Volume 35 — Number 2 — 2022 — ISSN 2190-6858

JLCL
Journal for Language Technology
and Computational Linguistics

Special Issue on Computational Linguistics for Political and Social Sciences

Edited by
Ines Rehbein, Gabriella Lapesa, Goran Glavaš and Simone Paolo Ponzetto

Imprint

Editor
Christian Wartena
Publication supported by the Gesellschaft für Sprachtechnologie und Computerlinguistik (GSCL)

Board of Directors
Committee and Advisory Board of the GSCL

Current issue
Volume 35 2022 Issue 2

Guest Editors
Ines Rehbein, Gabriella Lapesa, Goran Glavaš and Simone Paolo Ponzetto

Address
Christian Wartena
Hochschule Hannover
Expo Plaza 2
D-30539 Hannover
info@jlcl.org

ISSN
2190-6858

Publication
Mostly 2 issues per annum
Publication only electronically on jlcl.org

License
Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

Contents

Computational Linguistics for Political and Social Sciences (Editorial)
Ines Rehbein, Gabriella Lapesa, Goran Glavaš, Simone Paolo Ponzetto **iii**

Small Data Problems in Political Research: A Critical Replication Study
Hugo de Vos, Suzan Verberne **1**

Frame Detection in German Political Discourses:
How Far Can We Go Without Large-Scale Manual Corpus Annotation?
Qi Yu, Anselm Fliethmann **15**

Share and Shout: Proto-Slogans in Online Political Communities
Irene Russo, Gloria Comandini, Tommaso Caselli, Viviana Patti **33**

UNSC-NE:
A Named Entity Extension to the UN Security Council Debates Corpus
Luis Glaser, Ronny Patz, Manfred Stede **51**

Irene Russo, Gloria Comandini, Tommaso Caselli and Viviana Patti

Share and Shout: Proto-Slogans in Online Political Communities

Abstract

This paper proposes a methodology for investigating populism on social media by analyzing the emergence of proto-slogans, defined as nominal utterances (NUs) typical of a political community on social media.
We extracted more than 700.000 comments from the public Facebook pages of two Italian populist parties' leaders (Matteo Salvini and Luigi Di Maio) during the week preceding the 2019 European elections (i.e., from May 20 to May 26, 2019). These comments have been automatically clustered and manually annotated to find proto-slogans created by the parties' supporters.
Our manual annotation consists of four layers, namely:*Nominal Utterances* (NUs), a syntactic device widely used for slogans; *Slogans* for NUs with a slogan function; *Top-down/Bottom-up*, to recognize the slogans produced by the politicians and those produced by supporters; *Proto-slogans*, for NUs devoid of specific political content that nonetheless express partisanship and support for the leaders.

1 Introduction

Social media have increasingly become arenas of mainstream political discourse. Platforms like Facebook and Twitter offer politicians venues to express their views, aggregate supporters and critics, and reinforce identities (Stieglitz & Dang-Xuan, 2012; Stier, Bleier, Lietz, & Strohmaier, 2018).

The vast amount of comments on political topics daily produced by users can be monitored and analyzed using Natural Language Processing (NLP) tools to focus on relevant societal issues such as hate speech and fake news. However, apart from long comments that express more complex opinions, most comments on social media are characterized by the synthetic expression of a point of view. Analyzing this type of content is challenging: due to their brevity, a topic-based analysis of users' comments performs poorly. Nonetheless, short comments have a pragmatic function in online debates, and often through the use of nominal utterances (NUs) (Comandini & Patti, 2019; Comandini, Speranza, & Magnini, 2018), help to build a common view among supporters of the same party/politician.

NUs, intended as syntactic declarative constructions built around a nonverbal head, can be part of a shared vocabulary used to express the in-group sense of cohesion and belonging on political pages and fora. For example, the NUs *Italia agli Italiani* (*Italy to Italians*) and *Porti chiusi* (*Closed harbors*) uniquely characterize one of the political communities investigated in this paper, i.e. Lega Nord (LN, Northern League). Several

of these recurrent NUs are slogans carefully created by the politicians' communication staff and used by supporters to reinforce the sense of belonging to a community. However, political slogans can also emerge from supporters' interactions on social media such as Facebook. They can become a trademark of a political community on other social media, such as Twitter. We define this process as proto-slogan generation.

Proto-slogans are semi-fixed linguistic expressions realized by NUs; they express a generic stance - positive or negative - toward a target. They emerge in online environments, in communities of people sharing the same perspectives or points of view. We find that proto-slogans are a communicative device exploited by populist supporters, a stylistic feature usable in the future to detect emerging populist attitudes.

In this paper, we study online political communities, extracting comments from the public Facebook pages of two populist Italian party leaders, Matteo Salvini for the Lega Nord (LN, Northern League) and Luigi Di Maio for the Movimento 5 Stelle (M5S, Five Stars Movement), during the week preceding the 2019 European elections (i.e., from May 20 to May 26, 2019). At that time, these two leaders were both covering the position of deputy prime minister in the so-called yellow-green government. Their parties were gaining consensus, with the LN winning the European elections. However, both parties have lost consensus at the time of writing, and their leaders have changed their communication. These changes are mainly due to the new roles these leaders are now covering (Salvini being part of the ruling majority with no position in the government and Di Maio being Foreign Minister). Besides this difference, the data analyzed still represent a valuable tool for gaining insights into the communication strategies of populist leaders and parties.

