

Beyond Sentiment: Social Psychological Analysis of Political Facebook Comments in Hungary

**Márton Miháltz,
Tamás Váradi**

Research Institute
for Linguistics, Hungarian
Academy of Sciences
Benczúr utca 33, Budapest
H-1068 Hungary
mmihaltz@gmail.com
varadi.tamas
@nytud.mta.hu

**István Csertő, Éva Fülöp,
Tibor Pólya**

Institute of Cognitive
Neuroscience And Psychology,
Hungarian Academy of Sciences
Magyar tudósok körútja 2.
Budapest, H-1117 Hungary
{cserto.istvan,
fulop.eva,
polya.tibor}
@ttk.mta.hu

Pál Kóvágó

Institute of Psychology,
University of Pécs
Ifjúság útja 6, Pécs
H-7624 Hungary
kovago.pal
@ttk.mta.hu

Abstract

This paper presents the methodology and results of a project for the large-scale analysis of public messages in political discourse on Facebook, the dominant social media site in Hungary. We propose several novel social psychology-motivated dimensions for natural language processing-based text analysis that go beyond the standard sentiment-based analysis approaches. Communion describes the moral and emotional aspects of an individual's relations to others, while agency describes individuals in terms of the efficiency of their goal-orientated behavior. We treat these by custom lexicons that identify positive and negative cues in text. We measure the level of optimism in messages by examining the ratio of events talked about in the past, present and future by looking at verb tenses and temporal expressions. For assessing the level of individualism, we build on research that correlates it to pronoun dropping. We also present results that demonstrate the viability of our measures on 1.9 million downloaded public Facebook comments by examining correlation to party preferences in public opinion poll data.

1 Introduction

Social media (SM) is becoming an increasingly important channel for communications in politics. In Hungary, Facebook is the dominant SM platform, with 4.27M registered Hungarian users (59.2% penetration of 7.2M people with internet access, which represents 43% of the total population)¹. No political party or politician can afford to miss the opportunity of extending their influence by regularly publishing status update messages (posts) on their Facebook pages that are potentially accessible by all Facebook users (i.e., marked as “public”). Most political actors enable discussions (commenting) on their pages, which means other users are able to publicly respond to (post comments about) the original posts or to each other's responses. This constitutes a vast and vivid source of political or politics-inspired discussions, debates, expressions of sentiment, support or dissent. Most importantly, the participating social media users also happen to be real-life voters.

In this paper, we present a set of tools and resources that enable the collection and analysis of Hungarian public Facebook comments written in response to public posts published on the pages of Hungarian politicians and political organizations. Besides the identification of relevant entities and sentiment polarity in these messages, our investigations focused on methods for detecting and quantifying psychological and sociopsy-

¹ Source: <http://www.internetworldstats.com/europa.htm>

chological phenomena including *agency and communion*, *optimism/pessimism* and *individualism/collectivism*. These indicators are based on previous results in the area of social psychology research. The main contribution of this paper is the proposal of these new, social psychology-motivated dimensions for the analysis of attitudes expressed in social media that go beyond the standard sentiment (emotional polarity) analysis approaches. With these we hope to get better answers to questions like: what are the trends in the reactions to SM messages of political actors, and how do these correlate to real-life political actions and events, such as elections and votes? How do political communication and discussions shape the psychological states and social values of various SM user groups, such as supporters of political powers?

The rest of this paper is organized as follows: the next section presents our data sources and the methods of our analysis with respect to social psychology and the challenges presented by processing social media language. We then present preliminary empirical results that demonstrate the viability of our proposed approach by examining correlation with a real-life political event, general elections in Hungary in 2014.

2 Methods

There were three major election events in Hungary in 2014: general elections for seats in the National Assembly (Hungarian Parliament) in April, elections for seats in the European Parliament in May, and municipal elections in October. In order to focus on the debates surrounding these events, we collected the names of nominated and elected representatives and their political organizations involved in these campaigns (sources used: valasztas.hu (official election data) and Hungarian Wikipedia.) Using these names, we identified 1341 different Facebook pages that belong to Hungarian political organizations (parties, their regional and associated branches etc.) and politicians (candidates and elected representatives etc.) of years 2013 and 2014. We used both official pages (administered by the agents of the political actors the pages are about) and fan pages (administered by independent communities).

We used the Facebook Graph API to collect public posts and the associated public comments from these sources dated between October 2013 and September 2014 once a week. One week after each harvest, another script was used to check

for new comments that arrived to already downloaded posts. In total, our corpus contains 141K Facebook posts and 1.9 million comments from 226K users, constituting over 46 million running words.

