



ARL-TR-8904 • FEB 2020



# Analyzing Microblogs: Exploring the Psychology of Political Leanings

by Meaghan Bowman and Michelle Vanni

Approved for public release; distribution is unlimited.

## **NOTICES**

### **Disclaimers**

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.



# Analyzing Microblogs: Exploring the Psychology of Political Leanings

**Meaghan Bowman**

*University of California, Los Angeles, Oak Ridge Associated Universities  
Journeyman Fellow*

**Michelle Vanni**

*Computational and Information Sciences Directorate, CCDC Army Research  
Laboratory*

**REPORT DOCUMENTATION PAGE**

*Form Approved  
OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> February 2020		<b>2. REPORT TYPE</b> Technical Report		<b>3. DATES COVERED (From - To)</b> September 2019–February 2020	
<b>4. TITLE AND SUBTITLE</b> Analyzing Microblogs: Exploring the Psychology of Political Leanings				<b>5a. CONTRACT NUMBER</b> W911NF-20-2-0012	
				<b>5b. GRANT NUMBER</b>	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Meaghan Bowman and Michelle Vanni				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> CCDC Army Research Laboratory ATTN: FCDD-RLC-IT Adelphi, MD 20783-1138				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  ARL-TR-8904	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> Approved for public release; distribution is unlimited.					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> Microblogging has become increasingly popular for commenting on current events, spreading gossip, and encouraging individualism, which favors its low-context communication channel. These social-media (SM) platforms allow users to express opinions while interacting with a wide range of populations. Hashtags allow immediate identification of like-minded individuals worldwide on a vast array of topics. For example, in 2016 the Colombia government held a referendum on a draft peace agreement with the antigovernment militia, Fuerzas Armadas Revolucionarias de Colombia (FARC). Popular approval would provide seats in the legislature for former FARC members and immunity for drug-cartel members. Government polling had indicated strong public support of the referendum—which then failed. In this report, the output of Linguistic Inquiry and Word Count (LIWC), a program that associates psychological meaning with the frequency of use of specific words, may suggest the nature of individuals’ internal states and general sentiments. When applied to groupings of SM posts unified by a hashtag, such information can be helpful to community leaders during periods in which the forming of public opinion happens in parallel with the unfolding of political, economic, or social events. This is especially true when outcomes stand to impact the well-being of the group.					
<b>15. SUBJECT TERMS</b> Linguistic Inquiry and Word Count, LIWC, social media, polarization, microblogs					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>  UU	<b>18. NUMBER OF PAGES</b>  26	<b>19a. NAME OF RESPONSIBLE PERSON</b> Michelle Vanni
<b>a. REPORT</b> Unclassified	<b>b. ABSTRACT</b> Unclassified	<b>c. THIS PAGE</b> Unclassified			<b>19b. TELEPHONE NUMBER (Include area code)</b> (301) 394-0367

## **Contents**

---

<b>List of Figures</b>	<b>iv</b>
<b>List of Tables</b>	<b>iv</b>
<b>1. Introduction</b>	<b>1</b>
<b>2. Background</b>	<b>2</b>
<b>3. Data and Methods</b>	<b>3</b>
3.1 Translating the Microblogs	3
3.2 “Political Leanings” Categories	3
<b>4. Experimental Design</b>	<b>7</b>
<b>5. Results</b>	<b>8</b>
<b>6. Discussion</b>	<b>13</b>
<b>7. Conclusions</b>	<b>14</b>
<b>8. References</b>	<b>15</b>
<b>Appendix. Figures for Cross-Condition Intercategory Analysis</b>	<b>16</b>
<b>List of Symbols, Abbreviations, and Acronyms</b>	<b>19</b>
<b>Distribution List</b>	<b>20</b>

## List of Figures

---

---

Fig. 1	Online LIWC analysis of Google-translated SM postings in Condition A.....	8
Fig. 2	Online LIWC analysis of Google-translated SM postings in Condition B.....	10
Fig. 3	Online LIWC analysis of Google-translated SM postings in Condition C.....	12
Fig. A-1	Graph shows the four most relevant outputs from Linguistic Inquiry and Word Count (LIWC) based on no-advocate tweets compared with Twitter Media Average.....	17
Fig. A-2	Four most relevant LIWC-category outputs for neutral tweets compared with Twitter Media Average.....	17
Fig. A-3	Four most relevant LIWC-category outputs for yes-advocate tweets compared with Twitter Media Average.....	18

## List of Tables

---

---

Table 1	Summary of value differences; blue cells indicate stronger VoteNo signal.....	14
---------	---	----

## 1. Introduction

---

Microblogging is popular for commenting on current events, spreading gossip, and encouraging the individualism that favors low-context channels (Pennebaker and King 1999; Garcia-Gavilanes et al. 2013). These social media (SM) platforms allow users to express opinions and interact with a range of populations, and hashtags allow immediate identification of like-minded individuals on a vast array of topics. For example, in 2016 the Colombia government held a referendum on a draft peace agreement with Fuerzas Armadas Revolucionarias de Colombia (FARC), the antigovernment militia group. Stakes were high as agreement provisions included seats in the legislature for former FARC members and immunity for drug-cartel members. Notably, polling indicated strong public support for the agreement, which, surprisingly, failed on 2 October 2016.

Analysis provided by Linguistic Inquiry and Word Count (LIWC)—a program that calculates the degree to which words in psychologically meaningful categories are used—may suggest the nature of individuals’ internal states and sentiments. When applied to SM posts, such information can be helpful to community leaders during periods when the forming of public opinion happens at the same time as political, economic, or social events are unfolding, especially when outcomes stand to impact community well-being. In this study, we applied the online tools, Google Translate and University of Texas’s LIWC, to a 90-posting sample of a corpus of Colombian Spanish microblogs collected by Notre Dame University (NDU). On translated disjoint sets, identified by hashtag as being authored by advocates of voting “No”, advocates of voting “Yes”, and individuals choosing to refrain from using hashtags, values for LIWC’s Tone attribute showed significant effects. The lexeme “peace”, due to its frequency of use, emerged as relevant.