To investigate the linguistic behavior of these communities, we propose a semi-supervised methodology that combines K-means clustering and manual annotation for the identification of proto-slogans. Additionally, we compare slogans extracted from Facebook with slogans retrieved on Twitter in different periods to distinguish between attested and emerging slogans. This comparison validates what has been extracted from Facebook's public pages on Twitter, where linguistic choices can be crucial for identifying communities if there are no other metadata available (such as the information that a user follows a politician).

Contributions The main contributions of this paper can be summarized as follows:

- an operational definition of proto-slogan as a key aspect in bottom-up populist communication on the web (Section 3.1);
- a methodology that combines unsupervised approaches (i.e., clustering) and manual annotation to identify political proto-slogans (Sections 3.3 and 3.4).

Article Outline The remainder of this article is structured as follows: in Section 2, we describe related work on populism on social media, what characterizes the language of online communities, and how slogans have been linguistically analyzed. Section 3 focuses on the methodology, presenting a definition of proto-slogan. It also describes how data

from social media have been extracted and pre-processed. The results emerging from the manual annotation of automatically clustered instances are discussed. In Section 4, we provide some preliminary comparisons between Facebook and Twitter data, before concluding in Section 5.

2 Related Work

Populist discourses have been analyzed from different perspectives, collecting them in corpora that ease qualitative and quantitative analyses. However, it is still unclear if populist discourses can be linguistically identified - independently from the historical moment, the latitude, and the political orientation. Moreover, social media has had a disruptive effect on the propagation of populist discourses, affecting some of its features in a way still under scrutiny. This section reports on relevant literature about populism on social media and how language is shaped by communication in online communities in order to frame the reception of populist discourses in online contexts. In particular, slogans are presented as stylistic devices related to the emergence of a shared attitude among users.

2.1 Populism on Social Media

Social media are fundamental for understanding populist ideologies, which are mainly identified by their communication style (Kriesi, 2014; Aslanidis, 2016; Stanyer, Salgado, & Strömbäck, 2016). In particular, Facebook seems to be the preferred social network of populist parties (Ernst, Engesser, Buchel, Blassnig, & Esser, 2017).

In this work, we will adopt a broad definition of populism as a discourse based on the juxtaposition of two homogeneous and antagonistic groups: "the good people" (the in-group) VS "the bad elite/the foreigners" (the out-group) (Mudde, 2004; Rooduijn & Akkerman, 2017).

Charismatic leaders are particularly relevant for populist parties, and on Facebook they are often more popular than the official party's page (Bobba, 2019). Thus, to study populist rhetoric, it is preferable to focus on the rhetoric of political party leaders, analyzing how supporters react to it.

Populist leaders often adopt an emotional and straightforward communication style to be more persuasive and trigger a more emphatic response on social media (Oliver & Rahn, 2016). Indeed, it has been shown that emotionalized-style messages produced by Matteo Salvini on Facebook are more popular than his more neutral messages (Bobba, 2019).

Using these emotional messages and the direct connection with the public provided by social networks, populist leaders can forge close ties with their supporters, appearing more approachable (Jacobs & Spierings, 2016). Therefore, populist leaders can transform their Facebook pages into sheltered spaces for their fans, creating echo chambers in which aggressive tones can be cultivated (Ernst, Esser, Blassnig, & Engesser, 2018; Engesser, Fawzy, & Larsson, 2017).

Together with the sense of belonging to the in-group due to the general resentment toward the out-group (Hameleers, Reinemann, Schmuck, & Fawzi, 2019), this perceived intimacy with the leader creates a strong sense of being part of a homogeneous community, supportive of their leaders. In this way, populist leader's supporters may experience inter-group emotions, with each member experiencing emotions and taking action on behalf of the group (Smith & Mackie, 2008).

Previous computational linguistics studies of populism are scarce. Recently, (Huguet Cabot, Abadi, Fischer, & Shutova, 2021) present a crowdsourcing annotated dataset for populist attitudes that collects comments about news on Reddit that mention a set of social groups (i.e., immigrants and Muslims), manually classifying attitudes toward them as supportive, critical, or discriminatory. In detecting the overall stance of comments, their analysis does not target exclusively populistic content and how populist attitudes are expressed. Instead, our work starts with the assumption that comments on the chosen politicians' Facebook public pages are mainly supportive and sympathetic to the populist rhetoric. Thus, they constitute the ideal starting point for a stylistic investigation of the reception of populist discourse online.

2.2 The language of online communities

Computational analyses of language used in online communities revealed that talking in a particular way on social media reinforces our networks and sense of belonging (McCulloch, 2019). For example, the use of written slang on Twitter depends on the number of times people see the new word and if a member of their network uses it or not (Eisenstein, O'Connor, Smith, & Xing, 2014). The central members of the network introduce lexical innovations that are successfully adopted by other members if there is a subset of adopters with strong ties (Tredici & Fernández, 2018). Creating a shared vocabulary is both a prerequisite and a consequence of being part of a cohesive community, even online.

