'We the People' in 140 Characters:
Populism in Official and Unofficial Electoral Campaign Tweets

Bruno Castanho Silva
(Central European University)

Paris Aslanidis
(University of Macedonia)

Introduction

An often mentioned aspect of populism is the immediate, direct connection between leader and followers. However, as this is a very abstract idea, empirical studies have a hard time assessing and measuring it. At the same time, the advent of new media that use the internet as their backbone has furnished novel tools and platforms to bring agents and audiences closer together. We suggest that social media is one place where it is possible to observe a more direct relation between a politician and his or her supporters. Therefore, if it is possible to identify populist discourse in this kind of media, it would also be possible to study more thoroughly how the public reacts to it. Moreover, by looking at content published by supporters, we would be closer to seeing how populist discourse is produced and reproduced in a bottom-up perspective, enriching demand-side studies of the topic.

Research questions

Can we identify populist discourse in Twitter data?
*Can we employ Twitter as a tool to classify parties as populist?*

Method

Clause-based semantic text analysis was employed to extract quantitative data from political text. This method takes advantage of structural features of written and spoken language, focusing on the invariable (Subject)-Verb-(Object) grammatical structure of clauses in order to produce a dataset of political language where SVO "triplets" become the *coding unit*, enhancing reliability and providing increased resolution (Franzosi 2004; Popping 2000; Roberts 1997).

- Population: Tweets from three official accounts (@dilmabr, @atsipras, @EvoMPresidente).
- *STEP 1*: Data collection - Tweets downloaded from official accounts using R packages *twitteR* and *streamR*.
- *STEP 2*: Sampling - 100 tweets selected randomly from each twitter account within specific time-frame (last national election campaign).
- *STEP 3*: Text preparation - Each tweet was split into *n* clauses of (Subject)-Verb-(Object) format, employing rewrite grammar rules where applicable.
- *STEP 4*: Coding - Resulting set of clauses coded using coding scheme comprised of 8 codes to capture "people-centrism" (U01), "anti-elitism" (T01), and their combination (V01). Residual codes did not contribute to scoring.

Costs

- *Data collection*: 1 hour per twitter account, 3 hours in total
- *Text preparation*: 2.5 hours per twitter account, 7.5 hours in total
- *Coding*: 4 hours per twitter account, 12 hours in total