It is important to note the public voted 50.2% against the agreement and 49.8% in favor. While not a big loss, such overconfidence in the results of a referendum holds lessons in governance due to the possible retaliation that may have resulted. Members of the militant losing side, FARC, could have become hostile and angry; they could have lost trust with the government and sought retribution against the public. To avoid these scenarios, governments may do well to update the way they learn about the mindsets and emotional states of the citizenry, to gain a clear understanding of public sentiment before putting important issues up for referendum. If Colombian government officials had used SM analysis as a tool to better understand the public’s perspective, they could have modified agreements and policies before the vote and avoided the disappointment and confusion of the defeat that resulted.

## 2. Background

---

In 2016, the Colombian government was ready to make peace with the FARC antigovernment terrorist group. Conflict between the Colombia population and the FARC dated back to the 1960s when the FARC started cooperating with the Medellin drug cartel, kidnapping women, and rejecting the legitimate government. On 23 June 2016, the FARC signed a ceasefire agreement with the government, and together they started drafting a peace agreement. The government was sure that putting the draft agreement up for referendum would result in its passage into law. However, on 2 October 2016, the public rejected the referendum despite predictions of its success based on government polling. The peace research community, to include the NDU Peace Studies Institute (PSI), wanted to explore possible causes of—prevent future instances of—similar erroneous predictions.

SM has ushered in an era of unprecedented global information sharing. Many individuals use SM to engage with world news and to gain knowledge of public affairs. It is also a platform where individuals can openly express their opinions, thoughts, and feelings: a feature that motivates social-science researchers to collect and analyze these data on a grand scale (Liang and Dai 2013). The NDU PSI, assisted by the Interdisciplinary Center for Network Science and Applications (iCenSA) of the NDU Department of Computer Science, collected 280,936 Spanish microblogs related to the Colombian peace process. The corpus represents material produced by 34,190 users who posted during the period leading up to the 2016 Peace Referendum (Nigam et al. 2017).<sup>\*</sup> NDU used the full dataset for its polarization analysis and a subset of 2,142 posts, manually annotated for political stance concerning the referendum, based on hashtag, as a ground-truth dataset for training and testing an algorithm for automatic detection.

In this pre-pilot study, we chose from the full dataset a random sample of 30 posts from each subset, along with 30 posts considered neutral, for a total of 90 randomly selected posts. Thus, our categories were a) *VoteYes*, b) *VoteNo*, and c) *Neutral*. It should be noted that because “neutral” posts were categorized based on the absence of a hashtag, positions expressed in them occasionally may, nevertheless, have leaned in one or the other political direction.

Postings were composed in Spanish by Colombian users. We ran them through the online machine translation (MT) engine, Google Translate, and lightly edited the English output for readability, being careful to leave substantive segments intact.

---

<sup>\*</sup> The specific timeframe was the 21-day period of 11 September through 1 October 2016.

We then ran the postedited output through the online version of LIWC, henceforth *Online LIWC*, which only accepts English input.

### 3. Data and Methods

---

#### 3.1 Translating the Microblogs

---

The 90 Spanish postings that were analyzed for this study required translation before Online LIWC could appropriately be applied to process them. In our use of the online translator, we encountered the usual problems inherent to free black-box online text data-processing tools; that is, mismatched encoding schemes, out-of-vocabulary lexical items, and the spurious rendering of names and technical terms.\*

Hashtags, when included in the input to MT, were handled by first removing the hash character from the tag itself.† In a second step, the tag’s words were spaced out to allow the MT system to recognize individual words or phrases and to translate them accordingly. For example, the hashtag #*SíALaPaz* became *Sí a la paz*, meaning “yes to the peace”.

To interpret these postings correctly we examined the MT-produced translations for understandability. Further confounding issues noted in the first paragraph of this section, we observed a lack of fluency and fidelity in the MT output. This we attributed to the facts that a) postings are composed often by individuals whose language use reflects a disregard for the prescriptive grammar of the text on which MT systems train and b) the structure of such short postings is often telegraphic. The politically charged pre-referendum context lent itself to postings made in haste, with accents or letters dropped in the process. For study purposes, it was important to edit output containing easily corrected errors, such as those stemming from accent placement and resulting in drastically altered meanings as with *papa* (potato) being confounded with *papá* (dad) and *sí* (yes) with *si* (if). MT’d postings were thus reviewed, edited, and confirmed as belonging to one of the three referendum-related categories.

#### 3.2 “Political Leanings” Categories

---

We had selected the postings from the *VoteYes*, *VoteNo*, and *Neutral* hashtag-determined categories of NDU’s manually annotated corpus to investigate the possibility that category-specific, language-borne psychological features might

---

\* Time constraints prevented rigorous attention to these crucial linguistic issues, resolution of which we leave for future work in technology optimization for effectiveness.

† Hashtags, considered as meta-information or commentaries on posting content, were included for analysis in two of the experimental conditions.

be distinguishable in the postings of the three groups. We wanted to understand the impact, including potential impact, of category-feature associations that might be discovered. *VoteYes* postings contained hashtags expressing support for the draft agreement. Examples included *#SíALaPaz*, *#El2PorElSí*, *#Sí*, and *#ColombiaVotaSí*. Hashtags on *VoteNo* postings expressed opposition to the draft agreement and included examples such as *#VotoNo*, *#ColombiaVotaNo*, *#NoALasFARC*, and *#voteno*. *Neutral* posts, which, as noted, did not contain a hashtag, often originated from accounts associated with news-media outlets.

In an exploratory trial, we used 30 postings, 10 per category. We then expanded the set to 90 postings, 30 per category, and capped it there due to the time constraints of the study.\*

LIWC starts with a developer-created dictionary that indexes psychological attributes, such as Analytical Thinking, Clout, Authenticity, Emotional Tone, and Positive/Negative Emotion, to individual words identified as expressive of these attributes (Pennebaker et al. 2015). Users input their text to the LIWC system, which outputs the analysis. The program is offered as both a free online processor, Online LIWC, as well as a paid subscription, with access to a larger dictionary base. For this study, we used the online tool, which was free and unencumbered. LIWC performs a word-by-word analysis—based on its dictionary-indexed attributes—of psychological features in the input or *target* text. LIWC analysis is arrived at through comparison with averages of previously performed professional assessments of personal statements.