With a data-oriented analysis, Khalid and Srinivasan (2020) show that different communities have peculiar styles, and the stylistic choices of users are a good predictor of group membership, more than the topics discussed.

The adaptations at the stylistic level contribute to being well-received by a community. In (Tran & Ostendorf, 2016) the reception of content posted by users with positive feedback is investigated through a hybrid n-grams and topic models to characterize the style and the topic of language in Reddit online communities. Stylistic features have discriminatory power for distinguishing between communities: the style is a better indicator of community identity than the topic. The authors found a positive correlation between the community reception of a contribution and the style similarity to that community. On the contrary, this does not hold for topic similarity.

In this paper, we argue that nominal utterances are a stylistic device that characterizes the reception of populist discourse online, being the pragmatic choice that supports the creation of a shared vocabulary among supporters.

2.3 Slogans as Linguistic Devices

Slogans are usually short, expressive, and assertive utterances, easy to memorize and spread (Amălăncei, Cîrtiţă-Buzoianu, & Daba-Buzioianu, 2015). They are defined linguistically by their pragmatic function: expressing an idea memorably and economically. They can have a broad range of syntactic forms and can be characterized by their use of figures of speech and rhetorical devices, such as metaphors ("Imagination at Work" from General Electric implies the metaphor that General Electric is imagination), parallel constructions ("Melts in your mouth, not in your hands" from M&M's) or alliteration ("Don't dream it. Drive it" from Jaguar) (Alnajjar & Toivonen, 2020).

The slogans that have received attention in previous works are those used in advertising. Political slogans are less studied, although they generally follow the same advertising rules and have the same goal: influencing people's behaviors (Ferrier, 2014). Furthermore, political slogans usually convey a strongly supportive or condemning message towards a person or a political program/action, because voters are mainly influenced not by their conscious opinion on a politician's program, but by their feeling about a candidate or a party (Westen, 2007)).

The procedure primarily used in studying slogans is top-down, concentrating on pre-existing slogans professionally crafted by politicians or companies. On the contrary, a bottom-up approach is much more complicated, because it would require recognizing slogans in day-to-day communication focusing on linguistic features.

Top-down slogans have a pragmatic function: they are created to persuade others. On the other hand, bottom-up slogans, emerging as linguistic devices shared by like-minded people, have a different function: they are used to structure and enhance the cohesion of online communities. In this paper, we present a methodology to detect bottom-up slogans that, if widely adopted, can shed light on the emerging attitudes of political supporters online.

3 Methodology

This section presents a definition of proto-slogan and reports on how data have been extracted from Facebook and pre-processed before automatically clustering them. In addition, the details of the manual annotation performed on clustered data are documented.

3.1 Definition of proto-slogan

Regarding the general political rhetoric, it is possible to differentiate two different kinds of slogans: top-down and bottom-up. Top-down political slogans are produced by the politician's communication team and generally convey a complex message in a carefully crafted short, sharp form, such as *porti chiusi* (*closed harbors*); on the other hand, bottom-up political slogans are produced by the political electorate and generally convey a less complex message, as in *avanti tutta* (*full steam ahead*).

However, analyzing bottom-up political slogans, it would appear that some of the linguistic devices used by the political electorate, such as *forza Salvini* (*go Salvini*), are even simpler than the others, conveying an even less complex message. These very simple linguistic devices do not appear to be real slogans, because they convey only the user's stance (positive or negative) regarding a target, which is always explicitly mentioned.

Some of the linguistic devices used by the political electorate are slogan-like constructions that enhance cohesion inside the group. However, they are also simpler than real political slogans produced spontaneously by the politician's supporters, which still convey a relatively complex message.

For example, these slogan-like constructions usually appear as concise messages supporting (example 1) or denigrating (example 2) a politician or a group of people.

1. *Grande Di Maio!* [*Great Di Maio!*]

2. *Giornalisti venduti!* [*Corrupted journalists!*]

Short slogan-like bottom-up constructions that convey a basic message of support/denigration will be called proto-slogans, assuming that they are an embryonic form of a real slogan, since they convey a positive or negative stance, but not the more complex messages typical of slogans. More specifically, proto-slogans are classified as a subset of bottom-up slogans, as illustrated in Figure 1.

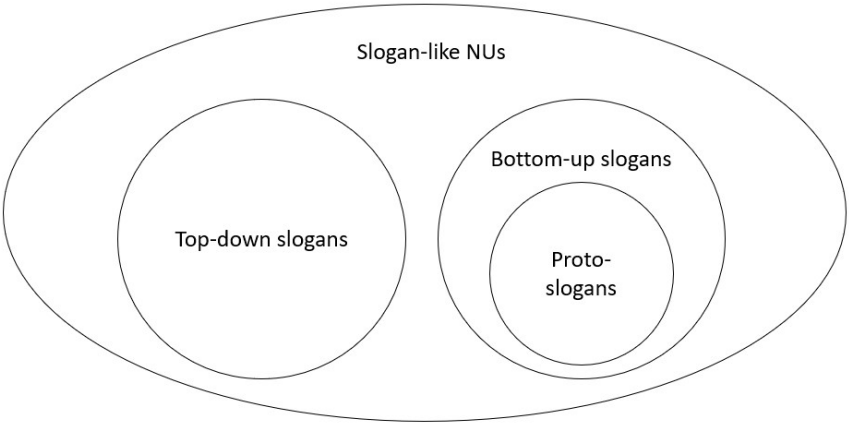


Figure 1: The subsets of slogan-like NUs

In this paper, we elaborate on the notion of proto-slogan as a specific device to build cohesion in online communities, proposing a methodology to identify them in social media.