In order to be able to analyze sentiment and the other sociopsychological measures we had to first process the comment messages using the following pipeline: segmentation and tokenization, morphological analysis, part-of-speech tagging and lemmatization. This was followed by the extraction of relevant entities and the identification of their party affiliations using custom lexicons compiled into finite state automata using the open source *NooJ*² tool (Silberstein 2005), which uses finite state automata compiled from custom grammars to annotate text. We also used NooJ to annotate expressions of sentiment and other sociopsychological phenomena using custom lexicons and grammars. The following sections give details about the background and development of these components.

2.1 Social Psychological Analysis

Scientific Narrative Psychology (SNP) is a complex approach to text-based theory building and longitudinal, quantitative assessment of psychological phenomena in the analysis of self and group narratives in the field of social, personality and clinical psychology (László 2008). Our methods for the sociopsychological content analysis of Facebook comments in this project builds on earlier work extending it to the domain of SM discourse in politics.

Current approaches to analyzing attitudes in social media messages mainly focus on one psychological viewpoint, emotional polarity (sentiment analysis), as in (Ceron et al. 2014; Chen et al. 2010; Costa et al. 2015; Hammer et al. 2014; O'Connor et al. 2010; Sobkowicz, Kaschesky and Bouchard 2012; Tumasjan et al. 2010). In addition to also measuring sentiment, we extend this framework by proposing several new aspects that offer insights into further psychological and social phenomena which might be important indicators in the assessment or prediction of the attitudes and behavior of online communities. In addition, while the majority of previous work on analyzing social media focuses on Twitter, such as (Asur et al. 2010; Balahur 2013; Costa et al. 2015; Habernal et al. 2013; Kouloumpis et al. 2011; Lietz et al. 2014; Sidorenko et al. 2013; Tumasjan et al. 2010), we use public comments

² <http://nooj4nlp.net/>

from Facebook, which is the dominant SM channel in Hungary.

According to social psychology theory social value judgments can be described along two lines (Abele and Wojciszke 2007; Abele et al. 2008). *Agency* describes an individual in terms of the efficiency of their behavior oriented to their personal goals (motivation, competence and control). *Communion* describes the moral and emotional aspects of an individual's relations to other group members, individuals or groups (cooperation, social benefit, honesty, self-sacrifice, affection, friendship, respect, love etc.). Both types have positive and negative dimensions, and can be used to describe social behavior, for example when relating to a political organization one supports or opposes. Our *agency and communion* NooJ component annotates comments using lexicons that contain 650 different expressions.

Individuals differ in the way past, present or future events dominate their thinking. When a person's thinking is dominated by the past, they are likely to view the world unchangeable. Thinking dominated by the present indicates the importance of realistically attainable goals, while future-dominated thinking usually sees open possibilities. We assume that optimistic people tend to talk more about the future and less about the past, while pessimists talk more about the past and less about the present, which is supported by previous studies (Habermas et al. 2008; Kunda 1999). This bears significance in situations when a person may choose to focus on any of the three temporal aspects. As an example, when making political decisions one might focus on either prior events leading up to the decision, on carrying out the decision itself or on the implied future consequences. Our NooJ grammar for *optimism/pessimism* annotates expressions of time using morphological information (verb tenses) and by recognizing some temporal expressions. Based on these, we calculated an optimism indicator (higher value, higher degree of optimism) using the ratio of present and future expressions to all expressions (past, present, future).

Individualism represents the importance of the category of the self when thinking about the world: individualistic societies keep the actions of the individual in focus, while collectivist societies focus on the actions of groups. Studies have shown a correlation between the usage/omission of personal pronouns (pronoun drop) and the levels of individualism in societies (Kashima and Kashima 1999). We extend this idea by assuming that pronoun drop can be used

to compare the level of individualism between groups within a society as well. Our *individualism/collectivism* NooJ grammar relies on only part-of-speech and morphological information to annotate personal pronouns and verbs or nouns with personal inflections. By calculating the ratio between the former and the latter we estimated the rate of actually versus potentially appearing pronouns which yielded a measure of individualism (higher score indicating higher degree of individualism).

We also measured *sentiment* in the comments by means of a NooJ automaton we created for the annotation of positive and negative emotions using a lexicon of 500 positive and 420 negative nouns, verbs, adjectives, adverbs, emoticons and multi-word expressions. It also uses a number of rules to treat elements of context that might affect polarity (e.g. negation).

