Analyzing Microblogs: Exploring the Psychology of Political Leanings

Meaghan Bowman

Abstract**—**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. The output of the analytic tool, 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. Here, we applied the online tools, Google Translate and the University of Texas's LIWC, to a 90-posting sample from a corpus of Colombian Spanish microblogs. On translated disjoint sets, identified by hashtag as being authored by advocates of voting "No," advocates voting "Yes," and entities refraining from hashtag use, we observed the value of LIWC's Tone feature as distinguishing among the categories and the word "peace," as carrying particular significance, due to its frequency of use in the data.

Keywords**—**Colombia peace referendum, FARC, hashtags, linguistics, microblogging, social media.

I. Introduction

SOCIAL media is becoming an outlet for individuals to express their opinions on worldly matters. The 2016 Colombian Peace referendum is no exception. It is important to note that the public voted 50.2% against the referendum 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 could have become hostile and angry; they could have lost trust with the government and sought retribution from 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 the Colombian government 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.

Meaghan Bowman is with the University of California, Los Angeles, United States (e-mail: meaghan_bowman@yahoo.com).

II. Background

In 2016, the Colombian government was ready to make peace with the FARC anti-government 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 June 23, 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 October 2, 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, to prevent future instances of, similar errorful 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 [1]. The NDU PSI, assisted by the Interdisciplinary Center for Network Science and Applications (iCenSA) of the NDU Department of Computer Science, succeeded in collecting 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 [2].[1] NDU used the full dataset for their 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

[1] The specific timeframe was the 21-day period from 11 September through 1 October 2016.

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

III. Data and Methods

A. 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.[2]

Hashtags, when included in the input to MT, were handled by first removing the hash character from the tag itself.[3] 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 (a) that 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) that 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, stemming from, e.g., accent placement, and resulting in drastically altered meanings, as with *papa*, 'potato' being confounded with *papá*, 'dad' and *sí*, 'yes' with *si*, 'if.' Machine translated postings were, thus, reviewed, edited and confirmed as belonging to one of the three referendum-related categories.

B. "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 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.[4]

C. LIWC Processing

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 [5]. 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 by utilizing 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:[5] (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. To convey the impact of these attributes, a brief description--based on reported studies--of what each one captures, follows here. To quote the authors of [6],

> 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 [6:30].

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. Attribute (1) "I" Words, accounts for self-references, such as "I," "me," and "myself," and reflects a focus on oneself and one's well-

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

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

[4] The main investigator on this effort, and first author of this Technical Note, was an ROTC cadet on a one-month internship with the U.S. Army Research Laboratory.

[5] 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.

being.[6] In our context, the expectation is that 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.[7] 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 in [6:32]. 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.[8]

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 [8]. In contrast with Attributes (1) to (5), scores for these 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 (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.[9]

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 [10:128]. 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 [10].

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 referendum. Researchers developed an algorithm for scoring this attribute based on these empirically-derived linguistic profiles, tested it for generalizability and predictability, and incorporated it into Online LIWC [8].

Finally, Attribute (9) Emotional Tone conflates Attribute (3) Positive 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 that the algorithm--used in 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" [8].

D. 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, we, for Condition A, 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 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

[6] Such references have been linked to depressive mood [6:26], experience of pain [6:30], the speech of younger individuals [6:36], and the making of requests [7].

[7] Language use in interactions reflecting status, dominance, coordination, closeness, group processes, and other features of social relations, has been the subject of many studies [6:33-35].

[8] While the set of attributes described up to this point can indeed serve to identify individual differences [6:36-37], it should be noted that such studies apply LIWC to data from individuals, rather than, as in the present case, data from virtual communities.

[9] See [9:1-6] for an overview of studies finding function words to be psychologically meaningful and for a definition of the CDI algorithm used for scoring Attribute (6) Analytic Thinking.

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.