Even if peculiar syntactic structures do not characterize slogans, slogans and especially proto-slogans are often realized syntactically as nominal utterances (Comandini et al., 2018), also known as fragments without an overt antecedent (Merchant, 2005). NUs are linguistic constructions without a verb in a finite form in their syntactic nucleus and are very common in informal spoken English (Merchant, 2005) (example 3) and Italian (Cresti, 1998) (example 4).

3. (After meeting Valentina at a social event, Katia says to her) *Nice dress, by the way!*

4. (When it begins to rain at the park, Monica says to her children) *Presto, tutti a casa!* [*Quick, everybody home!*]

NUs convey their content in a way that is expressive and informative but also very economic (Ferrari, 2011a, 2011b). It has been found that NUs are often used as a device to convey hate speech, as shown in the POP-HS-IT corpus, by frequently taking the form of a verbless, hateful slogan (Comandini & Patti, 2019), such as *TOLLERANZA ZERO* (*ZERO TOLLERANCE*) and *rimpatriare subito tutti gli immigrati irregolari* (*immediately repatriate all irregular immigrants*). Indeed, being without a verb in a finite form, NUs do not convey information about time, person, or aspect, creating messages similar to always valid maxims, mottoes, and, more importantly, actual slogans (Benveniste, 1990). Thus, not all NUs are slogans, but slogans are often NUs.

Therefore, many slogan-like constructions are NUs. Slogan-like NUs can be separated into two groups, as illustrated in Figure 1: top-down slogan-like NUs, produced by the politician's communication team, and bottom-up slogan-like NUs, produced spontaneously by the politician's supporters. Bottom-up slogan-like NUs also have an additional subset: proto-slogans, which convey a positive or negative stance towards a target.

3.2 Data Extraction and Preprocessing

This paper focuses on the online audience of Matteo Salvini and Luigi Di Maio, the two leaders of widely recognized populist parties, LN and M5S. In the selected period for the data collection, between May 20 and May 26, 2019, covering the last week of the political campaign before the 2019 European elections, their communication was primarily conveyed through social media. We used Netvizz (Rieder, 2013), a tool that crawls data from Facebook,[1] to extract posts and comments from the Facebook public pages of the two politicians. We have excluded all posts written by the leaders and the replies to comments by other users, focusing our analysis on direct comments

[1] Netvizz is no longer available because, from September 4, 2019, it has no more Page Public Content access.

FB page	Avg. post length	Avg. comm. lenght
Salvini	36.86±22.27	11.66±17.91
Di Maio	79.91±114.05	17.73±34.19

Table 1: Overview of the collected data from Salvini's and Di Maio Facebook pages between May 20th-26th, 2019. All numbers refer to tokens.

FB page	Comments	Eligible NUs
Salvini	565,411	201,179
Di Maio	135,022	42,064

Table 2: Eligible NUs after preprocessing

to the posts. Table 1 reports an overview of the extracted messages in terms of the average length of the posts published by the politicians (in tokens), the number of direct comments, and the average length of the comments.

As Table 1 shows, the two groups have similarities and differences. In both cases, we observe that the average length of the users' comments tends to be shorter than that of the posts published by the politicians. At the same time, it seems that the two groups of users (and ideally, the level of the interactions with their leaders) tend to differ, with users on Salvini's page producing shorter messages than those on Di Maio's. Since short comments are often verbless, we focus on nominal utterances (NUs) as syntactic declarative constructions built around a nonverbal head, framing them as the minimal unit of meaning in online communication. Therefore, we developed a preprocessing procedure to find out all NUs contained in the comments. Each comment has been preprocessed according to the following steps:

- its content has been sentence-splitted with NLTK (Bird, Klein, & Loper, 2009);
- its content has been PoS-tagged with TreeTagger (Schmid, 1995);
- finally, sentences that contain a verb in the finite form have been filtered out to include in the final dataset only the potential nominal utterances; sentences containing proper nouns other than *Matteo*, *Salvini*, *Luigi*, and *Di Maio* have been filtered out to exclude comments mentioning Facebook users.

The dimensions of the dataset before and after the preprocessing steps are reported in Table 2. The preprocessed data, more than 240k comments, have been the input for the clustering algorithm based on semantic similarity.

3.3 Aggregating Data Through Clustering

The amount of data after preprocessing is such that a manual exploration is not feasible (see Table 2 for details). We thus decided to aggregate the pre-processed messages using

clustering and perform manual annotation on aggregated data. Our approach is based on K-means clustering.

