that year. [LOCO ID : C005a9] Mark Zuckerberg Says That Social Media Giant Facebook Will Continue To Give A Voice To Holocaust Deniers [LOCO ID : C00b0f]

Table 3: Verbal and textual cues for CTs.

ments grouped into the CT category tend to contain formatting, spelling, and sentence and discourse structure anomalies. For this reason, we created a set of cues that can help the annotators make decisions. The cues listed below are the ones that the annotators listed when asked what they noticed in CT texts. However, note that the cues individually or in their entirety do not constitute a justification for labeling a document as CT. Instead, these cues are used as *supporting evidence* in the decision process. In order to classify a document as CT, we need verbal signs of a conspiracy belief.

We use the content and textual cues described below. The first set of cues focuses on content, examples are shown in the upper half of Table 3.

1. Contradictory opinion to mainstream opinion
Such cues consist of opinions that contradict opinions in the general domain. Note that this does not require elaborate fact checking.
2. Sensationalism
Headlines and content are written to excite strong emotions, often at the expense of correctness.

If you're anything like me... as soon as you hear the news about a big shooting or a terrorist attack in Europe or America, you roll your eyes and yawn. Then you go pop some popcorn and kick back in your recliner to watch the amusing theatrics that ALWAYS follow. [...] "Nope... today in 2018... anything that makes simultaneous nationwide headlines and is covered non stop for a week or even a couple days... is ALWAYS a faked hoaxed event. [...] I know it's hard to swallow... that they would prefer to use a hoax model over just really killing ppl. But they've been using the "hoax false flag" now since about 2008. And here is why they fake all of these events instead of just sending a patsy in and really killing victims.

The deep state learned their lesson after really killing ppl in the false flag of 9-11. The victims families could not be controlled or managed to say the things they wanted them to say or push the agenda they wanted pushed. [...] The McDonnell family – their daughter, Grace, was allegedly shot dead at Sandy Hook

Fake victims/no real deaths = crisis actors playing loved ones. Crisis actors instead of real heartbroken angry loved ones = no lawsuits and NO QUESTIONS. [...] [LOCO ID: C0443c]

Figure 5: Example of a CT document with clear CT language.

'Something's going on! Please!' Harrowing 911 calls from inside Sandy Hook Elementary School during massacre reveal staff desperately begging for help as dispatchers respond calmly. [LOCO ID: M1f6ae]

Figure 6: Example of a non-CT document with some of the verbal cues typical for CT.

3. Mentions of other conspiracy theories

CT documents often draw connections between different conspiracy theories.

There are also textual cues that are indicative of CTs, many of these cues are typically also used in other social media (as opposed to news reports). Examples are shown in the lower half of Table 3.

1. Extensive use of all caps
2. Atypical named entities
3. Unconventional use of punctuation
4. Frequent use of 1st and 2nd person pronouns
5. Frequent questions directed at the reader
6. Paraphrasing instead of direct quoting

Several documents in the LOCO corpus were written by or reference prominent conspiracy theory proponents such as Alex Jones and Infowars. If we were interested in conspiracy theories in general, such documents should be labeled as CT. Given our definition in (2), such documents are labeled as non-CT since they do not contain any language showing the belief in a CT. Figure 5 shows an example of a document that caused doubts based on our first definition in (1) but was considered a clear case of CT based on the new definition. In this document, we clearly see language relating to the conspiracy theory, e.g., "Fake victims/no real deaths = crisis actors playing loved ones. Crisis actors instead of real heartbroken angry loved ones = no lawsuits and NO QUESTIONS." Additionally, it shows a range of the cues we have identified: "The deep state", informal language, words in all caps to show emphasis, and repeatedly the hedge "allegedly".

In some cases, however, the verbal cues complicated the decision. Figure 6 shows an example containing

verbal cues of emotional language, ("harrowing") and quotations indicating panic. This language seems to imply that the "calm" response was inappropriate in that situation. Within the remainder of the document, however, there is no claim of a secret plot, etc. Consequently, we annotated this document as non-CT.

The lack of clarity in these documents may allow readers to impose their pre-existing beliefs or worldview; in this way, the CT is perpetuated in part because it can mean different things to different people, thus contributing to the multi-faceted collection of beliefs centered around one CT.

