Do I sound American?

How message attributes of IRA disinformation relate to Twitter engagement

-- Accepted for publication in Computational Communication Research --

Jiyoun Suk

Josephine Lukito

Min-Hsin Su

Sang Jung Kim

Chau Tong

Zhongkai Sun

Prathusha Sarma

* Jiyoun Suk is an Assistant Professor at the Department of Communication at the University of Connecticut. Josephine Lukito is an Assistant Professor at the School of Journalism and Media at The University of Texas at Austin. Min-Hsin Su is a Ph.D. candidate at the School of Journalism and Mass Communication at the University of Wisconsin-Madison. Sang Jung Kim is a Ph.D. candidate at the School of Journalism and Mass Communication at the University of Wisconsin-Madison. Chau Tong is a Postdoctoral Associate at the Department of Communication at Cornell University. Zhongkai Sun is an applied scientist at Amazon Alexa AI. Prathusha Sarma is an ML researcher at Curai.

Acknowledgment: The authors would like to thank Kaiping Chen, Dhavan Shah, and anonymous reviewers for their guidance and comments on this project.

Abstract

Ongoing research into how states coordinate foreign disinformation campaign has raised concerns over social media's influence on democracies. One example is the spread of Russian disinformation in the 2016 U.S. presidential election. Russia's Internet Research Agency (IRA) Twitter accounts have been known to deliver messages with strategic attempts and political goals. We use publicly-available IRA Twitter data created during and after the 2016 US election campaign (2016 and 2017) to examine the nature of strategic message features of foreign-sponsored online disinformation and their social media sharing. We use computational approaches to identify unique syntactic features of online disinformation tweets from IRA compared to American Twitter corpora, reflecting their functional and situational differences. More important, we examine what message features in IRA tweets across syntax, topic, and sentiment were associated with more sharing (retweets). Implications are discussed.

Keywords: Disinformation, Twitter, Internet Research Agency, corpus linguistics, computational social science

Do I sound American? How message attributes of IRA disinformation relate to Twitter engagement

Conspiracy theories, rumors, and the selective presentation of information have always been part of political campaigns and persuasion messages. However, their production within the digital landscape, especially combined with disinformation campaigns, has disrupted the healthy information environment and exacerbated epistemic failure (Bennett & Livingston, 2018).

One such disruptive force is the disinformation campaign by Russia's "troll farm," the Internet Research Agency (IRA). Employees of the troll farm operate hundreds of fake accounts that impersonate authentic U.S. persons or groups across social media. During the 2016 U.S. presidential election, IRA accounts interacted with 677,000 Americans on Twitter (Twitter, 2018), with the intent to "sow discord" among American public (United States v. Internet Research Agency LLC, 2018). To do so, the IRA has utilized influence operations known as disinformation campaigns, an intentionally coordinated strategy to disseminate false information, where false information encompasses a false context, fabricated identity, or imposter content (Fetzer, 2004). The IRA operates different types of accounts (Linville & Warren, 2020) and engages in strategic network positioning, such as by micro-targeting specific communities (Starbird et al., 2019). By blending their activities with those of legitimate users, a disinformation campaign's influence can go beyond one election and create social confusion about what sources of information are authentic.

Extant literature has been motivated by a desire to identify and cluster the profile and activity characteristics of disinformation actors by types, purposes, and strategies (Keller et al., 2020; Alizadeh et al., 2020). However, less attention has been paid to the sociolinguistic aspects of disinformation. As politically-motivated strategic agents, disinformation actors like the IRA

construct messages to maximize user engagement while disguising their identities.

Complementing previous research, we take a sociolinguistic perspective to examine how message construction could help foreign disinformation actors gain traction and build retweets. Focusing on Twitter, where the IRA operated for the longest (Howard et al., 2018), we first (a) document a linguistic profiling of IRA messages as different from non-IRA messages from the U.S. and (b) examine the relationship between IRA message features across linguistic styles (e.g., syntax), subject, and sentiment and user engagement via retweets.

This study makes several important contributions to the literature. First, we introduce sociolinguistic perspectives by studying linguistic features in anti-democratic discourses. We look at how certain functional and situational factors provide important contexts for IRA's language use, making IRA messages systematically different from non-IRA English tweets. In addition, by exploring the sociolinguistic features of popular and unpopular disinformation content, our results contribute to ongoing efforts to detect and combat disinformation that has yet paid much attention to sociolinguistics. Lastly, our study presents interdisciplinary perspectives, drawing theoretical insights from sociolinguistics, persuasion, and strategic communication, along with computational methods. Through highlighting the importance of contextualization of social media analysis and disinformation studies, we aim to advance a research agenda that leverages the strengths of computational methods to deepen our understanding of the *social* nature of the current information disorder.

Disinformation Campaign and Audience Engagement

Disinformation refers to “the distribution, assertion, or dissemination of false, mistaken, or misleading information in an intentional, deliberate, or purposeful effort to mislead, deceive, or confuse” (Fetzer, 2004, p. 231). Disinformation is persuasive and strategically employed to

manipulate the nature, intention, and goal of others. While not every piece of information disseminated by disinformation agents is factually inaccurate (Fallis 2009), the use of fabricated identities still “disinforms” the public and breeds distrusts (Wardle & Derakshan, 2017).

Embedded within a broader political communication system, social media contain structural vulnerabilities. The lack of gatekeeping, coupled with crowd- and algorithm-driven information flows (Bradshaw & Howard, 2018) allowed disinformation operatives to maximize their influence by crafting the “right” messages to the “right” audience, exploiting the already polarized American electorate. In an environment where information sources are often masked and contexts are collapsed (Pearson, 2020), traditional source cues may not act as credibility heuristics, and other message features and language styles can be especially important for increasing perceived authenticity.

One factor contributing to the success of disinformation campaigns is their ability to attract shares (or retweets). The scope of the influence lies in how effectively disinformation actors are able to motivate the audience to retweet their messages and trigger information cascade organically while disguising their identity. Users’ decisions about whether to share certain messages not only convey communicative and political power, but also carry far-reaching implications for disinformation campaigns. Online, the amplification of such messages grants credibility to the false accounts and increases the newsworthiness or popularity of messages created in an inauthentic context (Lukito et al., 2019). More troubling, pro-Russian online disinformation may change individuals’ opinions in support of pro-Russian positions, even when they had been inoculated (Zerback et al., 2020).

Troll messages are “powerful weapon in modern hybrid warfare” (Monakhov, 2020, p. 2). For state-backed anti-democratic agents seeking to shift power dynamics, language can be a

crucial resource for the strategic amplification of messages (Lundberg & Laitinen, 2020). Topics and sentiments that accompany words operate simultaneously in a message. Therefore, we consider how the sociological and linguistic elements in IRA messages—language style, topic, and sentiment—were related to more user engagement through retweets, above and beyond structural factors like follower numbers. In what follows, we first consider the linguistic features of IRA messages that we anticipate reflect their functional goals and situational factors, which are unique from non-IRA, U.S.-based English messages, and associate those features with user engagement. We also discuss how IRA messages with specific topics and targeted sentiments, embedded in linguistic elements, will be related to higher retweets.

What message features are related to the spread of IRA tweets?

Syntax: Under what language structure?