Such an approach has advantages and disadvantages for our task and, most importantly, our data. Results from K-means are easy to interpret and can be refined by manual inspection. At the same time, we are aware that K-means is not the best solution in an exploratory task, such as ours, where the number of clusters is not known, and it can hardly be assumed *a priori*. In this case, using known estimate methods such as the Elbow curve does not represent a solution.

We have addressed this issue by empirically validating the clusters of different sizes by using a sample of the data of 40k messages from Salvini's comments. First, each message representing an eligible NU has been converted as a 300-dimensions vector using FastText (Bojanowski, Grave, Joulin, & Mikolov, 2016). We then computed the pairwise cosine similarity scores between vectorized messages. The result is a N by N matrix of similarity scores.

Similarity scores below 0.6 were excluded and replaced with zeros to reduce noise in the data.[2] The matrix has been used as input to the K-means algorithm. [3]

We experimented with generating three groups of clusters of different sizes: 100, 150, and 200. Although none of them would correspond to an ideal amount of clusters for the aggregation of users' messages, their sizes allow for an easy and quick manual exploration of the data, ensuring a fine-grained level of analysis. We plotted their centroids and observed their distributions for each group of clusters. Quite interestingly, we could not find distinguishing differences or remarkable patterns. We finally selected 150 clusters as an appropriate level of aggregation to be subsequently manually annotated. Finally, we clustered the comments from Salvini's page daily, creating eight sets of comments, while aggregating those for Di Maio in three blocks. These differences are due to the number of messages available for the two politicians.

3.4 Manual annotation

Since centroids have been obtained by means of semantic similarity scores, focusing on them is a way to avoid annotating all the comments (a task that is not feasible) or annotating a not representative sample. The list of centroids (1,650 in total) obtained from K-means clustering has been manually annotated by two annotators with four annotation layers. These annotation layers are performed sequentially, and each of them is essential to understand the frequencies of NUs with different functions in each community. The agreement between annotators is calculated after discussion on divergent choices.

The first layer identifies NUs, which can be annotated following Comandini and Patti (2019) guidelines with a good agreement (0.96 in terms of Cohen's Kappa). We considered hashtags formed by two or more words as a single noun for this task, even

[2] With a similarity equals to zero, messages are considered to be very different.

[3] We used the K-means implementation available in the sci-kit learn Python library (Pedregosa et al., 2011).

if they contained a verb in a finite form. Most of these verbal hashtags are not used as VPs, but as nominal elements, linking the post to an "existing collective practice" (Zappavigna, 2015). The clause is excluded from the annotation when a NU has a coordinate clause with a verb in a finite form. Verbs in a non-finite form (infinitive, gerund, and participle) can be included in a NU, as they do not convey informations about Tense, Aspect and Mood.

The list below provides several examples of NUs retrieved in our dataset:

5. <NU> *bella intervista complimenti* </NU> [*Nice interview congrats*]

6. <NU> *forza salvini* </NU> *non pensare a sti dementi* [*go Salvini don't think about these idiots*]

7. <NU> *denunciare e sospendere il magistrato* </NU> [*to report and to suspend the magistrate*]

The second annotation layer recognized particular NUs with a slogan-like form, with a binary value (yes-no). As noticed in Section 2.3, an utterance is a slogan because of its purpose. Labeling an utterance produced by an anonymous user as a slogan is not a trivial or straightforward task, even if it is pretty simple to recognize political slogans created by politicians. Inter-annotator agreement for this level is 0.65 in terms of Cohen's Kappa, showing that recognizing slogans involves some form of subjective interpretation. Below we report examples of slogans in our dataset:

8. <NU> *L'Italia agli Italiani* </NU> [*Italy to Italians*]

9. <NU> *Orgogliosi della propria identità* </NU> [*Proud of our identity*]

10. <NU> *Forza Salvini* </NU> [*Go Salvini*]

The third layer has been applied only to those items previously annotated as slogans by both annotators, distinguishing between top-down and bottom-up slogans. Top-down slogans are created by the political leader or party, while fans spontaneously produce bottom-up slogans. Annotators reached a better agreement on this distinction (0.74 Cohen's Kappa). One example for each category is reported below:

11. <NU> *Porti chiusi* </NU> [*Closed harbors*] [top-down]

12. <NU> *Forza capitano* </NU> [*Go captain*] [bottom-up]

As illustrated by example 12, bottom-up slogan-like NUs tend to be semantically close to encouragements and cheers that characterize sports competitions. They generally do not convey complex meanings but endorse the leader's message; they are phatic expressions with a clear social function (Jacobson, 1960).

As Table 3 illustrates, these NUs are predominant in the annotated dataset. Not surprisingly, the set of top-down slogans annotated is smaller than the set of bottom-up

FB page	NUs	Top down slogans	Bottom up slogans
Salvini	926	22	204
Di Maio	369	5	57

Table 3: Eligible NUs after preprocessing

slogans: politicians' staff produce few slogans to communicate the politician's message. On the other hand, supporters use a broader set of NUs.