Text quantity affects the accuracy of LIWC processing; the greater the number of words in the input text, the greater the accuracy of the LIWC analysis. For this reason, for each “political leaning” category, we concatenated the 30 translated-and-reviewed postings, created one long text file, and input that file to the Online LIWC system along with the provenance specification, SM. Thus, our Online LIWC system output consisted of three analyses, one for each of the categories, *VoteYes*, *VoteNo*, and *Neutral*, each of which contained Online LIWC scores for nine attributes<sup>†</sup>: 1) “I” Words, 2) Social Words, 3) and 4) Positive and Negative Emotions, 5) Cognitive Processes, 6) Analytic Thinking, 7) Clout, 8) Authenticity, and 9) Emotional Tone.

---

\* The main investigator on this effort, and first author of this report, was a Reserve Officer Training Corps (ROTC) cadet on a one-month internship with the US Army Combat Capabilities Development Command Army Research Laboratory.

<sup>†</sup> We use the term “attribute” here, rather than the term “category” used in the LIWC literature, to distinguish between categories of political leanings and those of LIWC analysis.

To convey the impact of these attributes, a brief description—based on reported studies—of what each one captures, follows here. To quote authors Pennebaker and Tausczik (2010, 30):

The words we use in daily life reflect what we are paying attention to, what we are thinking about, what we are trying to avoid, how we are feeling, and how we are organizing and analyzing our worlds.

With this in mind, and, with a view to understanding ensembles of internal motivations for political leanings, we also hypothesize general associations between LIWC output percentages for select attributes and plausible characteristics of the SM users who produced the analyzed text. The Attribute 1) “I” Words accounts for self-references, such as “I”, “me”, and “myself”, and reflects a focus on oneself and one’s own well-being.\* In our context, the expectation is there will be a high percentage for this attribute in one category relative to others, to suggest one side’s extreme concern about the effects—either on oneself or on one’s family—of a particular outcome of the referendum. Attribute 2) Social Words contains words referring to activities in social encounters and persons with others as, for example, “mate”, “talk”, and “buddy”. A focus on, or concern for, others might be reflected by high scores for this attribute.† The words of Attributes 3) and 4) Positive and Negative Emotion can be analyzed in conjunction with 2) Social Words to reflect the quality of a relationship, as noted by Pennebaker and Tausczik (2010, 32). However, for this study, Online LIWC assessed words like “love”, “happy”, and “peace” along with “hate”, “anger”, and “upset” on their own as reflecting positive and negative sentiment, respectively. For Attribute 5) Cognitive Processes, nouns, verbs, and adverbs referring to, or reflecting, mental activity or logical concepts are identified; for example, “think”, “because”, “insight”, and “belief”. These scores may point to a focus on knowledge and consideration—not necessarily borne of experience—given to referendum-related topics and opinions and the individuals’ expression thereof.‡

Attributes 6) Analytic Thinking, 7) Clout, 8) Authenticity, and 9) Emotional Tone are described, with references, on the LIWC website’s *interpreting-liwc-output* page (LIWC 2015). In contrast with Attributes 1) to 5), scores for these

---

\* Such references have been linked to depressive mood, experience of pain, the speech of younger individuals (Pennebaker and Tausczik 2010; 26, 30, 36), and the making of requests (Pennebaker 2011).

† Language use in interactions reflecting status, dominance, coordination, closeness, group processes, and other features of social relations, has been the subject of many studies (Pennebaker and Tausczik 2010, 33–35).

‡ While the set of attributes described up to this point can indeed serve to identify individual differences (Pennebaker and Tausczik 2010; 36, 37), it should be noted that such studies apply LIWC to data from individuals, rather than, as in the present case, data from virtual communities.

attributes are largely dependent on the presence of linguistic artifacts in more than one category. We will take a closer look at these.

The dictionary for Attribute 6) Analytic Thinking contains closed-class function words rather than open-class substantives. Function words, such as prepositions, conjunctions, and articles, are required to refer to intricately organized objects and concepts. For this reason, their use serves to distinguish such references from other types of text (i.e., narratives and here-and-now reporting). In contrast with *dynamic* language, Attribute 6) measures *categorical* language, which reflects logical and hierarchical thinking processes. The application of this type of thinking to referendum issues may be reflected in high scores here. Eight function-word dimensions contribute to this attribute's categorical-dynamic index (CDI) calculation.\*

Attribute 7) Clout, refers to relative social status, with higher scores reflecting greater levels of confidence or leadership. This attribute is based on the use of two different types of pronouns: references to oneself (e.g., “I” and “me”) and to collectives and others (e.g., “us” and “they”). Inherently social, research has linked patterns in pronoun use with focus tendencies associated with social status; self-focus with lower status and collective, or other-focus with higher status (Kacewicz et al. 2014, 128). We hypothesized that lower or higher scores for one side would then suggest auto-perception of lower or higher status, or possibly influence, in the referendum debate. Five separate studies found pronoun use reflecting the social position and the algorithm developed to capture these findings were incorporated into Online LIWC (Kacewicz et al. 2014).

Attribute 8) Authenticity processed word-type constellations, which, research has indicated, are in play when speakers are being honest and revealing their true selves. These are word-category patterns associated with humility and vulnerability and include a) more self-references, showing ownership of the information; b) fewer negative emotion words, indicating feelings of comfort and the taking of responsibility; and c) more markers of cognitive complexity associated with judgment-based distinctions. For the present study, one side may have exhibited more indicators of being forthright—rather than duplicitous—in the run-up to the Colombian referendum. Researchers developed an algorithm for scoring this attribute based on these empirically derived linguistic profiles, tested it for

---

\* Pennebaker et al. (2014, 1–6) has an overview of studies finding function words to be psychologically meaningful and has a definition of the CDI algorithm used for scoring Attribute (6) Analytic Thinking.

generalizability and predictability, and incorporated it into Online LIWC (Newman et al. 2003, 666–668).