IV. Results

Language analysis on 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 processed, we analyze results for distinctions between Political Leaning Categories; Neutral Category results and Media Averages serve to scale our assessments.[10]

Starting with Condition A, which sought to conflate signals in the posting, reflecting internal state, with those in the hashtags, reflecting shared attitudes about the political question, we observe, in Fig. 1, on Attributes (2) Social Words, (4) Negative Emotion, and (5) Cognitive Process, Category values for VoteNo that 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.

For Attributes (6)-(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 .6, while on (6) Analytic Thinking, scores differed by 7.5 full points. The higher amount of language evidence for Attributes (6)-(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 that 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 (7) suggests higher VoteYes social status, leadership, and confidence. The discrepant results for Attributes (3) and (4), Positive and Negative Tone are magnified in (9). Taken together, results for (7) and (9) reflect an enormously high degree of leadership optimism on the part of both individuals and the voting block.

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 that, 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 that VoteNo microbloggers may, as individuals, perceive themselves personally as having high status and influence in this particular debate.

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 that 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 (4) Negative Emotion affecting the (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 above, 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.[11] We wanted to test for interference in the LIWC analytic process for all Attributes, but especially for (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. 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

[10] According to its webpage [8], LIWC processing is reliable for accurate calculations of true values for each category when data input is more than 100 words. 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.

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

or down for one or the other of the Categories.