The fourth level of annotation explicitly targets proto-slogans, with an inter-annotator agreement of 0.63: several slogan-like NUs (*in alto i cuori* (*lift up your hearts*), *sempre e per sempre* (*forever and ever*)) are not proto-slogans because they are hapax in the list of centroids and lack of specific content. We recognize as proto-slogans the following NUs:

- <NU> *via i ladroni* </NU> [*away the robbers*]
- <NU> *m5s tutta la vita* </NU> [*m5s for the rest of my life*]

In Table 4 we report how many NUs have been labeled as proto-slogans. Bottom-up NUs are proto-slogans when they express a positive or negative stance towards a discourse target (always explicitly mentioned), whose identity is common knowledge for the electoral base.

Source	Bottom-Up NUs	Proto-slogans
Salvini	196	102
Di Maio	57	25

Table 4: Proto-slogans after annotation

Comparing the bottom-up slogans and proto-slogans produced by the users to those produced by the politicians, it is clear that Salvini uses both these kinds of slogans very frequently, while Di Maio generally uses only top-down slogans. Salvini often uses bottom-up slogans such as *avanti tutta* (*full steam ahead*), which appears three times a week and it is also frequently used by Salvini's followers in the comments, often preceded by a proto-slogan such as *forza Matteo* (*go Matteo*). However, the slogans most used by Salvini, appearing at least once a day, are two proto-slogans, both with a positive stance towards Italy or Italians: *prima l'Italia* (*Italy first*) and *prima gli Italiani* (*Italians first*). These proto-slogans are not used in the comments by Salvini's followers, unlike top-down slogans such as *porti chiusi* (*closed harbors*). Thus, *prima l'Italia/gli Italiani*, while it conveys a political stance and it is used by a political leader, does not act as a top-down slogan.

Therefore, we may suppose that these proto-slogans act like a turn in an ongoing dialogue between Salvini and his followers, both of them expressing their support to each

other through proto-slogans: Salvini expresses a positive stance towards his followers, who in return express their support to him through proto-slogans such as *forza Matteo* (*go Matteo*).

Furthermore, Salvini refers to his followers as "Italians" using a very common populist strategy that identifies populist voters with "the people" and, in this case, with the Italian population as a whole. In this way, Salvini identifies his electorate with the Italian population, giving the impression of a much larger voter base and giving his electorate the perception that they are the real Italians, while their opponents are not.

4 Facebook and Twitter Data Comparison

Slogan-like NUs are specific linguistic items for a political community, when supporters use them. However, they display different frequency patterns over time, i.e., they emerge as more frequent in a specific period. Therefore, the relationship between the frequencies of bottom-up slogans on social media and proto-slogans need a more complex investigation based on more data.

We propose a qualitative classification of slogan-like NUs complementary to proto-slogans' characterization. In order to investigate this aspect, after extracting and annotating nominal utterances from Facebook public pages, the list of NUs was searched on Twitter with the help of GetOldTweets3 Python library in three different one-week time spans across 3 years (2019, 2018, 2017).[4] The aim of this analysis is the identification of three types of slogan-like NUs:

- Generic slogan-like NUs: nominal utterances whose content does not directly concern populism or are specifically related to the leaders. They can not be proto-slogans;

- Attested slogan-like NUs: specific to populist messages concerning Di Maio and Salvini, some attested slogan-like NUs are frequently used, but their presence varies through different periods. They tend to be proto-slogans, especially if they are bottom-up;

- Episodic slogan-like NUs: these NUs are linked to a specific event or period. However, they could still emerge as attested NUs if their use continues beyond a specific period. More data are needed to decide if they are proto-slogans or not.

Table 5 presents three examples with their frequencies in the different periods.

The presence of slogan-like NUs varies depending on their bottom-up or top-down nature. Facebook slogan-like NUs are mostly bottom-up and generally composed of encouragements to the party or, more often, to the leader. They usually display a very familiar and affectionate tone, referring to the leader by his first name. This behavior is

[4] The exact periods for each collection round are: 2019-11-20/2019-11-27, 2018-11-20/2018-11-27, and 2017-11-20/2017-11-27)

Examples	NU type	2017	2018	2019
sempre avanti (always ahead)	generic	115	108	104
avanti capitano (come on captain)	attested	4	45	30
#26maggiovotolega (#26mayIvoteLega)	episodic	1	0	0

Table 5: Types of NUs on Twitter

coherent with the perceived intimacy of Facebook communication, which makes leaders seem more approachable.

On Twitter, top-down slogans are more productive (see examples in Table5) and with longer lifespans, primarily if they are not referred to a specific event, being instead relevant in a more general way. Thus, top-down slogans usually are attested slogan-like NUs or episodic slogan-like NUs.