Communication styles are affected by social contexts and audience interactions. The persuasiveness of communication is shaped by the language structure and styles (Giles & Ogay, 2006). As Sornig (1989) said, “it is the *way* things are said (or done), irrespective of the amount of genuine information carried by an utterance” (p. 95).

Undergirding the operations of IRA agents are unique contextual and situational factors: First, they are strategically motivated to impersonate and deceive audiences; second, as they are backed by a foreign state, the IRA utilizes a language variation that is different from the targeted audience (i.e., the U.S. public). Literature suggests that linguistic styles, such as the use of pronouns or the length of messages, reflect authors’ emotions, social and political identity, and communicative motivations (Eberl, 2019; Perloff, 2021). Notably, deceptive messages may be distinct from authentic ones in their linguistic features (Newman et al., 2003). Considering the

functional goals and situational factors, IRA message's language structure likely differs from that of organic messages of non-IRA users.

While prior evidence has noted that Russian troll messages differ lexically from other English-speaking content (Boyd et al., 2018; Eberl, 2019), fewer studies have explored how linguistic manifestations relate to shares. Focusing on three prominent persuasion tactics in political messages—linguistic complexity (Bene, 2017; Heiss et al., 2019), linguistic similarity (Dyagilev & Yom-Tov, 2014), and use of personalization language (Alavidze, 2016; Proctor et al., 2011)—we investigate how these features correlate to the number of retweets among disinformation messages.

Linguistic complexity. Literature suggests that messages with a deceptive motivation tend to be shorter, less cognitively complex, and offer fewer details than truthful messages (Newman et al., 2003; DePaulo et al., 2003) because detailed messages are more cognitively demanding. Much less is known about whether linguistic complexity in political disinformation messages relate to higher engagement. While simpler messages can be rhetorically effective, syntactic complexity may also contribute to message readability and persuasiveness (Lowrey, 1998). On social media, where source cues are less visible, argument quantity and elaboration may function as information shortcuts, increasing the message's perceived strength and acceptance (Petty & Cacioppo, 1984). Supporting this cognitive heuristic perspective, research has found that politicians' longer social media posts, signaling greater deliberation and reasoning, tend to receive greater user engagement (Heiss et al., 2019). Similarly, syntactically complex news headlines gain more popularity on Twitter (Piotrkowicz et al., 2017). Based on the above, IRA tweets with more complex linguistic structures would be perceived as more persuasive and attract more retweets.

Linguistic similarity to American English. Accommodation in communication styles establishes intimacy, promotes social approval, and strengthens social ties (Giles & Ogay, 2006). In the online sphere with reduced social cues, linguistic similarities in lexicon, syntax, and structure can enhance solidarity and reduce social distance between communicators (Scissors et al., 2009). For disinformation campaigns, this means that syntactic similarity serves as heuristics that produce perceptions of familiarity and trustworthiness. The enhanced familiarity creates “an illusion of truth,” making the content more credible, especially in the absence of source-specific cues (Henkel & Mattson, 2011). In contrast, syntactic unfamiliarity can signal low trustworthiness of the interlocuter and decrease processing fluency, lowering engagement (Simmons, 2006).

Given that IRA agents come from a different language group from their target audience (i.e., the U.S. public), it is reasonable to expect that IRA tweets may be syntactically distinct from non-IRA U.S. ones. In such cases, “speaking like others” may be particularly important for IRA messages to gain credibility and popularity. As “speaking like others confirms both a respect for local conventions and communal bonds” (Jamieson, 2020, p. 120), IRA messages with higher syntactic similarity to the U.S. citizens are more likely to be retweeted.

Personal pronoun use. The use of personal pronouns is a critical linguistic element in persuasive political messages, delivering a sense of personalization and shared identity (Alavidze, 2016; Proctor et al., 2011). Using personal pronouns demands an understanding between interlocutors about the self, others, and the polarizing categories of “us vs. them” in the service of the speaker’s goals (Pennebaker, 2011). Given their intent, IRA accounts likely use first-person pronouns (I’s) to signal a false American identity, inviting attention to the self.

In addition, priming group identity may be especially useful given the polarized American political environment wherein group pronouns are not merely categorical references, but also social relationship indicators (Íñigo-Mora, 2004). The presence of group pronouns invokes in-group solidarity, creates personal relevance, and mobilizes reaction, as demonstrated in Donald Trump's strategic use of pronouns to evoke nationalism and mobilize support (Săftoiu & Toader, 2018). It may also generate norms that motivate groups to engage with content to achieve positive intergroup differentiation (Chilton, 2017), particularly on social media where engagement is often motivated by a shared group identity or affiliations (boyd et al., 2010). We therefore expect that IRA tweets with more personalization language markers, as indicated by the use of personal pronouns, will be retweeted more.

Personalization and modal verbs. Personalization can take on a more direct function when utilized in a call-to-action. For example, the popular phrase “[yes,] we can”—consisting of a first-person plural term and a modal verb—can mobilize voters and create a sense of political progress (Bista, 2009). Modal verbs are often used to reflect cultural values (Talmy, 1988). The combination of personal pronouns and modal verbs can therefore be linguistic markers that increase users' collective efficacy (i.e., be positively related to engagement).

In sum, syntactic features are one crucial element to understand the effectiveness or persuasiveness of a message. Given its functional and situational factors (i.e., deceptive motivation and divergent linguistic background), we examine how IRA tweets are distinct from non-IRA messages in the use of these language structures (RQ1) and further look at the relationship with user engagement via retweets (H1).

RQ1: How would IRA tweets display unique linguistic features in terms of (a) the level of syntactic complexity, (b) the use of American English syntax, (c) the use of personalization language?

H1: IRA tweets with (a) higher syntactic complexity, (b) more use of American English syntax, and (c) more personalization language markers (including personalization and modal verbs) will be associated with more retweets than IRA tweets without.

Subjects: What do IRA accounts tweet about?

In addition to language structure, it is important to examine what topics were discussed by the IRA. Generally, people's motivations to retweet are tied to building a new community around content that a user believes is worth sharing (boyd et al., 2010). Evidence suggests that tweets with newsworthy information are widely retweeted, including content about policy, social actors, or politics (Keib et al., 2018) or messages with practical or public value (Berger & Milkman, 2013), suggesting that online audiences may prioritize topics of high informational value, similar to the news selection process (Shoemaker & Cohen, 2006). That is, people retweet information and content they anticipate is of interest to their followers (Rudat & Buder, 2015).

Besides informational value, messages about controversial issues or social conflicts can also trigger attention and sharing. The focus on conflict may work in tandem with IRA's attempts to target specific communities. For example, IRA accounts exploited divisive topics such as police shooting, Islam and war, and race and religious identities (Badawy et al., 2019; Ghanem et al., 2019) which align with politically contentious agendas in the U.S. IRA accounts also impersonated local news, using journalistic tones, to focus predominantly on controversial and contentious topics such as gun violence and immigration (Bastos & Farkas, 2019). In light of the IRA accounts' tendency to capitalize on contentious U.S. political issues, and in connection with the identity-based news sharing patterns on social media (Marwick, 2018), we propose the following hypotheses:

H2: IRA tweets with news and/or information only will be associated with more retweets than IRA tweets without.

H3: IRA tweets on controversial political issues will be associated with more retweets than IRA tweets that do not refer to such issues.