After establishing our new annotation scheme, we conducted a second inter-annotator agreement experiment on 20 previously unseen documents from the CT subcorpus and 20 from mainstream. The inter-annotator agreement results are shown in rows 3 and 4 in Table 1 for the decision on CT vs. non-CT and in Table 2 for relatedness. Note that we reached a perfect agreement on the mainstream documents for CT vs. non-CT, thus corroborating our decision to trust the retrieval strategy for this subcorpus. For the CT documents, Fleiss' kappa increased from 0.466 to 0.696. For relatedness, we also see a marked improvement in Fleiss' kappa from 0.655 to 0.776 for CT documents and from 0.512 to 0.819 for mainstream documents, but we do not reach a perfect agreement. All scores correspond to substantial agreement based on Landis and Koch (1977).

7. Using the Annotation Scheme for Coronavirus Documents

The second round of annotation in Sandy Hook documents shows that annotators reach a high agreement in annotating for both conspiracy and relatedness. This leads to the next question, namely whether the annotation guidelines developed based on texts on Sandy Hook will also be relevant for the annotation of other

Seed	Category	# LOCO docs	# docs annotated	Conspiracy Rate	Related Rate
Sandy Hook	Conspiracy	364	364	0.615	0.615
Sandy Hook	Mainstream	1476	200	0.020	0.780
Coronavirus	Conspiracy	571	571	0.413	0.891
Coronavirus	Mainstream	1965	20	0	0.850

Table 4: Statistics of our annotations.

CTs, or whether we need to adapt the guidelines to new CTs.

To answer this question, we chose a second seed from the LOCO corpus: coronavirus. This choice was determined in the attempt to find a CT that is different enough from Sandy Hook. The coronavirus CT concerns ongoing events, unlike the Sandy Hook CT, where the focus event happened in 2012. Furthermore, while the majority of the narratives on Sandy Hook are centered around the event of the school shooting, there is no such core event for coronavirus. Our hypothesis is that the coronavirus texts are more diverse in topics than the Sandy Hook ones, therefore if the annotation guidelines are usable for coronavirus, they should also be usable for a wider range of CTs.

We conducted a third inter-annotator agreement experiment on 20 CT and 20 mainstream documents for the coronavirus seed. From the annotation results in rows 5 and 6 of Tables 1 and 2, we see a clear divergence. For CT vs. non-CT, both Fleiss’ kappa and Krippendorff’s alpha are considerably lower for these documents than for the second round of Sandy Hook documents (kappa: 0.577 vs. 0.696), clearly showing that the CT documents are structured differently in different CTs. For the mainstream documents, in contrast, the results are very similar to the second round of Sandy Hook annotations. For relatedness, the CT documents had similar trends to the second round of Sandy Hook while the mainstream documents reached lower scores (kappa: 0.517 vs. 0.751).

These differences can partly be explained by the differences in the success of the automatic retrieval strategies in LOCO (see below for more details): A much higher percentage of the documents in the CT subcorpus for the seed coronavirus are non-CT based on our definition. Additionally, in comparison to the Sandy Hook documents, a higher percentage of documents in both subcorpora for the coronavirus seed are related to the topic.

8. Overview of All Annotations

We re-annotated the documents from the first inter-annotator agreement experiment and continued annotating the remaining documents in the conspiracy subcorpus for both seeds. An overview of the complete set of annotations is shown in Table 4. Here the conspiracy rate refers to the percentage of documents of a subcorpus that were annotated as CT by our annotators. We see that for both seeds, the conspiracy rate is very low for mainstream documents (0.020 for Sandy

Hook and 0.0 for coronavirus). However, the rate is also rather low for the CT documents, showing that less than 2/3 of the documents in the Sandy Hook CT subcorpus, and less than half of the documents in the coronavirus CT subcorpus, actually contain CT language. The relatedness rate refers to the percentage of documents that were labeled as closely or broadly related to the seed CT by the annotators. Here we see a similar trend to the Sandy Hook CT subcorpora, a much higher rate for the coronavirus CT subcorpus, and lower rates for the mainstream subcorpora: 0.780 for Sandy Hook and 0.850 for coronavirus. These numbers show very clearly that the two retrieval strategies work qualitatively differently for different CT seeds.