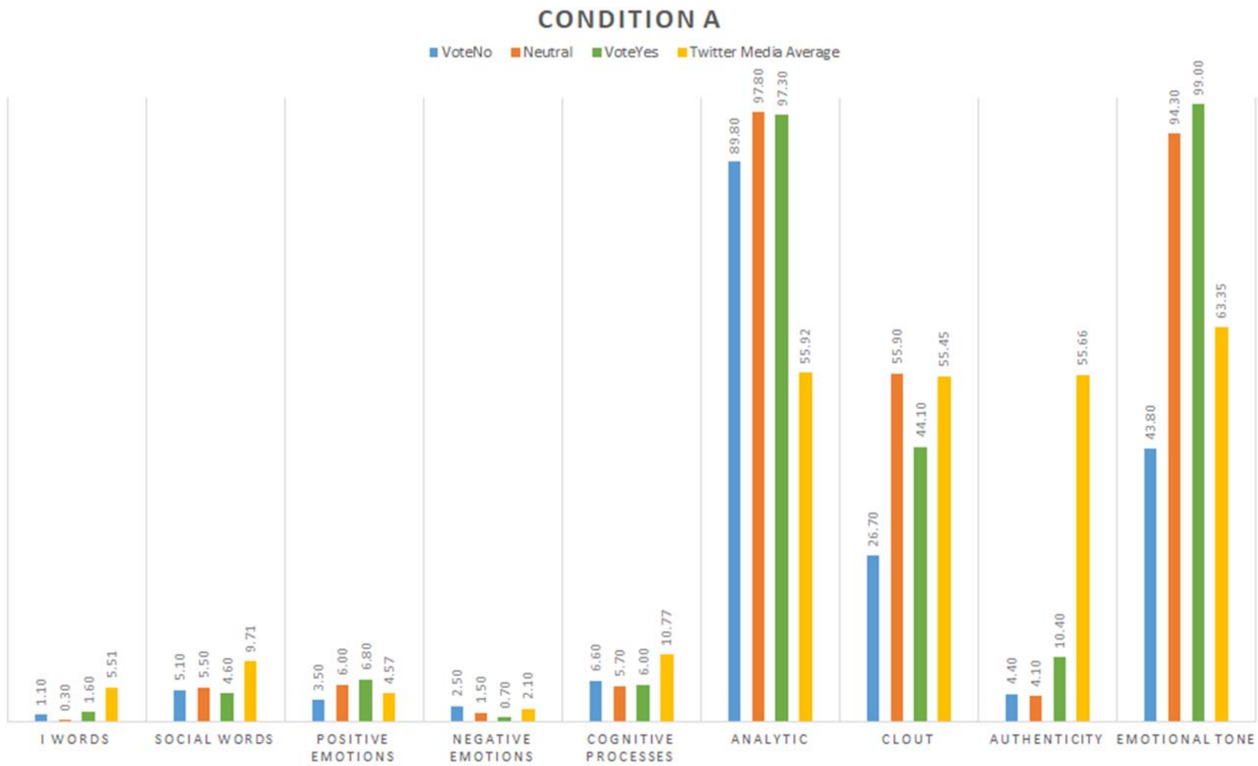

Fig. 1 LIWC Online analysis of Google-translated SM postings in Condition A

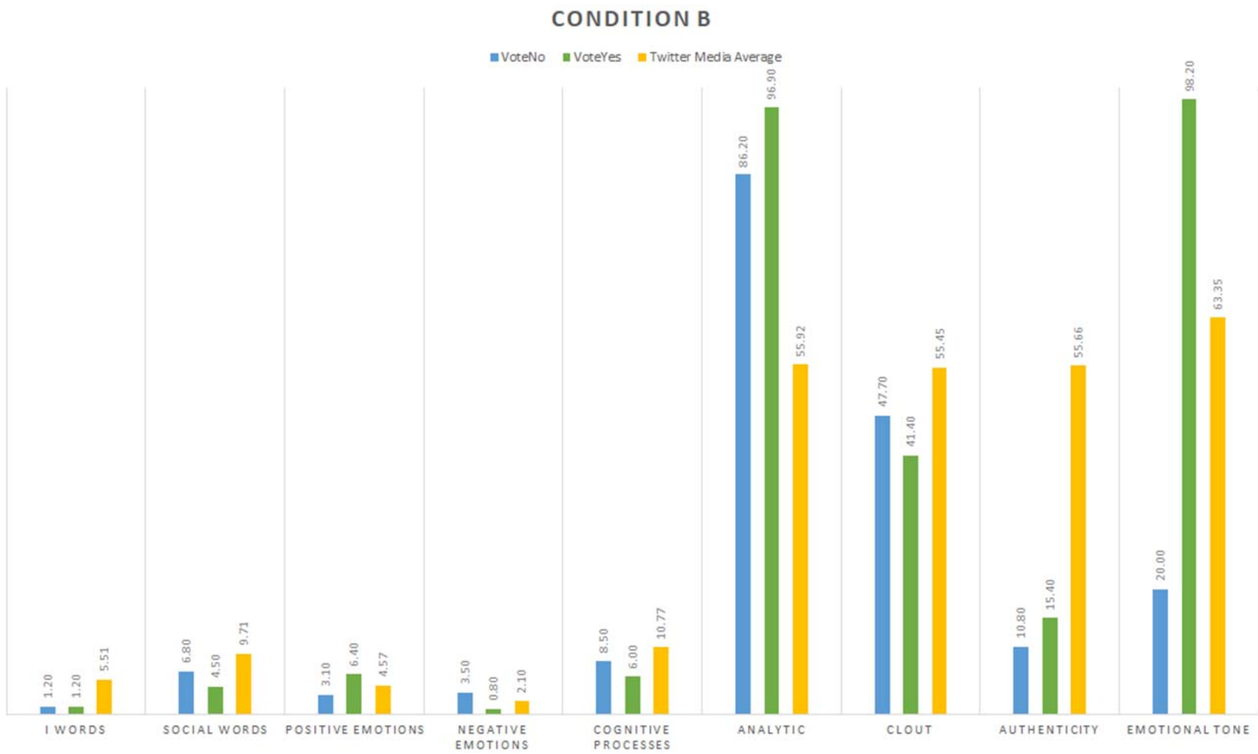

Fig. 2 LIWC Online analysis of Google-translated SM postings in Condition B

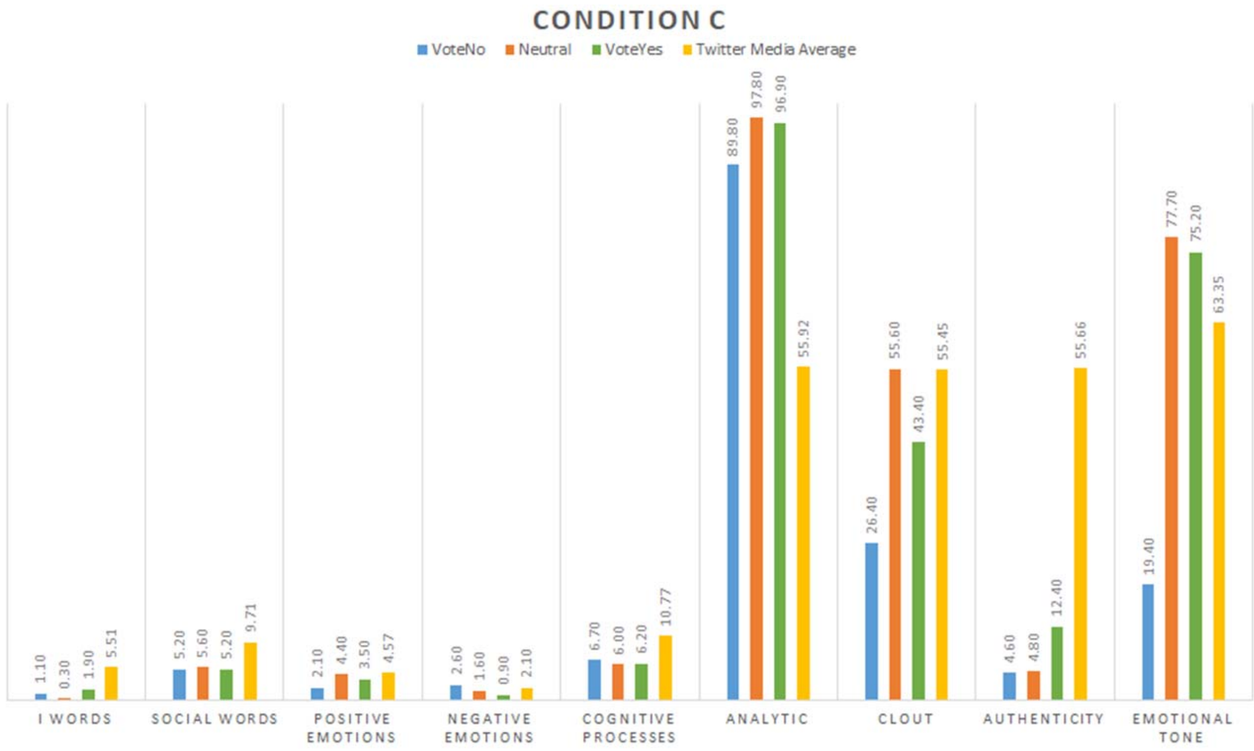

Fig. 3 LIWC Online analysis of Google-translated SM postings in Condition C

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 two full points for (8) Authenticity, indicating that 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.[12] A similar inference can be made about the greater-than-half-point VoteYes group value difference for (9) Emotional Tone: the original VoteYes positive tone is generally independent of the name's frequent occurrence. However, given that LIWC calculations for (9) conflate values for (3) and (4), it may be affected by values for (3) Positive Emotions, which for the VoteYes group decreases more than three points and for the VoteNo group decreases less than one-and-a-half points from the previous analysis, causing the value difference to be much smaller, and indicating that the initial (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 non-editorial 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 increases resulting from a reduction in total words.[13] 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.

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 (8) Authenticity, they increased, and for (3) Positive Emotions and (9) Emotional Tone, they decreased. Neutral values for (8) increased more than three times as much as VoteNo values, but only within a

[12] The VoteNo signal was even less dependent, as its (8) Authenticity value was actually slightly increased from the previous analysis, see Fig. 1.