Sentiments: How do they express it?

Beyond syntax and topics, effective communication messages convey social meanings. One way to deliver meanings is by expressing sentiment. Emotional appeals are part of effective persuasive tactics and are frequently employed in propaganda messages (Wardle & Derakshan, 2017). Evoking emotions can deceive and mislead, as strong sentiment can strengthen a message's persuasiveness, enhance attention and involvement, and dramatize and personalize political causes (Berger & Milkman, 2013).

As the affective intelligence theory posits, political messages with heightened emotions can mobilize engagement (Marcus et al., 2000). Evidence shows that politicians' social media posts with a negative tone were shared more as followers were mobilized to express and perform their political self (Bene, 2017; Heiss et al., 2019). Positive sentiment can also attract engagement and continued support from followers (Eytan et al., 2011). Though specific mechanisms through which sentiment mobilizes engagement may vary, there is a scholarly consensus that messages with heightened sentiment are more likely to gain engagement than those without. We therefore expect, given that the IRA operated during a politically contentious period, that IRA messages with heightened sentiment—positivity, negativity, or a presence of both—will be retweeted more.

Given the strategic nature of disinformation operation, IRA crafts messages to target specific audiences (Wardle & Derakshan, 2017). As such, IRA tweets likely directed their sentiment along party lines, especially focusing on the two candidates of the 2016 U.S. election (Hillary Clinton and Donald Trump) and their partisan supporters, in line with the IRA's goals to demotivate support for Clinton and promote support for Trump (Starbird et al., 2019).

There are reasons to expect that IRA tweets with sentiment toward different targets (Trump and Clinton) will be related to varying levels of retweets. At a personal level, conservatives are more vulnerable to the reception of disinformation messages than liberals (Hjorth & Adler-Nissen, 2019). Conservatives possess greater motivations for identity confirmation and cognitive stability (Boutyline & Willer, 2017), which can lead conservatives to share IRA messages with partisan cues and heightened identities. At an online network level, conservative networks are more fragmented and cloistered than their liberal counterparts (Faris et al., 2017), which can help disseminate IRA messages within a homogenous circuit. Coupled with prior evidence showing that IRA communication targeting conservatives outnumbered those against liberals (Howard et al., 2018; Linvill & Warren, 2020), we expect that the level of engagement of IRA tweets with different target groups will not be uniform. However, given the dearth of evidence on how specific combination of target *and* expressed sentiment will be related to engagement, we propose the following:

H4: IRA tweets that expressed sentiment, either (a) positive, (b) negative, or (c) both, will be associated with more retweets compared to tweets without such sentiments.

RQ2: How will IRA tweets with positive or negative sentiments against (a) Trump and (b) Clinton be associated with more retweets?

Lastly, we also examine how three dimensions of text—syntax, topic, and sentiment—work together to increase a message’s retweet-ability. Notably, the role of message features in communication differs across contexts (Newman et al., 2003), emphasizing an interrelationship between the style of language, the content of communication, and its emotional component. Similarly, we expect that the interaction of message features can relate to user engagement (e.g., retweets). While there is initial evidence for potential interactive effects in the domains of

customer engagement (e.g., Davis et al., 2019), little is known about strategic political communication like online disinformation. Therefore, we ask the following:

RQ3: How will syntactic features in IRA tweets interact with tweet subjects and sentiments to be associated with more retweets?

Method

Russian IRA Twitter data

In October 2018, Twitter released a full corpus of IRA tweets, including all accounts, related content, and engagement metrics, including the number of retweets. We focus on tweets that were created during and after the 2016 U.S. presidential election period, from January 1, 2016 to September 30, 2017, when the IRA activity was disclosed. To analyze the written text, we retained the original English tweets from IRA accounts, resulting in a dataset with 802,618 unique tweets posted by 1,049 unique IRA accounts. In the dataset, Twitter provided the number of retweets that IRA tweets received, *removing* the counts from suspended accounts, which excludes retweet behaviors from other IRA accounts. In our dataset, 38.6% of tweets ($N = 310,040$) received a minimum of one retweet; the maximum number of retweets was 121,190 (for more, see Supplementary Materials A.1.¹).

US Twitter data

To identify the systematic linguistic differences of IRA tweets (RQ1), we built a corpus of US-based tweets using the Twitter API V2. We first made sure that accounts from both IRA corpus and comparison corpus are comparable in terms of such characteristics as the number of accounts, verification status, and the number of followers. We sampled 1,049 accounts (the same number as the IRA accounts in the dataset) that are non-verified, based on the U.S., and used

¹ Supplementary Materials and R codes are available at <https://doi.org/10.17605/OSF.IO/6B4N5>.

English language, using keywords about U.S. politics and election²; in other words, the comparison corpus consists of English-speaking ordinary Twitter users who talked at least once about politics from January 1, 2016 to September 30, 2017 (the same time span as the IRA corpus). Of the tweets posted by these accounts, we took a sample to ensure a comparable distribution of tweets across the IRA and comparison corpus, resulting in a total number of 808,089 tweets (see Supplementary Materials A.2. for details).

Variable Construction

Syntax variables. The syntactic complexity of IRA tweets was measured by (1) the average dependency length and (2) counts of syntactic structures that increase cognitive processing: appositional modifiers, clausal complements, open clausal complements, and nested prepositions. This multidimensional approach is consistent with previous linguistics research (Lu, 2017), which has operationalized syntactic complexity in two ways: by dependency length (Temperley, 2007) and by identifying uses of complex syntactic units, the latter of which tends to be language-specific (Kuiken et al., 2019).

Then, we looked at six syntactic structures that are different between English and Russian languages: subject-verb structure (Slobin, 1969), copula verb conjugation (e.g., “to be”; Unlu & Hatipoglu, 2012), articles (e.g., “the”; Ionin & Wexler, 2002), indefinite pronouns (e.g., “all”), auxiliary verbs (e.g., “be”, “do”; Ionin & Wexler, 2002), and gerunds (e.g., verb-ing; Pazelskaya, 2012). To compare the similarity to American English syntax, we averaged the degree of correspondence to these six American-syntax structures for each IRA tweet. For personal

² The keywords we used include the following: “Clinton,” “Donald Trump,” “election,” “MAGA,” “syria,” “terrorism,” “terrorist,” “Ukraine,” “BlackLivesMatter,” “policebrutality,” “racism,” “lgbt,” “gunrights,” “guncontrol,” “refugee,” “immigration,” “muslim.” We relied on these keywords as political issues that were prominent during and after the 2016 US election and therefore largely targeted by the IRA (Freelon & Lokot, 2019; Freelon et al., 2020).

pronouns, we counted first-person singular, first-person plural, second-person, and third-person plural per tweet (see Supplementary Materials B for the selection of syntax variables).

Subject and sentiment variables. We used a supervised machine learning technique to classify tweets into topic and sentiment variables. Four trained graduate students coded the following variables, all of which obtained the satisfactory level ($>.70$) of Fleiss' Kappa (FK): (1) whether the tweet only delivered factual information or not ($FK = .71$); (2) whether the tweet was about important/controversial political topics in the U.S., such as racial issues ($FK = .78$), immigration ($FK = .85$), terrorism ($FK = .88$), foreign/international policy ($FK = .83$), or not; (3) whether the tweet contained positive, ($FK = .73$) or negative sentiment ($FK = .90$), or not; (4) if so, whether it was toward Donald Trump ($FK = .75$), Hillary Clinton ($FK = .75$) or not (see Supplementary Materials C for a detailed coding scheme).