To facilitate the creation of the custom lexicons for the NooJ grammars above, we created a sample corpus that contains 176K comments from 569 different Hungarian politics-related Facebook pages, totaling 5.45M words. The corpus was analyzed using our standard NLP tools. The lexicons for sentiment, agency and communion were constructed by 6 independent human annotators who coded words in the sample corpus that occurred with a frequency of 100 or more (about 3500 total) for each category. In the cases where at least 4 annotators agreed, a seventh annotator made the final decision.

2.2 Adapting NLP Tools to Social Media

All of the NLP tools that were used for preprocessing the comments were developed for a linguistic domain (using standard language texts, mostly newswire) that is different from the language used in Facebook comments. The latter has a high tendency for phenomena like typos and spelling errors, non-standard punctuation use, use of slang expressions, emoticons and other creative uses of characters, substitution of Hungarian accented characters by their unaccented variants etc. For this reason, our readily available tools suffered from degradation in performance. To overcome this problem, we employed a two-fold approach: we applied normalizations to the input and also extended our tools to adapt them to the SM language domain.

To properly investigate the problems arising from processing SM language, we created a corpus of 1.2 million Facebook comments (29M running words total), which was analyzed by the vanilla NLP tools. Unknown types with a fre-

quency of 15 or higher (about 14,000 types) were manually inspected (with reference to their contexts) to yield an overview of common problems that showed regularity and lists of unknown, frequent and important words and expressions.

Based on these findings, our tokenizer was augmented by pre- and post-processing routines that resolved some of the issues arising from non-standard use of spaces and punctuation. We also used lists to normalize commonly misspelled words and slang expressions. Unknown but frequent and important words were added to the morphological analyzer’s lexicon using analogous known words in the same morphological paradigms, which enabled the analysis of arbitrary inflected forms of these words.

For the identification of relevant named entities (names of persons (politicians) and organizations (parties)) we tested a maximum entropy classifier tool trained to resolve Hungarian named entities (Varga and Simon 2007). However, because of its low performance on Facebook comments, we made a decision to use custom, domain-specific, lexicon-based NE recognition, which relies on names, name variants, nicknames and party affiliations of relevant political actors collected from the development corpus described above.

2.3 Evaluation

In order to evaluate the reliability of our named entity, sentiment and social psychological annotations, we constructed two gold standard sets of 336 and 672 Facebook comments. Each set contained messages from all political parties’ Facebook pages in the same distribution as in the complete 1.9M comment corpus (FIDESZ-KDNP 25.2%, EGYÜTT-2014 19.3%, JOBBIK 19.2%, MSZP 16.6%, DK 12.5%, PM 4.2%, LMP 2.9%). In the smaller set, three human annotators annotated each comment for the political affiliations of named entities (persons and organizations), while in the larger set they identified expressions of sentiment, agency, communion and the linguistic markers used by our optimism and individualism measures. *Table 1* shows the results of evaluating the annotations produced by our system against these gold standards.

Annotation Type	Precision	Recall	F1
Party affiliations	98.36	57.14	72.29
Pos. sentiment	82.56	74.50	77.38
Neg. sentiment	67.03	53.68	59.62
Pos. agency	70.59	69.43	52.83
Neg. agency	65.79	25.51	36.76
Pos. communion	65.75	38.40	48.48
Neg. communion	96.39	41.45	57.97
Individualism: pers. pronouns	35.20	65.63	45.82
Individualism: inflections	77.27	94.74	85.12
Optimism: past	78.90	93.97	85.78
Optimism: present	31.40	92.54	46.88
Optimism: future	32.80	67.03	44.04

Table 1: Evaluation of annotations against the gold standards

The results show that while the performance of the annotations of party affiliations and sentiment, agency and communion expression are generally acceptable, there are serious issues with the annotations of linguistic markers for individualism and optimism. Preliminary investigations revealed problems with the manual coding of some markers in the gold standard sets, we are currently working on identifying these issues in order to be able to re-annotate the gold standard and repeat evaluation measurements for our optimism and individualism indicators.

We also evaluated the performance of sentiment analysis based on our sentiment annotations. We assigned a sentiment score to each sentence in each comment by subtracting the number of negative sentiment expressions from the number of positive sentiment expressions, and normalized by the number of words in the sentence. We then mapped this score to a 3-value sentiment polarity indicator: -1 if the sentiment score was negative, 0 if it was 0 (neutral), or 1 if it was positive. We also calculated sentiment polarity for each sentence in each comment in the gold standard set to compare against our automatically obtained polarity indicators (Table 2). Our system performed well above the baseline method, which worked by assigning the most frequent polarity value (neutral) observed in the gold standard.