Finally, Attribute 9) Emotional Tone conflates Attributes 3) Positive Emotion and 4) Negative Emotion, previously described. A significant difference between groups here would point up a strong contrast of emotions around referendum issues. This attribute is summarized well on the webpage. Its description states the algorithm—used in Cohn et al. (2004) and in Online LIWC—“puts the two dimensions into a single summary variable[,]” whose value is a number, and that “the higher the number, the more positive the tone” with numbers “below 50 suggest[ing] a more negative emotional tone” (LIWC 2015).

## 4. Experimental Design

---

We ran the postings in the three experimental conditions: A with hashtags, B without hashtags, and C with hashtags/without the word “peace”. We chose these formulations based on the results of initial probes. Recall that hashtag format consists of the pound sign, “#”, followed by words strung together without whitespace separators. Keeping in mind this format, the original language of the postings, Spanish, as well as Online LIWC reports of improved accuracy with increased data volume, for Condition A we broke apart the hashtags into whitespace-delimited words, translated posting+hashtag with Google Translate, and input English versions to Online LIWC. We thus aimed to enhance LIWC accuracy with the higher volume of input while, simultaneously, capturing internal state and shared attitude signals in posting and hashtag, respectively.

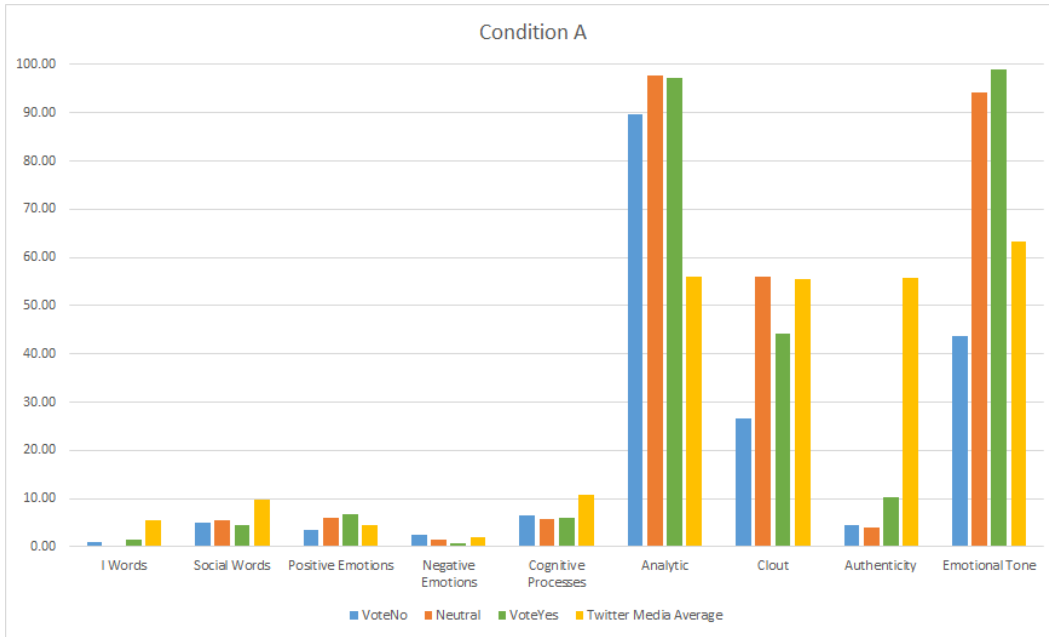
Condition B input to Google Translate consisted of the Spanish versions without hashtags; English output was fed to Online LIWC. Here, the focus was on the signal produced by the posting language without possible redundancy or bias introduced by the hashtag language.

In Condition C, hashtags were given the same treatment as in Condition A, with data preparation including the elimination of the word *paz*, “peace”. This condition was formulated as a check on LIWC values for Attributes 3), 4), and 9). Frequent referencing in both posting and hashtag data of the event in question, *el plebiscito por la paz*, “peace plebiscite”, a name also often abbreviated to merely *paz*, “peace”, increased occurrences of the word beyond the normal levels on which LIWC algorithms are calibrated. By removing it, we sought to attenuate possible skew resulting from the word’s inordinate frequency of occurrence.

## 5. Results

Language analysis of translated data allowed us to compare and contrast the internal states of SM users on two sides of a political question. In each condition, for each attribute, we analyze results for distinctions between Political Leaning Categories; Neutral Category results and Media Averages serve to scale our assessments.\*

Starting with Condition A, which sought to conflate signals in the posting, reflecting individuals' internal state, with those in the hashtags, reflecting shared attitudes about the political question, we observe the following in Fig. 1. On Attributes 2) Social Words, 4) Negative Emotion, and 5) Cognitive Process, category values for *VoteNo* are only slightly higher than those for *VoteYes*, reflecting perhaps a personal and collective *VoteNo* propensity for sadness or anger, cogitation, and concern for others. By contrast, for the other attributes, *VoteYes* values were higher. The difference was slight for Attribute 1) "I" Words and Attribute 3) Positive Emotion, possibly indicating slightly more *VoteYes* individual and group concern for themselves and their families, concerning the referendum, and a somewhat more upbeat *VoteYes* perspective.