A top-down episodic slogan made for the European election like *#domenicavotolega* (*#sundayIvoteLega*) is well-attested several months later, probably because it is still relevant for the next Italian Regional elections, planned on a Sunday too. Similarly, the generic, encouraging hashtag *#iostoconsalvini* (*#Istaywithsalvini*), an attested slogan, has been productive in every period considered. On the contrary, the more specific and episodic *#26maggiovotolega* (*#26mayIvoteLega*) is significantly less used after the European elections. Twitter displays some of LN's and M5S's main leitmotifs: the slogan-like NUs *porti chiusi* (*closed harbors*) and *tutti a casa* (*everybody home*). In 2018, *porti chiusi* had been used often to answer Matteo Salvini's tweets, while in 2019 appeared more frequently in free-standing tweets. *Porti chiusi* is an example of an attested slogan-like NU that is distinctive for a political community but can also be used to address this community, criticizing its members. Bottom-up slogan-like NUs are generally present on Twitter, but they show some peculiar differences from those on Facebook. Firstly, particularly familiar generic slogan-like encouragements like *forza matteo* (*go matteo*), very frequent on Facebook, are rare on Twitter, and they never appear in answers to Matteo Salvini's tweets. The less informal *forza salvini* (*go salvini*), *avanti capitano* (*come on captain*) and *forza capitano* (*go captain*) are far more frequent. Still, while on Facebook, they were placed inside the private echo chamber of the leader's page. They do not appear in answers to Matteo Salvini's tweets on Twitter, but they are characteristic of independent tweets. Most of the bottom-up generic slogan-like NUs, like *noi tutti con te* (*all of us with you*), are not attested on Twitter, but there are a few notable exceptions, such as *avanti tutta* (*full steam ahead*), *sempre avanti* (*always forward*) or *vergogna* (*shame*).

However, this investigation is still preliminary since it has not been possible to ensure that tweets with bottom-up generic slogan-like NUs, such as *forza capitano* (*go captain*), are unquestionably referred to LN. If the user explicitly mentions the politician, disambiguation is possible. Otherwise, the tweet could be used to support a football team.

5 Conclusions and Future Work

Political communication on social media can be investigated with real data available on Twitter and Facebook public pages. This paper introduces the concept of proto-slogan as an economical device used to build and reinforce the in-group sense of belonging in online political communities. We introduced a methodology for identifying NUs peculiar to a political community on social media. These NUs extracted from centroids, derived from the Facebook public page of Matteo Salvini and Luigi Di Maio, are often slogan-like.

The political party or leader creates top-down slogans, and they are generally more linked to the party's program. Instead, the supporters produce bottom-up slogans, which we define as proto-slogans, and they are usually less specific and more linked to informal encouragements.

Recognizing these slogan-like NUs makes it possible to recognize supporters of a specific populist political party, even when their messages are not otherwise contextually linked to it. Even if less specific, bottom-up slogan-like NUs are still recognizable on Twitter. They can uncover political support without explicit political content. However, refining the automatic recognition of NUs is still necessary since informal computer-mediated communication typically shows a substandard variety of Italian. For example, some verbs in the finite form may appear inside a NU, since they have a non-standard spelling.

Our analysis represents the first step toward identifying stylometric patterns in the populist electorate's informal writing on social media. We aim to characterize political affiliation in language even when explicit political themes are not mentioned. It would be advisable to remind that this kind of author profiling could have some ethical issues, but the final goal would not be monitoring opinions expressed on the web.

Instead, we believe that public and open research on these topics would be helpful to show and make transparent for everyone what commercial systems - that often do not share their approaches with the scientific and the civil communities - can do with publicly available data.

References

Alnajjar, K., & Toivonen, H. (2020). Computational Generation of Slogans. *Natural Language Engineering*.

Amălăncei, B.-M., Cîrtiţă-Buzoianu, C., & Daba-Buzioianu, C. (2015). Looking for the best slogan: an analysis of the slogans of the 2016 Romanian parliament campaign. *Studies and Scientific Researches. Economics Edition*, *26*.

Aslanidis, P. (2016). Is populism an ideology? A refutation and a new perspective. *Political Studies*, *64*.

Benveniste, É. (1990). *Problemi di linguistica generale.* Milano, Italia: Mondadori.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python* (1st ed.). O'Reilly Media, Inc.

Bobba, G. (2019). Social media populism: features and 'likeability' of lega Nord communication on facebook. *European Political Science*, *18*.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.

Comandini, G., & Patti, V. (2019). An impossible dialogue! Nominal utterances and populist rhetoric in an Italian Twitter corpus of hate speech against immigrants. In *Proceedings of the third workshop on abusive language online.*

Comandini, G., Speranza, M., & Magnini, B. (2018). Effective communication without verbs? Sure! Identification of nominal utterances in italian social media texts. In *Proceedings of the fifth italian conference on computational linguistics (clic-it 2018), torino, italy, december 10-12, 2018.* (Vol. 2253). CEUR-WS.org.

Cresti, E. (1998). Gli enunciati nominali. In M. T. Navarro (Ed.), *Italica matritensia: atti del IV convegno SILFI Società internazionale di linguistica e filologia italiana (Madrid, 27-29 giugno 1996).* Firenze: Cesati.

Eisenstein, J., O'Connor, B. T., Smith, N. A., & Xing, E. P. (2014). Diffusion of lexical change in social media. *PLoS ONE*, *9*.

Engesser, S., Fawzy, N., & Larsson, A. O. (2017). Populist online communication: introduction to the special issue. *Information, Communication and Society*, *20*.