Using the 4000 labeled tweets, we applied a supervised machine learning method to predict unlabeled tweets' categories (see Figure 1). The supervised machine learning includes two steps: first, finding an appropriate method to convert a short text (tweet) to a numerical vector; second, building high-quality models for each category. In our case, we chose InferSent Sentence Encoder (Conneau et al., 2017), which is a sentence encoder model to represent semantic information of English sentences, to convert tweets to vectors. InferSent is trained on natural language inference tasks so that it can learn universal English sentence representations, and recent studies show that this pre-trained model generalizes well on different tasks (Conneau & Kiela, 2018). We applied InferSent model to our dataset which produced a meaningful numerical vector with a dimension of 4,096, that represents every tweet.

Figure 1.

Supervised Machine Learning Process

We used logistic regression as our basic classification method and trained different logistic regression models for each of our variables. For each category, a Grid-Search with 5-folds cross-validation was used on the training set to select the best hyper-parameters (l1 or l2 penalty, regularization strength, and class weight) for these categories' logistic regression model. As some categories were extremely skewed in distribution, we applied an oversampling method "SMOTE (Synthetic Minority Over-Sampling Technique)" to the training process (Chawla et al., 2002), which allowed us to synthesize training data for the smaller category. For our final models, precision scores ranged from .81 to .99 and F1 scores from .53 to .80 (see Supplementary Materials D for details).

Analysis Strategy

To answer RQ1, we employed a register analysis. This is a corpus linguistics method that compares the use of pervasive syntactic features in multiple corpora (e.g., Staples et al., 2018). Typically, a register analysis is conducted by comparing two corpora of different registers to identify whether syntactic features are used more in one corpus compared to the other; for this reason, chi-square tests are common (Freddi, 2005; Crawford & Csomay, 2015).

To conduct this analysis, we first tokenized both Twitter corpora using an R wrapper for the Python library *spaCy*, then employed register analysis by conducting chi-square tests, a popular approach to comparing two datasets in corpus linguistics (e.g., Liu & Myers, 2020; Degaetano-Ortlieb et al., 2012). We then used multilevel negative binomial modeling via the R package *glmmTMB* to answer the hypotheses and RQs, given the non-independence and nested structure (tweets within accounts) of our dataset. As our dependent variable is the number of retweets (a count variable) and its distribution is highly skewed (positively), a multilevel negative binomial analysis was theoretically and statistically appropriate.

Results

RQ1 asks how IRA tweets would display distinct linguistic characteristics compared to US-based tweets. The comparison process serves a dual purpose of not only documenting syntactic differences in IRA tweets but also guiding a syntax variable construction for the subsequent predictive analysis.

First, we tested the difference in syntactic complexity. Results show that IRA corpus generally had shorter sentences and simpler syntax structures. For example, the average dependency length of IRA tweets ($M = 2.27$) was smaller than the US ones ($M = 3.49$), $t(1, 1377740) = 398.66, p < .001$. As Table 1 illustrates, the IRA corpus also had fewer counts with complex syntax structures. For example, tweets in the IRA corpus used fewer modifier, fewer clausal complements, and fewer open clausal complements, and fewer nested propositions. Consistent with prior evidence on deceptive messages, IRA tweets were also less likely to use cognitively complicated and longer structures.

For American-English syntax similarity, we looked at six language features that are mainly different between English and Russian languages. Our results show that the IRA corpus

had less correspondence to American-English syntax features (see Table 1). Overall, tweets in the IRA dataset had more tweets with incorrect subject-verb structure, including misplaced verbs-before-subjects and objects-before-subject structure.³ IRA also had more incorrect copula-verb conjugations, fewer tweets with articles (e.g., “a”, “the”), fewer indefinite pronouns (e.g., “all”, “any”), auxiliary verbs (e.g., “be”, “do”, “have”), and gerunds (e.g., verb-ing; see Supplementary Materials B for examples). As expected, IRA agents which employ a Russian native, English-as-a-second-language variation (that differs from its target audience), showed fewer correspondences to the syntactic features of the American comparison corpus.

Finally, we looked at personal pronouns. Generally, we find that American tweets used more personal pronouns, including first-person singular (“I”), first-person plural (“we”), second-person (“you”), third-person singular (“s/he”), and third-person plural (“they”). Overall, the effect sizes of the relationships (see Cramer’s Vs in Table 1) suggest that the magnitude of associations is low to moderate. However, given that the impact of subtle linguistic differences can be substantial (Pennebaker, 2011), even small differences would be meaningful.

Table 1.

Linguistic Differences Between IRA Tweets and US Tweets

	# in IRA tweets	# in US tweets	Chi-Square	Cramer’s V
Syntactic complexity				
Appositional Modifier	173,008	177,399	38,317**	0.153
Clausal Complements	135,715	195,217	93,084**	0.238
Open Clausal Complements	83,345	103,507	39,809**	0.156
Nested Prepositions	229,764	243,675	72,206**	0.210

³ For incorrect object-verb inversion, another type of subject-verb-object error, the American corpus had more tweets with such cases ($\chi^2 = 25.69, p < .001$), likely due to incomplete sentence structures in tweets; however, this was the least frequent word-order inversion.

American English-syntax similarity				
Incorrect S-V Inversion				
Verb before Subject	26,584	21,032	2,476**	0.038
Object before Verb	9,047	9,898	2,488**	0.390
Object before Subject	1,344	236	168**	0.010
Incorrect Copula Verb Conjugation	502	113	577.34**	0.019
Use of Article	167,585	398,844	154,238**	0.307
Use of Indefinite Pronouns	7,423	21,560	1,366.61**	0.028
Use of Auxiliary Verbs	246729	347640	145,863**	0.298
Use of Gerunds	186,368	277,274	47,933**	0.171
Personal Pronoun				
First-person Singular	5,499	8,970	3,916.24**	0.048
First-person Plural	36,405	129,855	80,660**	0.222
Second-person	35,228	89,945	57,776**	0.188
Third-person Singular	30,995	79,101	28,918**	0.132
Third-person Plural	19,034	51,004	33,873**	0.143

Note. For Verb before Subject structure, grammatically correct inverted sentences, declarative sentences, passive voice, and questions were excluded.

* $p < .01$, ** $p < .001$

Multilevel modeling

For our main analysis—regarding how features across syntax, topic, and sentiment in IRA tweets are associated with retweets—we conducted random slope multilevel modeling. The user handle was set as the group level and the account life span (i.e., the timespan that the account has been active before its disclosure) as the random slope. For syntax variables, we included 1) the

average dependency length of a tweet as a syntactic complexity,⁴ 2) the degree of correspondence to American-syntax similarity, and 3) counts of personal pronouns. Subject and sentiment variables were entered, along with several controls, including account life span, number of followers, number of hashtags, number of mentions, presence of visuals (videos, images, or animated GIFs), presence of URLs, and hour/day of posting (see Supplementary Materials E for details).