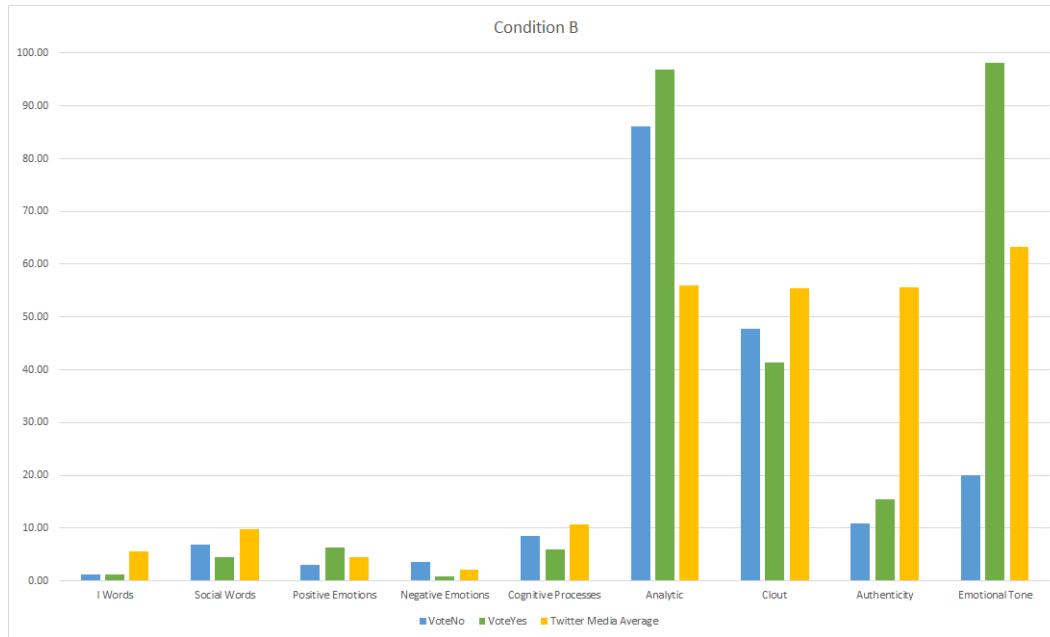


**Fig. 1 Online LIWC analysis of Google-translated SM postings in Condition A**

\* According to its webpage, LIWC processing is reliable for accurate calculations of true values for each category when data input is more than 100 words (LIWC 2015). Because tweets had a 140-character limit, in the interest of accuracy we concatenated the microblogs of each category in a batch-process. Without running individual tweets through LIWC processing, it was impossible to yield a standard deviation, without which statistical significance cannot be calculated.

For Attributes 6) to 9), those requiring complex calculations involving multiple types of linguistic artifacts, the category-based percentage differences were greater. Notably, the percentages themselves are generally much higher; percentage differences are also much larger. For example, on Attribute 5) Cognitive Processes, the percentages diverged by only 0.6, while on Attribute 6) Analytic Thinking, scores differed by 7.5 full points. The higher amount of language evidence for Attributes 6) to 9) may obscure the fact that while the attribute-specific percentages are higher—due to this increased amount of evidence—the percentage differences are also, in fact, larger. `VoteYes` scores for Attribute 8) Authenticity was more than double those for `VoteNo`, indicating the former’s political stance was also a true belief rather than merely expeditious support of the agreement. Attributes 7) and 9), Clout and Emotional Tone, showed the most divergent results. The disparity in Attribute 7) suggests higher `VoteYes` social status, leadership, and confidence. The discrepant results for Attributes 3) and 4), Positive and Negative Tone, are magnified in Attribute 9). Taken together, results for Attributes 7) and 9) reflect an enormously high degree of leadership optimism on the part of both individuals and the voting bloc.

Recall that Condition B sought, by filtering out hashtag information from tweet input, to focus attention on the individuals’ internal state rather than any shared political stance. As can be seen in Fig. 2, higher `VoteNo` values are reported for simple Attributes 2) Social Words, 4) Negative Emotion, and 5) Cognitive Processes. This means, for this sample, internal states for `VoteNo` microbloggers, in contrast to `VoteYes` counterparts, underscore more individual social relations and negative sentiment—possibly reflecting some qualitative aspect of the social relations—along with more mental activity involving logical concepts related to the issues, for individuals personally. At the same time, the highest `VoteNo` difference here—about three times greater than for the other attributes—is for the complex Attribute 7) Clout. This result suggests `VoteNo` microbloggers may, as individuals, perceive themselves personally as having high status and influence in this particular debate.



**Fig. 2 Online LIWC analysis of Google-translated SM postings in Condition B**

Higher *VoteYes* values can be seen for simple Attribute 3) Positive Emotions and complex Attributes 6) Analytic, 8) Authenticity, and 9) Emotional Tone. This suggests the sample’s *VoteYes* microbloggers’ positive attitude is personal, independent of political stance. That said, a *VoteYes* / *VoteNo* category-value comparison for these attributes indicates *VoteYes* individuals in this sample are personally more logical and hierarchical thinkers, with honest, true, and extremely positive sentiment—due to the large difference in Attribute 4) Negative Emotion affecting the Attribute 9) Emotional Tone difference spike of 78.2%—concerning the problem under scrutiny in the referendum, also interpretable as an optimistic attitude about solutions.

As noted previously, in Condition C we eliminated the word *paz* (peace) in posting+hashtag input to test the effects of its frequent occurrence as a constituent of the referendum title, *el plebiscito por la paz* (peace plebiscite), and as an abbreviation for the title.\* We wanted to test for interference in the LIWC analytic process for all attributes but especially for Attributes 3) Positive Emotions and 9) Emotional Tone, into the lexicons for which the word was integrated.

Condition C can be considered a variation of Condition A, with both taking as input data consisting of posting+hashtag. So, we compare and contrast those results here.

---

\* We removed the word only when it was used in the posting to refer to the referendum. This meant we removed 10 occurrences from the 30 concatenated *VoteNo* postings and 28 from the 30 *VoteYes* postings.

Values and value differences for Attributes 1) “I” Words, 2) Social Words, 4) Negative Emotions, 5) Cognitive Processes, 6) Analytic Thinking, and 7) Clout are relatively unchanged from previous analyses, with a few fractional percentage points up or down for one or the other of the categories.