	#Sent.	Acc.
All sentences in GS	1295	-
Neutral polarity (baseline)	920	71.04
Correctly identified polarity	1096	84.63

Table 2: Accuracy of sentiment polarity recognition

3 Experiments

We conducted several experiments to test the viability of our proposed agency, communion, optimism and individualism measures. We examined how well these could indicate changes in public attitude towards major political powers in the context of the April 2014 parliamentary elections in Hungary. We processed our corpus of 1.9 million Facebook comments using the above tools to calculate scores and indicators for each comment. We grouped comments for each political party’s Facebook pages, aggregated results for each month and compared them to the results of a traditional public opinion survey³. We used monthly party popularity (support) data available from confident voters. Since we did not have any information available about the party preferences of our Facebook comment authors, we operated under the simple assumption that the majority of commenters communicating on a given party’s Facebook page are the supporters of that party. This means that indicators measured from the comments posted at the Facebook page of a given political party were assumed to characterize the attitudes of the supporters of that party.

To assess our optimism and individualism indicators, we first correlated their values to party popularities. We expected that higher degrees of individualism would indicate higher responsibility for party choices, which would imply higher party popularity rates. We found nearly significant positive correlation for individualism ($r=.22$, $p=.052$). However, for our optimism indicator we measured negative correlation ($r=-.22$, $p=.055$) with party popularity, which did not support our hypothesis that a higher rate of optimism would indicate a higher ability to make party choices. This might be explained by the assumption that past events also play an important role in political preferences.

We also examined how values of our optimism and individualism measures behaved before and after the parliamentary elections in

April 2014. Both indicators showed notable changes in the time period immediately following the elections. Individualism levels increased, which might be explained by the decline of the significance of cooperation and unity after elections within politically organized groups. The levels of optimism also showed a change after the elections, but only increased on pages related to the winning party (FIDESZ-KDNP), and decreased on the pages of all other parties. This might be explained by the different experiences of success and failure: success leads to higher optimism, while defeat leads to decrease in optimism.

Our hypotheses about the relationship between party popularities and agency/communion were based on two observations in social psychology. First, the phenomenon of intergroup bias refers to a pattern when members of a group tend to overrate their own group while devaluing outside groups in situations of intergroup competence or conflict. Second, while people judge members of outside groups primarily through the aspect of communion, they tend to evaluate themselves and other members of their own groups via the aspect of agency. Based on these, we expected to find significant negative correlation between both positive agency and negative communion on the one hand and party popularity on the other: low or decreasing support represents a threatening situation to group identity that leads to compensation manifesting in the overrating of one’s in-group and the devaluation of out-groups.

In the 6-month period preceding the parliamentary elections, we found negative correlation between positive agency (number of identified positive agency expressions normalized by total number of tokens in the comments in the time period) and party popularity ($r=-.429$, $p=.05$). We also found strong negative correlation ($r=-.677$, $p=.05$) between party popularity and agency polarity score (difference of positive and negative agency normalized by sum of positive and negative agency) in the same period. After the elections, while there was no correlation between party popularity and agency, there was a high rate of negative correlation with negative communion ($r=-.574$, $p=.01$) and communion polarity score ($r=-.454$, $p=.05$). This also supported our initial hypothesis: the lower the popularity of a given party the stronger the devaluation of other parties through negative communion linguistic content. This might serve to protect

³ http://www.tarki.hu/hu/research/elect/gppref_table_03.html

threatened identity and build group cohesion in parties with less than expected success.

We also found that average positive agency was higher than average negative agency in the whole time period, the difference being significant ($p=.001$, using Student's t-test). Average negative communion was also significantly ($p=.001$) higher than average positive communion. Looking at the changes between before and after elections the rate of average positive agency showed significant decrease ($p=.01$). This might be linked to the fact that acquiring and keeping power is a more crucial issue in the tense competition before elections than in the subsequent period.

4 Conclusion

We presented our experiments to collect and analyze Facebook comments in Hungarian politics using novel sociopsychological measures that extend the possibilities for the assessment of attitudes expressed in text beyond sentiment analysis. We found that our proposed indicators for agency and communion are valid tools from a psychological perspective and can be useful for detecting changes in opinion on social media sites of political groups. While our individualism and optimism measures showed mixed results, they also show potential to bring new sides to SM text analysis in politics.

All the resources (complete corpus of 1.9M comments with full annotation, ontology of relevant political actors) and the source codes of our tools to process them are available for download⁴.

5 Acknowledgements

We wish to thank the thorough work of the conference reviewers. Special thanks to Anikó Hanák for her helpful suggestions. Work described in this paper was supported by the TrendMiner project, which was co-funded by the FP7-ICT programme of the European Commission (project reference nr. 287863).