[13] Our surmise is that the Online LIWC algorithm for Clout not only omits the word, '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 [10].

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. The VoteNo value difference in Condition A, which was over three times, for (3), that for VoteYes, was, in 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 the "name effect" impacted postings in the VoteYes more than the VoteNo Category, when measured against Neutral Category postings.

For (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 a tenth 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 (9) Emotional Tone, the "name effect" had a greater impact on the VoteNo than the VoteYes Category, when measured against Neutral Category values.

V. Discussion

While Neutral postings served as a gauge on value levels, the VoteYes and VoteNo Category differences, reported on in Table I, where VoteNo values are subtracted from VoteYes values, resulting in negative values when the VoteNo signal for the attribute was stronger. This is the case only for Attributes (2) Social Words, (4) Negative Emotions, (5) Cognitive Processes and, in Condition B, for (7) Clout. Note as well the divergent VoteYes values in Condition B for Attribute (6) Analytic Thinking and (9) Emotional Tone as well as, in Condition C, for (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.

TABLE I
Summary of Value Differences

	"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

Shaded cells indicate stronger VoteNo signal.

VI. Limitations of This Study

The Army ROTC program allows four weeks for cadets to complete their internship research projects. The 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. Future work will entail a deeper investigation into these 90 tweets or analysis of the entire collection, to verify these emerging signals.

APPENDIX: FIGURES FOR INTER-CATEGORY ANALYSIS

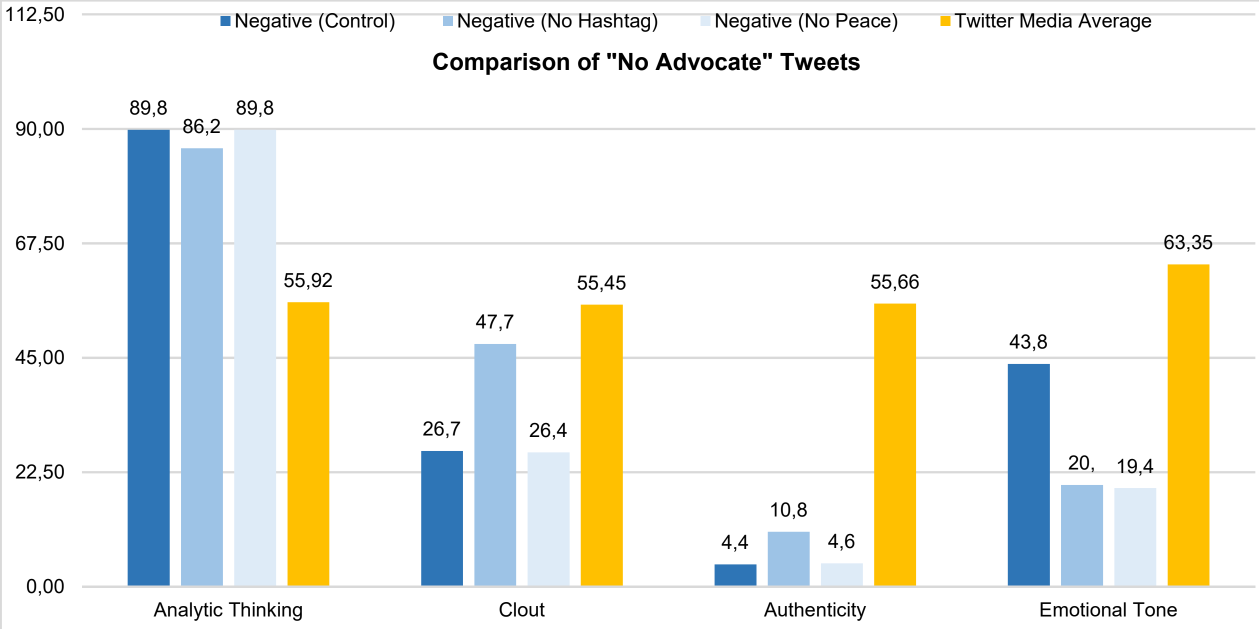

Fig. 4 The four most relevant outputs from LIWC based on no-advocate tweets compared with Twitter Media Average