Values and differences for remaining Attributes 3) Positive Emotions, 8) Authenticity, and 9) Emotional Tone, however, warrant a closer look. `VoteYes` group results showed a value-difference increase relative to `VoteNo` results of almost 2 points for Attribute 8) Authenticity, indicating the original stronger `VoteYes` signal, evident in words expressing humility and indicative of feelings of comfort, such as the one eliminated, was quite independent of any specific name reference.\* A similar inference can be made about the >0.5-point `VoteYes` group-value difference for Attribute 9) Emotional Tone: the original `VoteYes` positive tone is generally independent of the name’s frequent occurrence. However, given that LIWC calculations for Attribute 9) conflate values for Attributes 3) and 4), it may be affected by values for Attribute 3) Positive Emotions, which for the `VoteYes` group decreases more than 3 points and for the `VoteNo` group decreases less than 1.5 points from the previous analysis, causing the value difference to be much smaller; also, it indicates the initial Attribute 9) Positive Emotions’ `VoteYes` group value depended fairly heavily on the frequent occurrence of the referendum’s name.

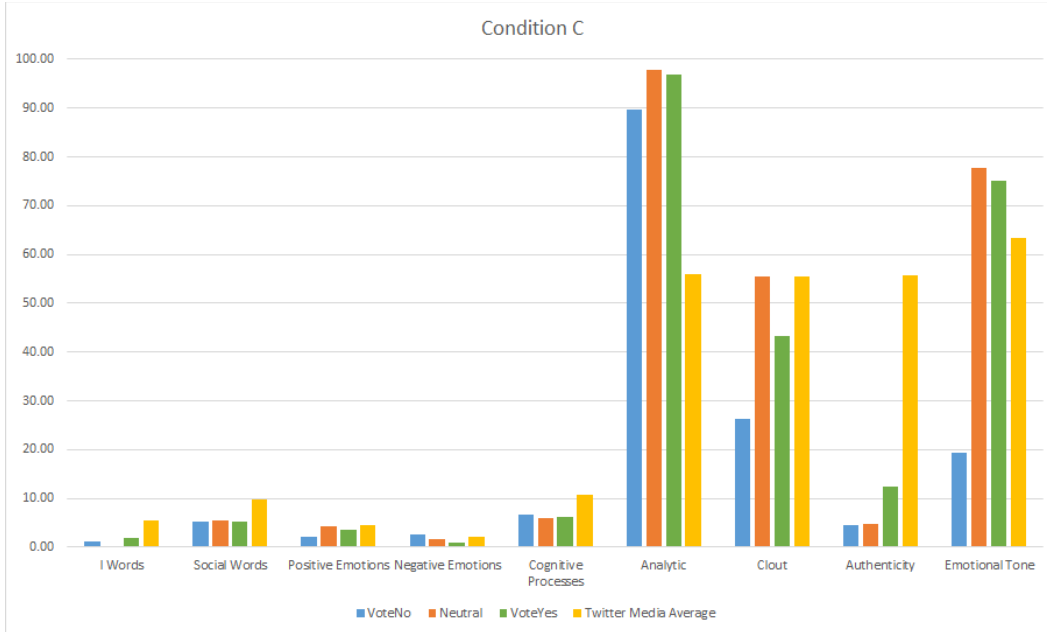
Conditions A and C include, as well, analyses of 30 `Neutral` postings, defined as being on the topic of the referendum but without hashtags. We consider this category relative to the other two and to itself. In both conditions, Attribute 1) “I” Words’ values were negligible, as expected in noneditorial media, which reports events in third person. In going from A to C, there was a similar dynamic for Attributes 2) Social Words, 4) Negative Emotions, and 6) Analytic Thinking. Percentages increased only very slightly or not at all, to allow closing in by higher `VoteYes` values resulting from a total-words decrease, which was less pronounced for `VoteNo`, permitting the latter difference to remain unchanged. While Attribute 5) Cognitive Processes’ `Neutral` values were lower than those for other categories, between A and C the differences were reduced due to a `Neutral` increase greater than those for the other categories. These changes were expected, based on the experimental design.

`Neutral` results for Attribute 7) Clout in Condition C posed a challenge for interpretation due to across-the-board value decreases, rather than the expected

---

\* The `VoteNo` signal was even less dependent, as its (8) Authenticity value was actually slightly increased from the previous analysis. See Figure 1.

increases resulting from a reduction in total words.\* The `VoteNo` value difference, however, remained unchanged as the decreases were the same in both categories and the `VoteYes` value difference, where the decrease was steeper, was slightly larger. (See Fig. 3.)



**Fig. 3 Online LIWC analysis of Google-translated SM postings in Condition C**

When the referendum name was filtered out of the data, all Category values for attributes affected by the name’s meaning changed in the same direction. For Attribute 8) Authenticity, they increased, and for Attributes 3) Positive Emotions and 9) Emotional Tone, they decreased. `Neutral` values for Attribute 8) increased more than three times as much as `VoteNo` values, but only within a point of each other as they were fairly low and at a level that was less than half of `VoteYes` values. `VoteYes` category percentage increases were then more than twice the `Neutral` category increases. Thus, Attribute 8) referendum-name interference affected `Neutral` more than `VoteNo` but much less than `VoteYes` category postings, due to a much higher volume of referendum-name mentions in the latter set of postings.

Distinct dynamics characterize the `Neutral` category for Attributes 3) Positive Emotions and 9) Emotional Tone values. For Attribute 3), the `VoteNo` value difference in Condition A, which was over three times that for `VoteYes`, was in

\* Our surmise is the Online LIWC algorithm for Clout not only omits the word for “peace” but also links pronoun type counts to related linguistic properties undiscoverable without specific word types, such as the ones assigned to those word occurrences that were removed in the framing of Condition C. For details on the LIWC algorithm, see Kacewicz et al. (2014).

Condition C slightly reduced, while the difference for `VoteYes` was made slightly larger. That is, it came in at a value lower than the `Neutral` value by more, indicating that “name effect” impacted postings in the `VoteYes` more than the `VoteNo` category, when measured against `Neutral` category postings.