References

Abele, A. E. and Wojciszke, B. 2007. Agency And Communion From The Perspective of Self Versus Others. *Journal of Personality and Social Psychology* 93:751-763.

Abele, A. E.; Cuddy, A. J. C.; Judd, C. M.; and Yzerbyt, V. Y. 2008. Fundamental Dimensions of Social Judgment. *European Journal of Social Psychology* 38:1063-1065.

Asur, S.; Huberman, B.A. 2010. Predicting the Future with Social Media. *Web Intelligence and Intelligent Agent Technology* 492-499.

Balahur, A. 2013. Sentiment Analysis in Social Media Texts. *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 120-128.

Ceron, Andrea; Curini, Luigi and M Iacus, Stefano. 2014. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society* 16:340-358.

Chen , Bi, Leilei Zhu, Daniel Kifer, and Dongwon Lee. 2010. What Is an Opinion About? Exploring Political Standpoints Using Opinion Scoring Model. *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence* 1007-1012.

Costa, Jean M. R., Rahmtin Rotabi, Elizabeth L. Murnane and Tanzeem Choudhury. 2015. It is not only about Grievances - Emotional Dynamics in Social Media during the Brazilian Protests. *Proceedings of the Ninth International AAAI Conference on Web and Social Media* 594-597.

Csertő, I., and László, J. 2013. Intergroup evaluation as an indicator of emotional elaboration of collective traumas in national historical narratives. *Sociology Study* 3: 207-224.

Habermas, T.; Ott, L. M.; Schubert, M.; Schneider, B.; and Pate, A. 2008. Stuck in the Past: Negative Bias, Explanatory Style, Temporal Order, and Evaluative Perspective in Life Narratives of Clinically Depressed Individuals. *Depression and Anxiety* 25: 1091-1269.

Habernal, I.; Ptáček, T and Steinberger, J. 2013. Sentiment Analysis in Czech Social Media Using Supervised Machine Learning. *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 65-74.

Hammer, Hugo Lewi, Per Erik Solberg and Lilja Øvrelid. 2014. Sentiment classification of online political discussions: a comparison of a word-based and dependency-based method. *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* 90-96.

Kashima, E. S., and Kashima, Y. 1998. Culture and language: The case of cultural dimensions and personal pronoun use. *Journal of Cross-Cultural Psychology* 29:461-486.

⁴ <http://corpus.nytud.hu/trendminer>

- Kouloumpis, E.; Wilson, T.; and Moore, J. 2011. Twitter sentiment analysis: The good the bad and the omg!. *ICWSM* 11:538-541.
- Kunda, Z. 1999. *Social Cognition. Making Sense of People*. Cambridge, Mass.: MIT Press.
- László, J. 2008. *The Science of Stories: An introduction to Narrative Psychology*. London, New York: Routledge.
- László, J.; Csertó, I.; Fülöp, É.; Ferenczhalmy, R.; Hargitai, R.; Lendvai, P.; Péley, B.; Pólya, T.; Szalai, K.; Vincze, O.; and Ehmann, B. 2013. Narrative Language as an Expression of Individual and Group Identity: The Narrative Categorical Content Analysis. *SAGE Open* 3:1-12.
- Lietz, H.; Wagner, C.; Bleier, A. and Strohmaier, M. When Politicians Talk. 2014. Assessing Online Conversational Practices of Political Parties on Twitter. *International AAAI Conference on Weblogs and Social Media* 285-294.
- O'Connor, B.; Balasubramanyan, R.; Routledge, B. R.; and Smith, N. A. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *ICWSM* 11:122-129.
- Sidorenko, W.; Sonntag, J.; Krüger, N.; Stieglitz, S. & Stede, M. 2013. From newspaper to microblogging: What does it take to find opinions? *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* 81-86.
- Silberstein M. 2005. NooJ: a Linguistic Annotation System for Corpus Processing. *HLT/EMNLP 2005* 10-11.
- Sobkowicz, P.; Kaschesky, M.; and Bouchard, G. 2012. Opinion Mining in Social Media: Modeling, Simulating, and Forecasting Political Opinions in The Web. *Government Information Quarterly* 29: 470-479.
- Tumasjan, A.; Sprenger, T. O.; Sandner, P. G.; and Welpe, I. M. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM* 10:178-185.
- Varga, D.; E. Simon. 2007. Hungarian named entity recognition with a maximum entropy approach. *Acta Cybernetica* 18(2):293-301.