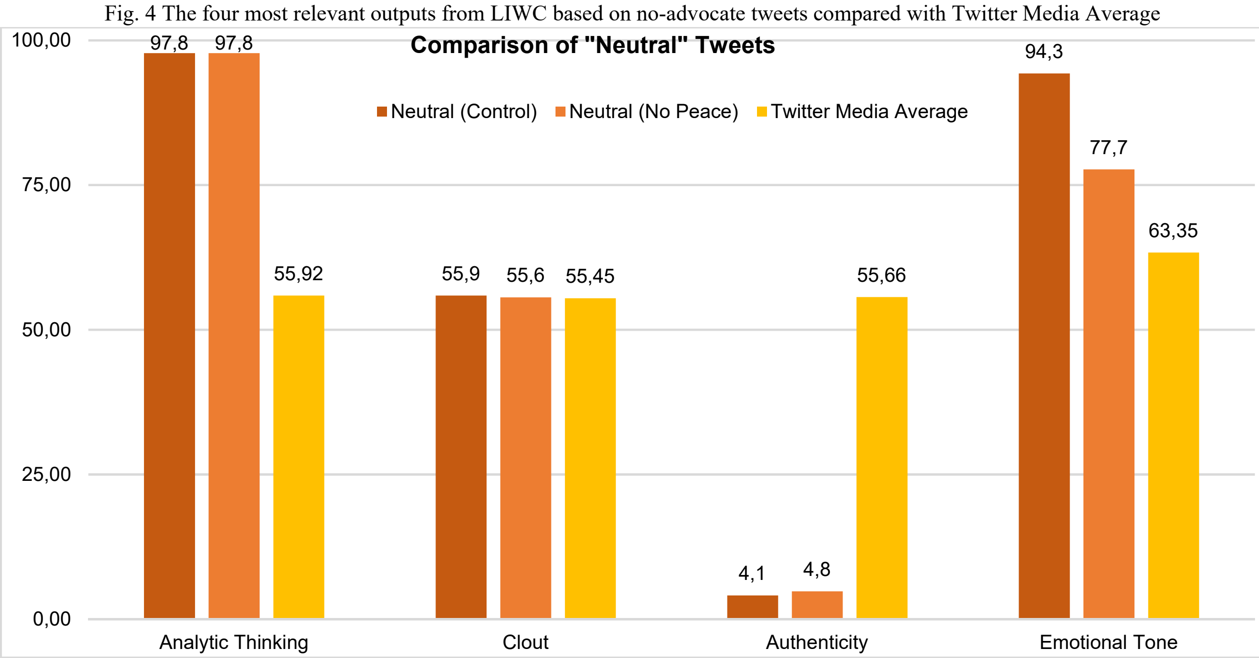

Fig. 5 Four most relevant LIWC-category outputs for neutral tweets compared with Twitter Media Average