9. Conclusion and Future Work

We have investigated the annotation of documents in the LOCO corpus for the presence of language that indicated a belief in conspiracy theories. Our experiments show that the automatic retrieval methods used to create the LOCO corpus reach different levels of conspiracy content and relatedness for the two seeds that we used for our investigation. We also find that distinguishing between CT and non-CT is a difficult and subjective task. Our annotation guidelines can help with consistent decisions across different annotators and can be used across different CTs. We do notice a deterioration of inter-annotator agreement in some metrics, but these can be partly explained by the underlying differences in terms of the ratios of conspiracy content and relatedness. However, this needs deeper probing. For the future, we are considering an extension of the CT annotation to include a concept similar to CThinking by Introne et al. (2020), to better handle documents such as the one in Figure 3. More generally, we plan to use the annotated texts for creating a classifier to detect CT language across different CTs. The annotations will be publicly available at <https://github.com/zytian9/locoAnnotations>.

10. Acknowledgments

This work is based on research supported by US National Science Foundation (NSF) Grants #2123635 and #2123618.

11. Bibliographical References

Banas, J. and Miller, G. (2013). Inducing resistance to conspiracy theory propaganda: Testing inoculation and metainoculation strategies. *Human Communication Research*, 39(2):184–207.

- Barkum, M. (2013). *A Culture of Conspiracy*. University of California Press.
- Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009). The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226.
- Castelo, S., Almeida, T., Elghafari, A., Santos, A., Pham, K., Nakamura, E., and Freire, J. (2019). A topic-agnostic approach for identifying fake news pages. In *Companion Proceedings of the 2019 World Wide Web Conference*, pages 975–980.
- Douglas, K. M. and Sutton, R. M. (2018). Why conspiracy theories matter: A social psychological analysis. *European Review of Social Psychology*, 29(1):256–298.
- Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., and Deravi, F. (2019). Understanding conspiracy theories. *Political Psychology*, 40:3–35.
- Fong, A., Roozenbeek, J., Goldwert, D., Rathje, S., and van der Linden, S. (2021). The language of conspiracy: A psychological analysis of speech used by conspiracy theorists and their followers on Twitter. *Group Processes & Intergroup Relations*, 24(4):606–623.
- Introne, J., Korsunskaya, A., Krsova, L., and Zhang, Z. (2020). Mapping the narrative ecosystem of conspiracy theories in online anti-vaccination discussions. In *SMSociety’20: International Conference on Social Media and Society*, pages 184–192.
- Kwon, S., Cha, M., and Jung, K. (2017). Rumor detection over varying time windows. *PloS one*, 12(1):e0168344.
- Landis, J. R. and Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159–268.
- Lopez Long, H., O’Neil, A., and Kübler, S. (2021). On the interaction between annotation quality and classifier performance in abusive language detection. In *Proceedings of the Conference on Recent Advances in NLP (RANLP)*, Online.
- Miani, A., Hills, T., and Bangerter, A. (2021). LOCO: The 88-million-word language of conspiracy corpus. *Behavior Research Methods*.
- Papacharissi, Z. (2016). Affective publics and structures of storytelling: Sentiment, events and mediality. *Information, Communication & Society*, 19(3):307–324.
- Samory, M. and Mitra, T. (2018). ‘The government spies using our webcams’: The language of conspiracy theories in online discussions. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–24.
- Seelig, M., Funchion, J., Trego, R., Kübler, S., Verdear, D., Wuchty, S., Uscinski, J., Klofstad, C., Murthi, M., Premaratne, K., and Diekman, A. (2022). The dark side of Twitter: A framing analysis of conspiratorial rhetoric stoking fear. Presentation at the 72nd Annual ICA Conference: One World, One Network? Online.
- Smith, N. and Graham, T. (2019). Mapping the anti-vaccination movement on Facebook. *Information, Communication & Society*, 22(9):1310–1327.
- Uscinski, J. E., Parent, J., and Torres, B. (2011). Conspiracy theories are for losers. In *APSA 2011 Annual Meeting*.
- Wood, M. J. (2018). Propagating and debunking conspiracy theories on Twitter during the 2015–2016 Zika virus outbreak. *Cyberpsychology, Behavior, and Social Networking*, 21(8):485–490.
- Zhao, X., Liu, J. S., and Deng, K. (2013). Assumptions behind intercoder reliability indices. *Annals of the International Communication Association*, 36(1):419–480.

12. Language Resource References

- Miani, A. et al. (2021). *Language Of Conspiracy Corpus (LOCO)*. Available from <https://osf.io/snpcg>.