Table 2*Multilevel Modeling Predicting Retweet Counts of IRA Tweets*

	B	SE	IRR
Fixed Parts			
(Intercept)	-4.322**	.242	0.013
Subject			
News/Information (1) vs Opinion (0)	0.197**	.006	1.218
Political issues	0.126**	.006	1.133
Sentiment			
Positive	-0.044**	.009	0.957
Negative	0.188**	.007	1.206
Mixed (Both positive and negative)	0.202**	.018	1.223
Target of Sentiment			
Positive toward Trump	0.144**	.020	1.155
Negative toward Trump	-0.007	.038	1.008
Positive toward Clinton	-0.077*	.027	0.926
Negative toward Clinton	0.167**	.010	1.181
Syntax			
Sentence complexity (Dependency length)	0.006**	.001	1.006
American English-syntax similarity	0.064**	.004	1.066
Personal pronoun use	-0.003	.006	1.003
Personalization and modal verbs	0.092**	.019	1.096
Control			
Account life span	-0.004**	.001	0.996
Number of followers	0.000**	.000	1.000
Visuals	0.997**	.007	2.710
Number of URLs	-0.207**	.008	0.813
Day of posting (weekday vs weekend)	-0.109**	.005	0.897
Hour of posting	-0.000**	.000	1.000
Number of hashtags	0.005**	.003	1.054

⁴ We used the number of dependencies as an indicator of sentence complexity in the multilevel modeling, due to a high correlation between the number of complex structures and dependencies.

Number of mentions	0.027**	.002	1.027
Random Parts			
σ^2		3.797	
$\tau_{00,user\ handle}$		22.790	
$\tau_{11,account\ life\ span}$		0.0004	
Observations (Tweets)		802618	
Observations (User handles)		1049	
Deviance		2416330	
* $p < .01$, ** $p < .001$			

H1 proposed that IRA tweets with (a) higher syntactic complexity, (b) more use of American English syntax, and (c) more personalization language markers (including with modal verbs) were associated with more retweets than IRA tweets without. Our results show that IRA tweets with more complex sentences had 1.006 times the rate of retweets than tweets without, and IRA tweets with more similarity to American English syntax had 1.066 times the rate of retweets than those without (see Table 1). The personal pronouns use was not a significant factor predicting more retweets, but IRA tweets with personal pronoun-modal verb combinations had 1.096 times the rate of retweets than those without. This supports H1(a)(b), but not H1(c).

We also hypothesized that IRA tweets with information-only would be associated with more retweets (H2) and IRA tweets on controversial political issues would be associated with more retweets (H3). Our findings show that IRA tweets delivering information/news only had 1.218 times the rate of retweets than those without. IRA tweets on contentious political issues also had 1.133 times incident rates of retweets, supporting H2 and H3.

H4 proposed that IRA tweets that expressed sentiment—(a) positive, (b) negative, or (c) both—would be associated with more retweets compared to tweets without such sentiment. Our analysis shows that negative tweets had 1.206 times the rate of retweets than those without negative sentiment, whereas positive tweets had 0.957 times the rate of retweets than those without such sentiment, supporting H4(b) but not H4(a). Tweets with both sentiments (positive

and negative) had 1.223 times the rate of retweets than those without, supporting H4(c). RQ2 asked how IRA tweets with positive or negative sentiment against (a) Trump and (b) Clinton would be associated with retweets. IRA tweets with negative sentiment toward Clinton had 1.181 times the rate of retweets, while for IRA tweets about Trump, those with positive sentiment had 1.155 times the rate of retweets than those without. IRA tweets with positive sentiment about Clinton had 0.926 times the rate of retweets. Taken together, IRA tweets with the presence of negativity had significantly higher expected retweet rates, and it was more pronounced when the conveyed negativity was directed toward Clinton.

RQ3 asked about the interaction between syntactic, subject, and sentiment features of IRA tweets. Our findings show that negative tweets with more complexity predicted higher expected rates of retweets ($IRR = 1.010, p < .001$). Figure 2 shows that IRA tweets with negative sentiment were more likely to receive more retweets when associated with higher dependency structures, indicating that long, articulated IRA tweets with negative sentiment garnered more retweets. Interestingly, call-to-actions (measured as a combination of personal pronouns and modal verbs) with higher dependency structures were associated with fewer retweets ($IRR = 0.984, p < .01$). In other words, when IRA tweets had call-to-action markers (combination of personal pronouns and modal verbs), the number of retweets was higher when they were written in shorter, less complicated structures (full tables are in Supplementary Materials F).

Figure 2

Interaction Plots Predicting Retweet Counts of IRA Tweets

(a) Predicted incidents for retweet counts from negative sentiment and sentence complexity

Note. Shades are 95% CI levels.

(b) Predicted incidents for retweet counts from call-to-action marker and sentence complexity

Note. Shades are 95% CI levels.

Discussion

Using a combination of computational approaches, this study examined the nature of foreign-sponsored online disinformation messages and explored how syntax, topic, and sentiment features were related to Twitter sharing. First, by documenting the syntactic features in IRA tweets that differed from authentic U.S.-based users, we find that IRA tweets were syntactically simpler and demonstrated features distinct from tweets by authentic users. This may reflect their deceptive intents (Newmann et al., 2003) and the challenges of producing disinformation for a different language group (Giles & Ogay, 2006).

Importantly, our findings provide insights into how sociolinguistic features in IRA messages were related to their likelihood of being retweeted. Supporting prior evidence on how linguistic similarity builds higher trust (Scissors et al., 2009), IRA tweets with higher similarity to American English-syntax structures were retweeted more. Those with more complicated structures tended to receive more retweets, indicating the potential role of argument length and linguistic complexity as persuasion heuristics in source-blind environments like social media (Pearson, 2020). Since syntactic complexity is often accompanied by reasoning and elaboration, well-reasoned messages likely evoked favorable feelings and appeared more persuasive (Heiss et al., 2019). It is important to note that while IRA tweets overall had simpler language structures and deviated from standard American-English syntax, those that demonstrated syntactic complexity or "sounded more like" authentic US Twitter users tended to attract more retweets, enabling them to blend their activities with those of authentic users. This indicates the important role of syntactic cues in message sharing and highlights the difficulty of distinguishing disinformation messages from authentic ones, especially those capable of gaining traction.

For research on disinformation campaigns, our study highlights potential interactions among sociolinguistic features. Specifically, messages expressing negative sentiment were more powerful when accompanied by elaborated content, whereas pronoun-modal combinations received more retweets when conveyed through simpler languages. Well-articulated IRA tweets emphasizing negative sentiment likely enhanced the credibility and stimulated sentiment heuristics, producing more engagement. On the contrary, mobilizing tweets attracted more retweets in shorter and simpler language structures. Lastly, while personal pronouns alone did not elicit more retweets, their use with mobilization cues were significantly related to retweets, suggesting that the effect of pronouns may be moderated when associated with certain pragmatic functions. These results confirm that the effect of communication styles is not constant but contingent on content context (Newmann et al., 2003).

Our findings also aligned with current scholarship showing that divisive topics (e.g., issues with racial or ideological implications) were associated with more retweets. IRA tweets with only news or information were retweeted more, resonating with prior findings of informational value and virality (e.g., Keib et al., 2018). It is noteworthy that news/information delivered by the IRA is likely a subjective selection tailored to a target audience, some of which may be factually incorrect. Troublingly, several IRA accounts presented as local news (Farkas & Bastos, 2018), which may create a false sense of credibility and induce confusion about information trustworthiness.