Histograms with the percentage of populist clauses in each candidate 100-Tweets sample, and wordclouds from these tweets
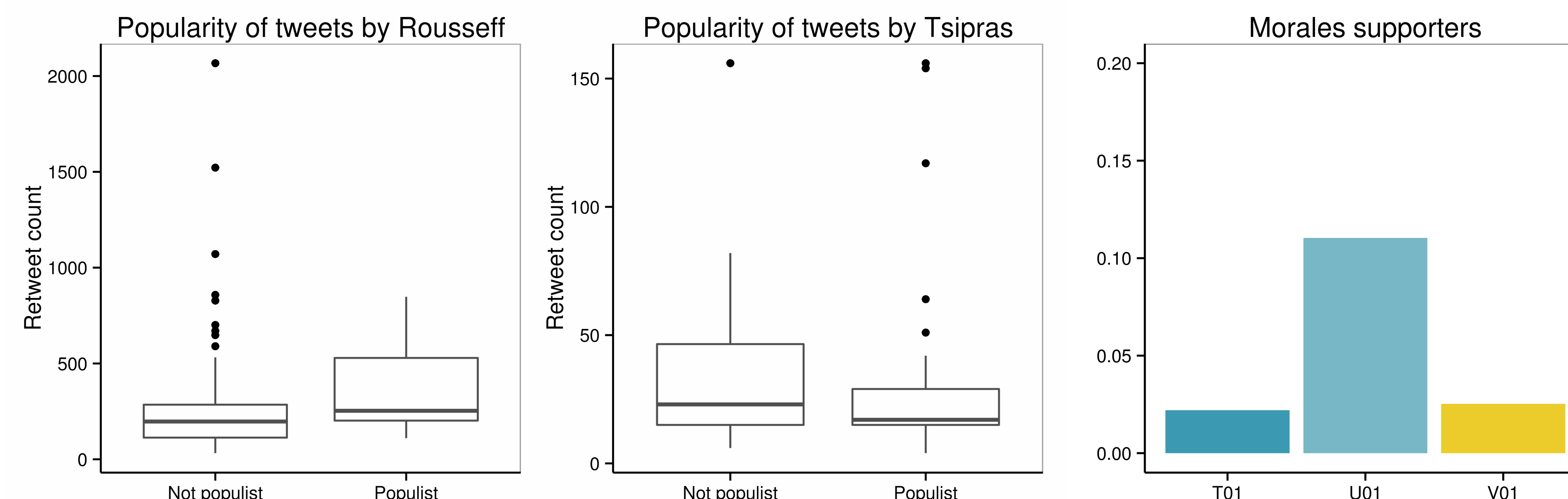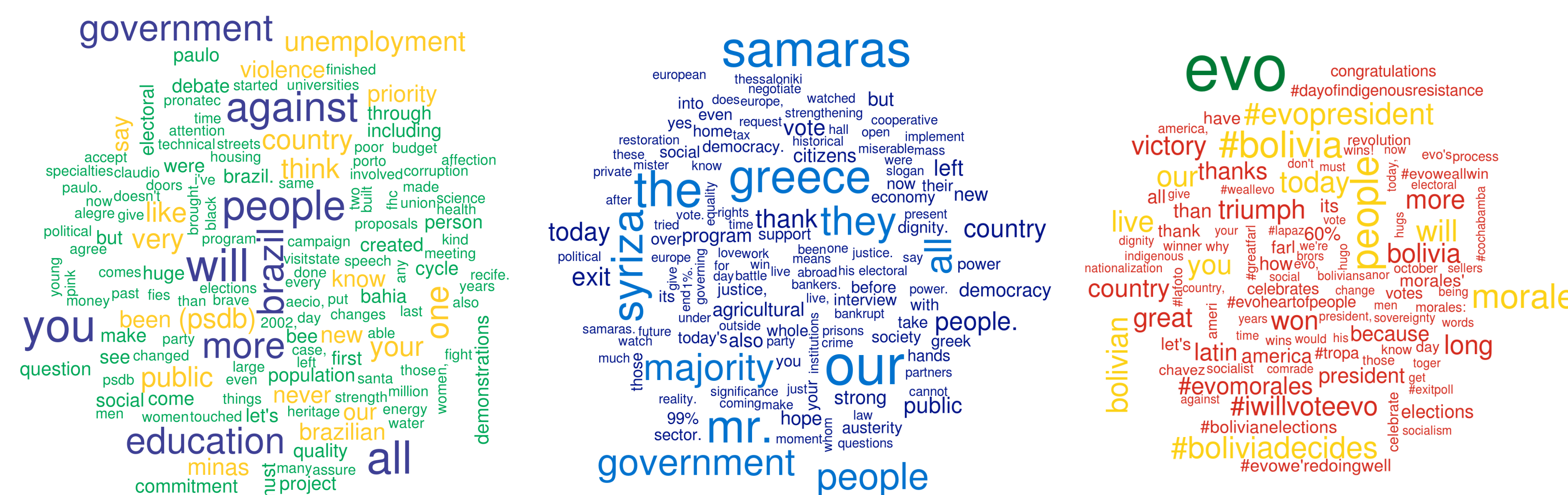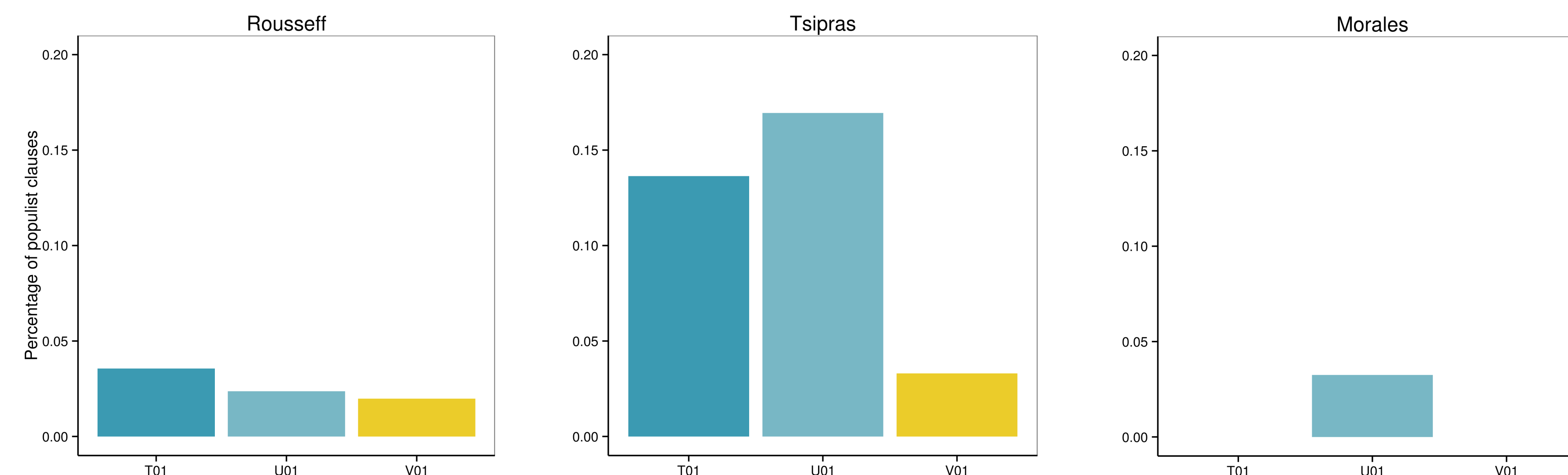
Number of retweets of each status update by Rousseff and Tsipras, divided by whether tweet contained a clause coded as U01, T01 or V01 (populist) or not (non-populist)