Ernst, N., Engesser, S., Buchel, F., Blassnig, S., & Esser, F. (2017). Extreme parties and populism: an analysis of facebook and twitter across six countries. *Information Communication and Society*, *20*.

Ernst, N., Esser, F., Blassnig, S., & Engesser, S. (2018). Favorable opportunity structures for populist communication: Comparing different types of politicians and issues in social media, television and the press. *The International Journal of Press/Politics*, *24*.

Ferrari, A. (2011a). Enunciati nominali. *Enciclopedia dell'Italiano*. Retrieved from `http://www.treccani.it/enciclopedia/enunciati-nominali_(Enciclopedia_dell'Italiano)/` (`http://www.treccani.it/enciclopedia/enunciati-nominali_(Enciclopedia_dell'Italiano)/`)

Ferrari, A. (2011b). Stile nominale. *Enciclopedia dell'Italiano*. Retrieved from `http://www.treccani.it/enciclopedia/stile-nominale_(Enciclopedia-dell'Italiano)/` (`http://www.treccani.it/enciclopedia/stile-nominale_(Enciclopedia-dell'Italiano)/`)

Ferrier, A. (2014). *The advertising effect. How to change behaviour.* South Melbourne: Oxford University Press University Press.

Hameleers, M., Reinemann, C., Schmuck, D., & Fawzi, N. (2019). The Persuasiveness of Populist Communication. Conceptualizing the Effects and Political Consequences of Populist Communication From a Social Identity Perspective. In C. Reinemann, J. Stanyer, T. Aalberg, F. Esser, & C. H. de Vreese (Eds.), *Communicating Populism. Comparing Actor Perceptions, Media Coverage, and Effects on Citiziens in Europe.* New York and London: Routledge.

Huguet Cabot, P.-L., Abadi, D., Fischer, A., & Shutova, E. (2021). Us vs. Them: A Dataset of Populist Attitudes, News Bias and Emotions. In *Proceedings of the*

16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. Online: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/2021.eacl-main.165

Jacobs, K., & Spierings, N. (2016). *Social media, parties, and political inequalities.* Basingstoke: Palgrave Macmillan.

Khalid, O., & Srinivasan, P. (2020). Style matters! investigating linguistic style in online communities. In *nternational aaai conference on web and social media.*

Kriesi, H. (2014). The populist challenge. *West European Politics*, *37*(2). Retrieved from https://doi.org/10.1080/01402382.2014.887879 doi: 10.1080/01402382.2014.887879

McCulloch, G. (2019). *Because internet: Understanding the new rules of language.* Riverhead Books.

Merchant, J. (2005). Fragments and ellipsis. *Linguistics and Philosophy*, *27*.

Mudde, C. (2004). The Populist Zeitgeist. *Government and Opposition*, *39*(4).

Oliver, J. E., & Rahn, W. M. (2016). Rise of the trumpenvolk: Populism in the 2016 election. *The ANNALS of the American Academy of Political and Social Science*, *667*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*.

Rieder, B. (2013). Studying Facebook via data extraction: the Netvizz application. In *Proceedings of the 5th Annual ACM Web Science Conference.* New York, NY, USA: ACM.

Rooduijn, M., & Akkerman, T. (2017). Flank attacks: Populism and left-right radicalism in western europe. *Party Politics*, *23*.

Schmid, H. (1995). Improvements in part-of-speech tagging with an application to German. In *Proceedings of the ACL SIGDAT-Workshop.*

Smith, E. R., & Mackie, D. M. (2008). Intergroup emotions. In L. F. B. M. Lewis J. M. Haviland-Jones (Ed.), *Handbook of emotions.* New York: Guilford.

Stanyer, J., Salgado, S., & Strömbäck, J. (2016). Populist actors as communicators or political actors as populist communicators: Cross-national findings and perspectives. In T. Aalberg, F. Esser, C. Reinemann, J. Strömbäck, & C. H. de Vreese (Eds.), *Populist Political Communication in Europe.* New York: Routledge.

Stieglitz, S., & Dang-Xuan, L. (2012). Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, *3*.

Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018). Election campaigning on social media: Politicians, audiences, and the mediation of political communication on facebook and twitter. *Political Communication*, *35*.

Tran, T., & Ostendorf, M. (2016). Characterizing the language of online communities and its relation to community reception. In *Emnlp.*

Tredici, M. D., & Fernández, R. (2018). The road to success: Assessing the fate of linguistic innovations in online communities. *ArXiv*, *abs/1806.05838*.

Westen, D. (2007). *The political brain. the role of emotion in deciding the fate of the nation*. New York: PublicAffairs.

Zappavigna, M. (2015). Searchable talk: the linguistic functions of hashtags. *Social Semiotics*, *25*.

Correspondence

Irene Russo
Institute for Computational Linguistics "A. Zampolli" (ILC-CNR)
University of Pisa
irene.russo@ilc.cnr.it

Gloria Comandini
Italian Institute of German Studies
University of Trento
gloria.comandini@unitn.it

Tommaso Caselli
Center for Language and Cognition
University of Groningen
t.caselli@rug.nl

Viviana Patti
Department of Computer Science
University of Turin
patti@di.unito.it