Regarding sentiment and persuasion, our findings join work that considers the *target* of sentiment, particularly in an era of political microtargeting (e.g., Park et al., 2021). Specifically, we show that not *all* negativity has "news value" worth sharing (Mueller & Saeltzer, 2020). In our case of IRA operations, while tweets with negative sentiment toward Clinton were associated

with more retweets, those with positive sentiment towards Trump also garnered more retweets compared to neutral ones, corresponding to IRA's goal to amplify pro-Trump rhetoric.

On one hand, this calls for more attention to the highly contingent role of emotion in persuasion research, which traditionally did not differentiate the target of sentiment. On the other hand, these findings also inform how IRA activities could leverage asymmetrical American partisan ecology and polarized partisan behavior. Research shows that disinformation operations tend to promote conformity and normative influence (Wardle & Derakhshan, 2017). To the extent that pro-Trump and anti-Clinton IRA messages motivated subconscious "moral systems" and salient rhetoric among conservatives (Lakoff, 2010), sharing such content may serve as useful tactics for identity performance (Wardle & Derakhshan, 2017). Moreover, through publicly expressing conformity to group norms, the act of retweeting contributes to a strengthened partisan self (Klein et al., 2007). This tendency is likely more pronounced in cloistered conservative communities than liberal ones (Faris et al., 2017). Given that the Twitter population has a liberal bias (Wojcik & Hughes, 2019), it is striking that the IRA disinformation campaign could amplify conservative-leaning messages compared to liberal ones.

While focusing on the IRA disinformation campaign on Twitter, our study has broader implications for contemporary challenges surrounding the strategic disguise of identity and intention in computer-mediated contexts. Many IRA activities, as outlined above, capitalized on existing strategies of internet trolling, impersonation, and astroturfing that have been widely applied across contexts, including social activism initiatives and public relations management. For instance, the IRA frequently "operate as agents of chaos" by exploiting hot-button issues and provoking others emotionally, demonstrating typical trolling behaviors (Hardaker, 2010); or they may exploit bottom-up grassroots efforts to create a false impression of widespread support for a

cause that is in actuality not there, benefiting from common astroturfing tactics (Lock et al., 2016). Apparently, as with foreign disinformation operations, these practices have led to shared concerns about distorted perceptions of social reality, undermining trust and transparency, and disrupting the integrity of digital media ecology (Zerback & Töpfl, 2021). We believe our approach to identify sociolinguistic limitations and strategic and deceptive properties can inform broader research agendas on inauthentic agents and their hidden persuasive intents.

This study contains several limitations. First, we did not consider the nature of those retweeting IRA tweets. While the current dataset by Twitter does not include “retweeter” information, given the evidence on online political homophily and IRA’s strategy (Starbird et al., 2019), we expect IRA tweets were circulated differently along party lines. Investigating how these IRA tweets reached different audiences would be an important avenue for further investigation. Secondly, we focused on the IRA’s original messages only. While retweeting messages written by political figures or pundits can also be an important strategy and, relatedly, while we cannot rule out the possibility of IRA tweets being generated by automated agents, our study sheds light on the nature and reach of human-produced messages constructed by foreign-sponsored online disinformation agents. Lastly, we acknowledge that the political issues in our study are only a subset of issues that were largely covered during the 2016 U.S. election cycle. Besides practical infeasibility to cover all potentially relevant issues, it is noted that our study chose to focus on issues that achieved a satisfactory level of machine learning performance in order not to undermine the validity of our results.

In conclusion, our findings add to ongoing scholarly conversations about disinformation by integrating novel theoretical and methodological insights. By uncovering syntactic patterns reflecting IRA’s situational motivations, we highlight sociolinguistic characteristics inherent to

anti-democratic discourses. Given the global reach of disinformation campaigns, we hope our approach contributes to future research on troll messages in various linguistic and sociocultural settings and intervention efforts against state-backed disinformation spread.

Furthermore, our findings deepen the understanding of how a highly coordinated state-backed disinformation campaign operated in fueling the spread of messages with certain syntactic, topical, and sentiment features, leveraging the polarized and fragmented American online ecology. We underscore the importance of contextualization reflected in such anti-democratic discourses and their dissemination. As more retweeted IRA messages unwittingly lend credibility and trust to the false account, retweetability may help inauthentic narratives circumvent the newsroom gatekeeping process (Lukito et al., 2019). This is a vicious cycle of how the disinformation campaign can exacerbate the disruption of democratic discursive norms, undermine legitimacy in platforms and institutions, and lower trust in other social groups, facilitating alternative information systems that endanger “normal democratic order” (Bennett & Livingston, 2018).

While the current study focused on Russian IRA’s activities targeting the U.S., such political meddling by disinformation agents is not a geographically-bounded phenomenon, and the Russian IRA is not the only information operation attempting to interrupt foreign politics. Foreign disinformation campaigns also produce conspiratorial and partisan messages on topics like the COVID-19 outbreak and elections in Latin America and North Africa (e.g., Moreno, 2020). We hope future studies build on our research to investigate the nature of disinformation messages and their disruptions in online information flows in comparative contexts.

References

- Alavidze, M. (2016). The use of pronouns in political discourse. *International Journal of Arts & Sciences*, 9(4), 349-355.
- Alizadeh, M., Shapiro, J. N., Buntain, C., & Tucker, J. A. (2020). Content-based features predict social media influence operations. *Science Advances*, 6(30), 1-13. <https://doi.org/10.1126/sciadv.abb5824>
- Badawy, A., Addawood, A., Lerman, K., & Ferrara, E. (2019). Characterizing the 2016 Russian IRA influence campaign. *Social Network Analysis and Mining*, 9(1), 31-43. <https://doi.org/10.1007/s13278-019-0578-6>
- Bastos, M., & Farkas, J. (2019). “Donald Trump is my president!”: The Internet Research Agency propaganda machine. *Social Media+ Society*, 5(3). 1-13. <https://doi.org/10.1177/2056305119865466>
- Bene, M. (2017). Go viral on the Facebook! Interactions between candidates and followers on Facebook during the Hungarian general election campaign of 2014. *Information, Communication & Society*, 20(4), 513–529. <https://doi.org/10.1080/1369118X.2016.1198411>
- Bennett, W., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2), 122–139. <https://doi.org/10.1177/0267323118760317>
- Berger, J., & Milkman, K. L. (2013). Emotion and virality: What makes online content go viral?. *GfK Marketing Intelligence Review*, 5(1), 18-23. <https://doi.org/10.2478/gfkmir-2014-0022>

- Bista, K. (2009). On “Yes, We Can”: Linguistics power and possibility. *Journal of English for Specific Purpose*, 3(24), 34-50.
- Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology*, 38(3), 551–569.
<https://doi.org/10.1111/pops.12337>
- boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *2010 43rd Hawaii International Conference on System Sciences* (pp. 1-10). <https://doi.org/10.1109/HICSS.2010.412>
- Boyd, R., Spangher, A., Fourney, A., Nushi, B., Ranade, G., Pennebaker, J., & Horvitz, E. (2018). Characterizing the Internet Research Agency’s social media operations during the 2016 U.S. presidential election using linguistic analyses. *PsyArXiv*. Retrieved from http://test.adamfourney.com/papers/boyd_psyarxiv2018.pdf
- Bradshaw, S., & Howard, P. (2018). Challenging truth and trust: A global inventory of organized social media manipulation. *The Computational Propaganda Project*. Retrieved from <http://comprop.oii.ox.ac.uk/research/cybertroops2018/>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(2002), 321-357. <https://doi.org/10.1613/jair.953>
- Chilton, P. (2017). “The people” in populist discourse: Using neuro-cognitive linguistics to understand political meanings. *Journal of Language and Politics*, 16(4), 582–594.
<https://doi.org/10.1075/jlp.17031.chi>
- Conneau, A., & Kiela, D. (2018). Senteval: An evaluation toolkit for universal sentence representations. *arXiv*. <https://doi.org/1803.05449>.

- Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised learning of universal sentence representations from natural language inference data. *arXiv*.
<https://doi.org/1705.02364>
- Crawford, W., & Csomay, E. (2015). *Doing corpus linguistics*. Routledge.
- Davis, S. Horváth, C., Gretry, A., & Belei, N. (2019). Say what? How the interplay of tweet readability and brand hedonism affects consumer engagement. *Journal of Business Research*, 100(2019), 150–164. <https://doi.org/10.1016/j.jbusres.2019.01.071>
- Degaetano-Ortlieb, S., Lapshinova-Koltunski, E., & Teich, E. (2012, May). Feature discovery for diachronic register analysis: A semi-automatic approach. In *LREC* (pp. 2786-2790).
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 74–118. <https://doi.org/10.1037/0033-2909.129.1.74>
- Dyagilev, K., & Yom-Tov, E. (2014). Linguistic factors associated with propagation of political opinions in twitter. *Social Science Computer Review*, 32(2), 195–204.
<https://doi.org/10.1177/0894439313506850>
- Eberl, M. (2019). Trollspeak: Who do Russian trolls tweet like? *Journal LIPP*, 7(2019), 49–59.
<https://doi.org/10.5282/journallipp/4880>
- Eytan, T., Benabio, J., Golla, V., Parikh, R., & Stein, S. (2011). Social media and the health system. *The Permanente Journal*, 15(1), 71-74.
- Fallis, D. (2009). A conceptual analysis of disinformation. In *Proceedings of iConference* (pp. 1-8).
- Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S.*

- Presidential Election*. Retrieved from
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3019414
- Farkas, J., & Bastos, M. (2018, July). IRA propaganda on Twitter: Stoking antagonism and tweeting local news. In *Proceedings of the 9th International Conference on Social Media and Society* (pp. 281-285). <https://doi.org/10.1145/3217804.3217929>
- Freddi, M. (2005). Arguing linguistics: Corpus investigation of one functional variety of academic discourse. *Journal of English for Academic Purposes*, 4(1), 5-26.
<https://doi.org/10.1016/j.jeap.2003.09.002>
- Freelon, D., & Lokot, T. (2020). Russian disinformation campaigns on Twitter target political communities across the spectrum. *Harvard Kennedy School (HKS) Misinformation Review*. <https://doi.org/10.37016/mr-2020-003>
- Freelon, D., Bossetta, M., Wells, C., Lukito, J., Xia, Y., & Adams, K. (2020). Black trolls matter: Racial and ideological asymmetries in social media disinformation. *Social Science Computer Review*, 0894439320914853. <https://doi.org/10.1177/0894439320914853>
- Fetzer, J. (2004). Disinformation: The use of false information. *Minds and Machines*, 14(2), 231–240. <https://doi.org/10.1023/B:MIND.0000021683.28604.5b>
- Ghanem, B., Buscaldi, D., & Rosso, P. (2019). TexTrolls: Identifying Russian trolls on Twitter from a textual perspective. *arXiv*. <https://doi.org/1910.01340>.
- Giles, H. & Ogay, T. (2006). Communication accommodation theory. In B.B. Whaley & W. Samter (Eds.), *Explaining Communication: Contemporary Theories and Exemplars* (pp. 293-310). Mahwah, NJ: Erlbaum.

- Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, 6(2), 215–242.
<https://doi.org/10.1515/jplr.2010.011>
- Heiss, R., Schmuck, D., & Matthes, J. (2019). What drives interaction in political actors' Facebook posts? Profile and content predictors of user engagement and political actors' reactions. *Information, Communication & Society*, 22(10), 1497–1513.
<https://doi.org/10.1080/1369118X.2018.1445273>
- Henkel, L., & Mattson, M. (2011). Reading is believing: The truth effect and source credibility. *Consciousness and Cognition*, 20(4), 1705–1721.
<https://doi.org/10.1016/j.concog.2011.08.018>
- Hjorth, F., & Adler-Nissen, R. (2019). Ideological asymmetry in the reach of pro-Russian digital disinformation to United States audiences. *Journal of Communication*, 69(2), 168–192.
<https://doi.org/10.1093/joc/jqz006>
- Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., & François, C. (2018). The IRA, social media and political polarization in the United States, 2012–2018. *Project on Computational Propaganda*. Retrieved from <https://digitalcommons.unl.edu/senatedocs/1/>
- Íñigo-Mora, I. (2004). On the use of the personal pronoun we in communities. *Journal of Language and Politics*, 3(1), 27–52. <https://doi.org/10.1075/jlp.3.1.05ini>
- Ionin, T., & Wexler, K. (2002). Optional verb raising in L2 acquisition: Evidence from L1-Russian learners of English. In *Proceedings of the GALA 2001 Conference* (pp. 1–9).
- Jamieson, K. (2020). *Cyberwar: How Russian hackers and trolls helped elect a President what we don't, can't, and do know*. Oxford University Press.

- Keib, K., Himelboim, I., & Han, J. (2018). Important tweets matter: Predicting retweets in the #BlackLivesMatter talk on twitter. *Computers in Human Behavior*, 85(2018), 106–115.
<https://doi.org/10.1016/j.chb.2018.03.025>
- Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication*, 37(2), 256-280.
<https://doi.org/10.1080/10584609.2019.1661888>
- Klein, O., Spears, R., & Reicher, S. (2007). Social identity performance: Extending the strategic side of SIDE. *Personality and Social Psychology Review*, 11(1), 28-45.
<https://doi.org/10.1080/10584609.2019.1661888>
- Kuiken, F., Vedder, I., Housen, A., & De Clercq, B. (2019). Variation in syntactic complexity. *International Journal of Applied Linguistics*, 29(2), 161-170.
<https://doi.org/10.1111/ijal.12255>
- Lakoff, G. (2010). *Moral politics: How liberals and conservatives think*. University of Chicago Press.
- Linville, D. L., & Warren, P. L. (2020). Troll factories: Manufacturing specialized disinformation on Twitter. *Political Communication*, 37(4), 447–467.
<https://doi.org/10.1080/10584609.2020.1718257>
- Liu, D., & Myers, D. (2020). The most-common phrasal verbs with their key meanings for spoken and academic written English: A corpus analysis. *Language Teaching Research*, 24(3), 403-424. <https://doi.org/10.1177/1362168818798384>
- Lock, I., Seele, P., & Heath, R. L. (2016). Where grass has no roots: The concept of ‘shared strategic communication’ as an answer to unethical astroturf lobbying. *International*