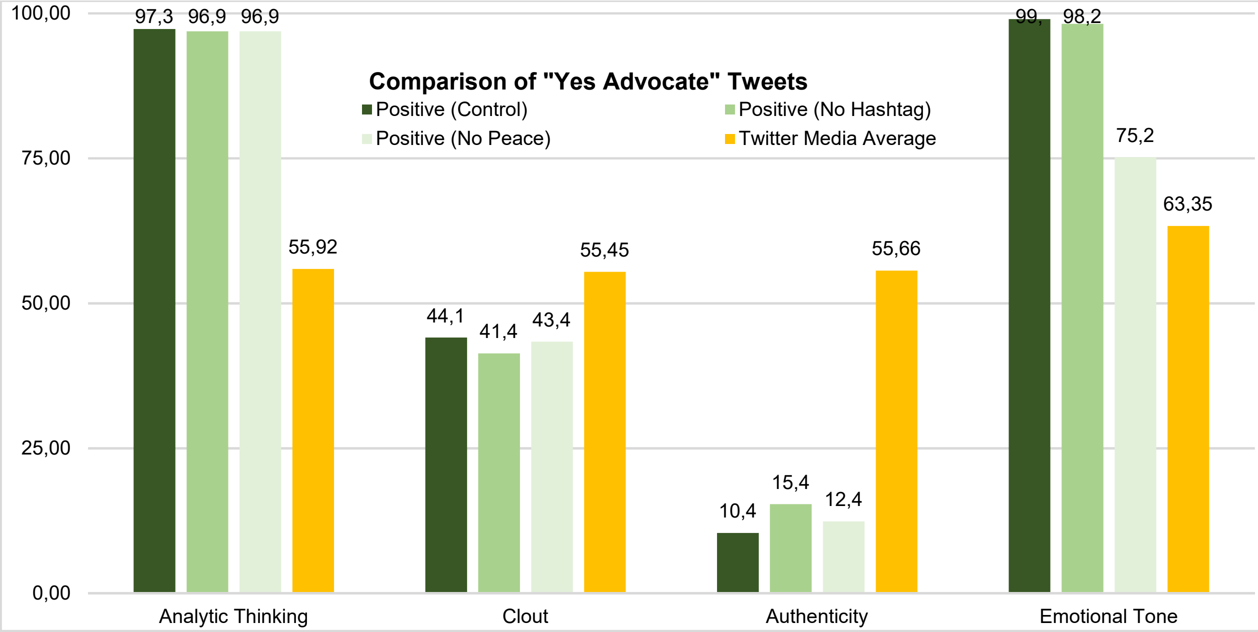

Fig. 6 Four most relevant LIWC-category outputs for yes-advocate tweets compared with Twitter Media Average

REFERENCES

[1] Liang, Po-Wei and Bi-Ru Dai. 2013. Opinion Mining on Social Media Data. In Proceedings of IEEE 14th International Conference on Mobile Data Management.

[2] Nigam, Aastha, Henry K Dambanemuya, Madhav Joshi, and Nitesh Chawla. 2017. Harvesting Social Signals to Inform Peace Process Implementation and Monitoring. *Big Data* 5:4:1-19.

[3] Pennebaker, James W. and Laura A. King. 1999. “Linguistic Styles: Language Use as an Individual Difference.” *Journal of Personality and Social Psychology* 77:6:1296-1312.

[4] Garcia-Gavilanes, Ruth, Damele Quercia, and Alejandro Jaimes. 2013. Cultural Dimensions in Twitter: Time, Individualism and Power. In Proceedings of the 7th International AAAI Conference on Weblogs and Social Media.

[5] Pennebaker, J.W., R.L. Boyd, K. Jordan, & K. Blackburn. 2015. Development & psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin.

[6] Pennebaker, J.W. and Y.R. Tausczik. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29:1:24–54.

[7] Kacewicz, E., J. Pennebaker, M. Davis, M. Jeon, and A.C. Graesser. 2014. Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, *33*:2:125-143. https://doi.org/10.1177/0261927X13502654

[8] Cohn, M.A., M.R. Mehl, and J.W. Pennebaker. 2004. Linguistic Markers of Psychological Change Surrounding September 11, 2001. *Psychological Science*, 15:687-693. https://doi.org/10.1111/j.0956-7976.2004.00741.x

[9] Newman, M.L., J.W. Pennebaker, D.S. Berry, and J.M. Richards. 2003. Lying Words: Predicting Deception from Linguistic Styles. *Personality and Social Psychology Bulletin*, *29*:5: 665–675. https://doi.org/10.1177/0146167203029005010