For Attribute 9) the much higher values, along with the algorithm’s handling of words expressing both positive and negative emotions, told a different story. The `VoteNo` value in Condition A was lower than the `Neutral` value by over 10 times the percentage that the `VoteYes` value was higher than `Neutral`. In Condition C, the `VoteNo` value dipped by more than half, increasing the value difference by more than 0.1 while the `VoteYes` value dropped to lower than `Neutral` by about half of what it had been above the `VoteYes` value. The high levels and closeness of the `Neutral` and `VoteYes` values in both conditions indicate that extremely high values for Emotional tone resulted from the “name effect” and that the signal shared a similar degree of strength in both categories. By contrast, the even greater value difference with `VoteNo` postings, likely due to the higher-negative-lower-positive levels of emotion words in them, suggests that for Attribute 9) Emotional Tone, the “name effect” had a greater impact on the `VoteNo` than the `VoteYes` category, when measured against `Neutral` category values.

## 6. Discussion

---

While `Neutral` postings served as a gauge on value levels, the study’s findings are summarized in the `VoteYes` and `VoteNo` category differences, listed in Table 1, where `VoteNo` values are subtracted from `VoteYes` values, resulting in negative values when the `VoteNo` signal for the attribute is stronger. This is the case only for Attributes 2) Social Words, 4) Negative Emotions, 5) Cognitive Processes and, in Condition B, for Attribute 7) Clout. Note also the divergent `VoteYes` values in Condition B for not only Attribute 6) Analytic Thinking but also Attribute 9) Emotional Tone and, in Condition C, Attribute 3) Positive Emotion. For individuals posting in the `VoteNo` category, there is a marked focus on social connection, more negative emotion, cogitation, and a sense of empowerment while the `VoteYes` contingent is positive, analytical, and optimistic. Controlling for the referendum name, we note that `VoteYes` category posters, as a group, present as less positive, mildly less analytic, and with a more normalized tone than when the name is included in postings.

**Table 1 Summary of value differences; blue cells indicate stronger VoteNo signal**

Attributes Conditions	“I” words	Social words	Positive emotions	Negative emotions	Cognitive processes	Analytic thinking	Clout	Authenticity	Emotional tone
A	0.5	-0.5	3.3	-1.8	-0.6	7.5	17.4	6.0	55.2
B	0.0	-2.3	3.3	-2.7	-2.5	10.7	-6.3	4.6	78.2
C	0.8	0.0	1.4	-1.7	-0.5	7.1	17.0	7.8	55.8

## 7. Conclusions

Army ROTC programs allow four weeks for cadets to complete internship research projects, which was a limitation faced by one of the writers of this report. This limited length of time permitted only a preliminary analysis of the 90 microblogs and use of the online, rather than the paid subscription version, of the LIWC program. We envision future work to entail a deeper investigation into these 90 tweets and analysis of the entire collection to verify these emerging signals. See the Appendix for alternative views of the data, which also invite on-going analysis.

## 8. References

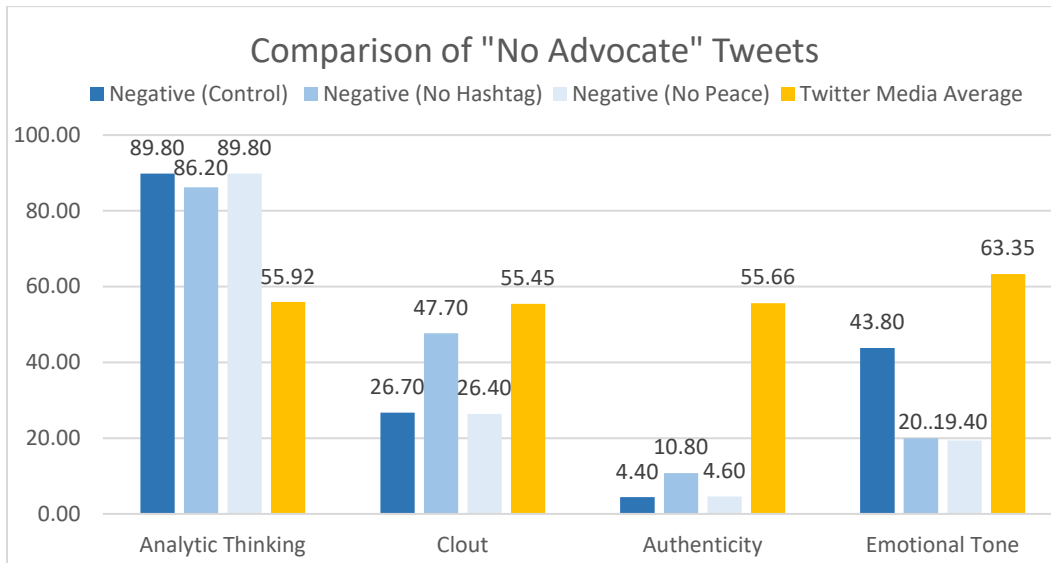
---

- Cohn MA, Mehl MR, Pennebaker JW. Linguistic markers of psychological change surrounding September 11, 2001. *Psych Sci.* 2004;15:687–693. <https://doi.org/10.1111/j.0956-7976.2004.00741.x>.
- Garcia-Gavilanes R, Quercia D, Jaimes A. Cultural dimensions in twitter: time, individualism and power. In: *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*; 2013.
- Kacewicz E, Pennebaker J, Davis M, Jeon M, Graesser AC. Pronoun use reflects standings in social hierarchies. *J Lang Soc Psych.* 2014;33(2):125–143. <https://doi.org/10.1177/0261927X13502654>
- Liang PW, Dai BR. Opinion mining on social media data. In: *Proceedings of IEEE 14th International Conference on Mobile Data Management*; 2013 June.
- LIWC. Linguistic inquiry and word count: discover LIWC 2015. Pennebaker Conglomerates, Inc. [accessed 2019 Sep 6]. <https://liwc.wpengine.com>.
- Newman ML, Pennebaker JW, Berry DS, Richards JM. Lying words: predicting deception from linguistic styles. *Pers Soc Psych Bul.* 2003;29(5):665–675. <https://doi.org/10.1177/0146167203029005010>.
- Nigam A, Dambanemuya HK, Joshi M, Chawla NV. Harvesting social signals to inform peace process implementation and monitoring. *Big Data.* 2017;5(4):337–355.
- Pennebaker JW. *The secret life of pronouns: what our words say about us*. New York (NY): Bloomsbury; 2011.
- Pennebaker JW, Boyd RL, Jordan K, Blackburn K. *The development and psychometric properties of LIWC2015*. Austin (TX): University of Texas at Austin; 2015.
- Pennebaker JW, Chung CK, Frazee J, Lavergne GM, Beaver DI. When small words foretell academic success: the case of college admissions essays. *PLoS ONE.* 2014;9(12):e115644. <https://doi.org/10.1371/journal.pone.0115844>.
- Pennebaker JW, King LA. Linguistic styles: language use as an individual difference. *J Pers Soc Psych.* 1999;77(6):1296–1312.
- Pennebaker JW, Tausczik YR. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psych.* 2010;29(1):24–54.