Histogram with percentage of populist clauses in 200 tweets by supporters of Morales

Sample

All tweets posted during last electoral campaign in Bolivia, Brazil, and Greece, by personal accounts of the three winning candidates: re-elected presidents Evo Morales and Dilma Rousseff, and Syriza leader Alexis Tsipras. Also parsed tweets from supporters of Morales, identified by use of specific hashtags. Parsing was done through the Twitter API, which claims to give a 1% representative sample of total statuses posted with the search term. The sample covers election day – October 12, 2014. Both replies and retweets were excluded for all samples. Total number of tweets was 2167 from Morales, 1119 from Rousseff, and 500 from Tsipras. Sample of partisans' tweets contains 19,839 from Morales' supporters. A 200-tweets sample was randomly selected from these.

These three leaders were chosen for substantive and practical reasons: (a) at least two are well-known cases of populism, and we expect to find a degree of populism in their discourse, (b) Twitter API limits access to 3200 most recent status entries for each user, meaning we were constrained to recent elections.

Tentative conclusions

a. Identifying populist discourse in Twitter data using human-based coding is feasible, even though context is largely missing.
b. Level of populism is lower compared to traditional data sources (e.g. speeches, manifestos). Morales, especially, exhibits very little 'twitter populism'. *Possible reasons*: medium limitations (e.g. 140 chars), audience nature, usage as informational channel.

Experiences with other methods

We also tried using computer-assisted content analysis (CATA), mostly with negative results.

- *Topic models*: Commonly used topic models algorithms, such as Latent Dirichlet Allocation, were used on tweets by the leaders. No meaningful dimensions were identified from the terms in each topic - neither for populism or other concepts.
- *Dictionary-based word associations*: small dictionary with populism-related terms (e.g. *people, elite, U.S., names of opposition parties*) was created and employed to look for word associations frequencies with each term. No meaningful results obtained.

References

Franzosi, Roberto. (2004). *From Words to Numbers: Narrative, Data, and Social Science*. Cambridge: Cambridge University Press.
Popping, Roel. (2000). *Computer-Assisted Text Analysis*. London: Sage
Roberts, Carl W. (Ed.). (1997). *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences From Texts and Transcripts*. Mahwah, NJ: Lawrence Erlbaum Associates.