- Journal of Strategic Communication*, 10(2), 87-100.
<https://doi.org/10.1080/1553118X.2015.1116002>
- Lowrey, T. M. (1998). The effects of syntactic complexity on advertising persuasiveness. *Journal of Consumer Psychology*, 7(2), 187-206. https://doi.org/10.1207/s15327663jcp0702_04
- Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, 34(4), 493-511. <https://doi.org/10.1177/0265532217710675>
- Lukito, J., Suk, J., Zhang, Y., Doroshenko, L., Kim, S. J., Su, M. H., Xia, Y., Freelon, D., & Wells, C. (2020). The wolves in sheep's clothing: How Russia's internet research agency tweets appeared in US news as vox populi. *The International Journal of Press/Politics*, 25(2), 196-216. <https://doi.org/10.1177/1940161219895215>
- Lundberg, J., & Laitinen, M. (2020). Twitter trolls: a linguistic profile of anti-democratic discourse. *Language Sciences*, 79(2020), 1-14. <https://doi.org/10.1016/j.langsci.2019.101268>
- Marcus, G., Neuman, W., & MacKuen, M. (2000). *Affective intelligence and political judgment*. University of Chicago Press.
- Marwick, A. (2018). Why do people share fake news? A sociotechnical model of media effect. *Georgetown Law Technology Review*, 2(2), 474-512.
- Monakhov, S. (2020). Early detection of internet trolls: Introducing an algorithm based on word pairs / single words multiple repetition ratio. *PLOS ONE*, 15(8), e0236832. <https://doi.org/10.1371/journal.pone.0236832>

- Moreno, E. (2020, April 21). Russia, China, Iran using similar themes in disinformation campaigns: report. *The Hill*. Retrieved from <https://thehill.com/policy/international/493851-russia-china-iran-using-similar-themes-in-disinformation-campaigns>
- Mueller, S. D., & Saeltzer, M. (2020). Twitter made me do it! Twitter's tonal platform incentive and its effect on online campaigning. *Information, Communication & Society*, 1-26. <https://doi.org/10.1080/1369118X.2020.1850841>
- Newman, M., Pennebaker, J., Berry, D., & Richards, J. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675. <https://doi.org/10.1177/0146167203029005010>
- Park, S., Strover, S., Choi, J., & Schnell, M. (2021). Mind games: A temporal sentiment analysis of the political messages of the Internet Research Agency on Facebook and Twitter. *New Media & Society*, 14614448211014355. <https://doi.org/10.1177/14614448211014355>
- Pazelskaya, A. (2012). Verbal prefixes and suffixes in nominalization: Grammatical restrictions and corpus data. *Oslo Studies in Language*, 4(1), 245-261. <https://doi.org/10.5617/osla.153>
- Pearson, G. (2020). Sources on social media: Information context collapse and volume of content as predictors of source blindness. *New Media & Society*, 1461444820910505. <https://doi.org/10.1177/1461444820910505>
- Pennebaker, J. (2011). The secret life of pronouns. *New Scientist*, 211(2828), 42-45. [https://doi.org/10.1016/S0262-4079\(11\)62167-2](https://doi.org/10.1016/S0262-4079(11)62167-2)
- Perloff, R. M. (2021). *The dynamics of political communication: Media and politics in a digital age*. Routledge.

- Petty, R., & Cacioppo, J. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, 46(1), 69–81. <https://doi.org/10.1037/0022-3514.46.1.69>
- Piotrkowicz, A., Dimitrova, V., Otterbacher, J., & Markert, K. (2017). The impact of news values and linguistic style on the popularity of headlines on Twitter and Facebook. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 767-774. Retrieved from <https://ojs.aaai.org/index.php/ICWSM/article/view/14979>
- Proctor, K., Lily, I., & Su, W. (2011). The 1st person plural in political discourse—American politicians in interviews and in a debate. *Journal of Pragmatics*, 43(13), 3251–3266. <https://doi.org/10.1016/j.pragma.2011.06.010>
- Rudat, A., & Buder, J. (2015). Making retweeting social: The influence of content and context information on sharing news in Twitter. *Computers in Human Behavior*, 46(2015), 75-84. <https://doi.org/10.1016/j.chb.2015.01.005>
- Săftoiu, R., & Toader, A. (2018). The persuasive use of pronouns in action games of election campaigns. *Language and Dialogue*, 8(1), 21-42. <https://doi.org/10.1075/ld.00003.saf>
- Scissors, L. E., Gill, A. J., Geraghty, K., & Gergle, D. (2009, April). In CMC we trust: The role of similarity. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 527-536). <https://doi.org/10.1145/1518701.1518783>
- Shoemaker, P., & Cohen, A. (2006). *News around the world: Content, practitioners, and the public*. Routledge.
- Simmons, E. (2006). The grammars of reading. *English Journal*, 95(5), 48-52. <https://doi.org/10.2307/30046588>

- Slobin, D. (1969). Universals of grammatical development in children. In G. Flores d'Arcais & W. Levelt (Eds.), *Advances in psycholinguistics* (pp. 174–186). North Holland Publishing.
- Sornig, K. (1989). Some remarks on linguistic strategies of persuasion. In R. Wodak (Ed.), *Language, power and ideology: Studies in political discourse* (pp. 95-114). Benjamins
- Staples, S., Biber, D., & Reppen, R. (2018). Using corpus-based register analysis to explore the authenticity of high-stakes language exams: A register comparison of TOEFL iBT and disciplinary writing tasks. *The Modern Language Journal*, 102(2), 310-332.
<https://doi.org/10.1111/modl.12465>
- Starbird, K., Arif, A., & Wilson, T. (2019). Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1-26. <https://doi.org/10.1145/3359229>
- Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, 12(1), 49-100.
https://doi.org/10.1207/s15516709cog1201_2
- Temperley, D. (2007). Minimization of dependency length in written English. *Cognition*, 105(2), 300-333. <https://doi.org/10.1016/j.cognition.2006.09.011>
- Twitter (2018, January 19). Update on Twitter's review of the 2016 US election. Retrieved from https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html
- Unlu, E., & Hatipoglu, Ç. (2012). The acquisition of the copula be in present simple tense in English by native speakers of Russian. *System*, 40(2), 255-269.
<https://doi.org/10.1016/j.system.2012.04.002>
- Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. *Council of Europe report*, DGI (2017), 9.

Retrieved from <https://firstdraftnews.com/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-de%CC%81sinformation-1.pdf?x29719>.

Wojcik, S., & Hughes, A. (2019, April 24). Sizing up Twitter users. *Pew Research Center*.

Retrieved from <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>

Zerback, T., & Töpfl, F. (2021). Forged examples as disinformation: The biasing effects of political astroturfing comments on public opinion perceptions and how to prevent them. *Political Psychology*. 1-20. <https://doi.org/10.1111/pops.12767>

Zerback, T., Töpfl, F., & Knöpfle, M. (2020). The disconcerting potential of online disinformation: Persuasive effects of astroturfing comments and three strategies for inoculation against them. *New Media & Society*, 23(5), 1080-1098.
<https://doi.org/10.1177/1461444820908530>