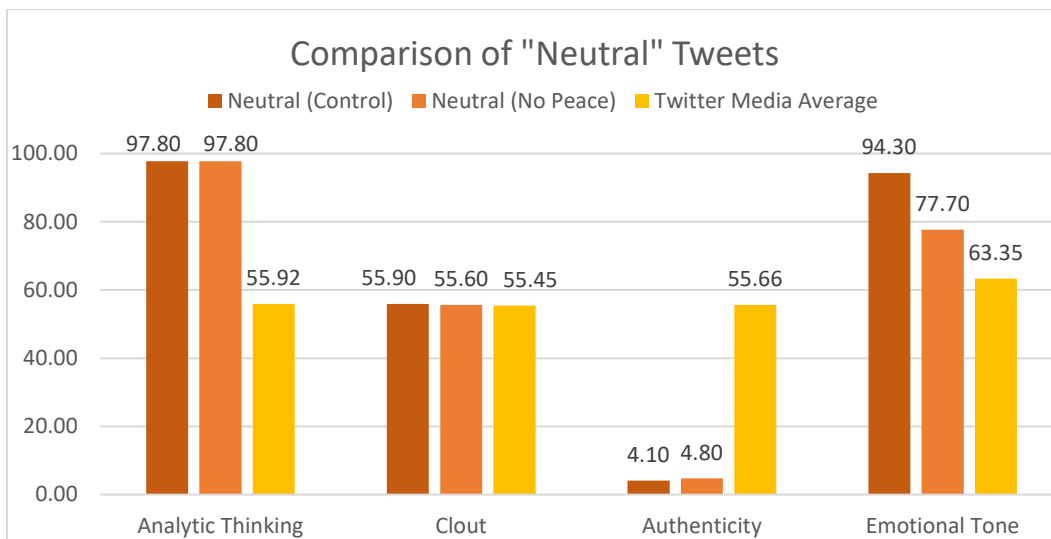
## **Appendix. Figures for Cross-Condition Intercategory Analysis**

---

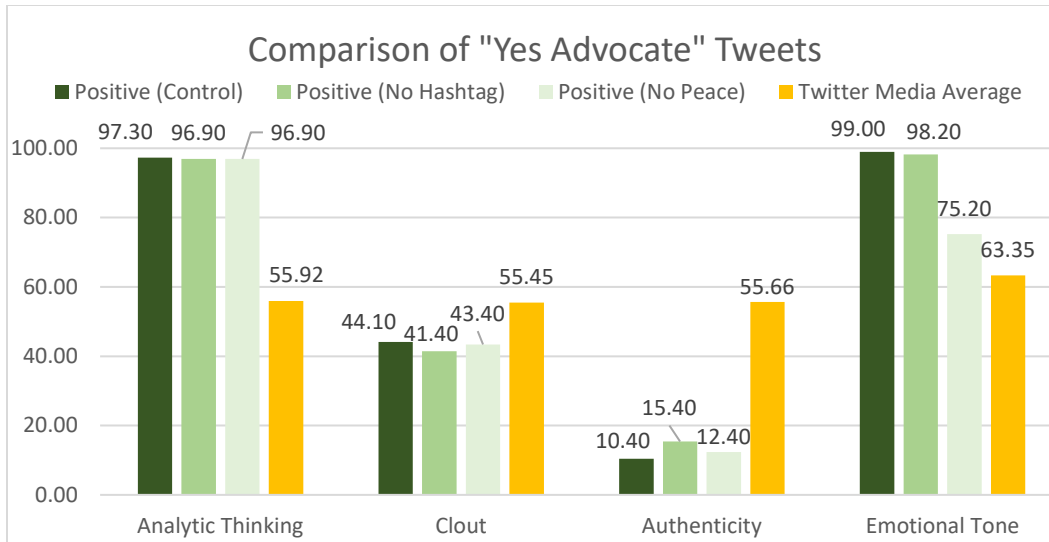
---



**Fig. A-1** Graph shows the four most relevant outputs from Linguistic Inquiry and Word Count (LIWC) based on no-advocate tweets compared with Twitter Media Average



**Fig. A-2** Four most relevant LIWC-category outputs for neutral tweets compared with Twitter Media Average



**Fig. A-3 Four most relevant LIWC-category outputs for yes-advocate tweets compared with Twitter Media Average**

## List of Symbols, Abbreviations, and Acronyms

---

CDI	categorical-dynamic index
FARC	Fuerzas Armadas Revolucionarias de Colombia
iCenSA	Interdisciplinary Center for Network Science and Applications
LIWC	Linguistic Inquiry and Word Count
MT	machine translation
NDU	Notre Dame University
PSI	Peace Studies Institute
ROTC	Reserve Officer Training Corps
SM	social media

1 DEFENSE TECHNICAL  
(PDF) INFORMATION CTR  
DTIC OCA

1 CCDC ARL  
(PDF) FCDD RLD CL  
TECH LIB

5 CCDC ARL  
(PDF) FCDD RLC IT  
M VANNI  
R HOBBS  
M BOWMAN  
C BONIAL  
FCDD RLD E  
K TAYLOR