The pragmatics of political messages in Twitter communication

Jurģis Šķilters[1], Monika Kreile[2], Uldis Bojārs[1], Inta Brikše[1], Jānis Pencis[1] and Laura Uzule[1]

[1] University of Latvia, Faculty of Social Sciences, Riga, Latvia
[2] University of Oxford, Oxford, United Kingdom
{jurgisskilters, uldis.bojars, janis.pencis}@gmail.com, monkru@mac.com, inta.brikse@lu.lv, laurauzule@inbox.lv

Abstract. The aim of the current paper is to formulate a conception of pragmatic patterns characterizing the construction of individual and collective identities in virtual communities (in our case: the Twitter community). We have explored several theoretical approaches and frameworks and relevant empirical data to show that the agents building virtual communities are 'extended selves' grounded in a highly dynamic and compressed, linguistically mediated virtual network structure. Our empirical evidence consists of a study of discourse related to the Latvian parliamentary elections of 2010. We used a Twitter corpus (in Latvian) harvested and statistically evaluated using the Pointwise Mutual Information (PMI) algorithm and complemented with qualitative and quantitative content analysis.

Keywords: Twitter, virtual identity, social science, political messages.

1 Introduction

In this paper, we explore the pragmatics of political messages in Latvian Twitter communication during the 2010 general election.

The results contain a topical analysis of election discussions as well as an analysis of hashtags and retweeted messages. The fast pragmatic dynamics in Titter communication can be observed through hashtags, showing a rapid reaction of Twitter users to the elections, while top retweets support the findings of content analysis with regard to political sentiment. Content analysis reveals the possibility of significant discrepancies in terms of the cognitive and physical distances between a group and its individual members in their identity generation processes. In view of the results, we propose a hypothesis that reveals correlations between a group and its individual members, the richness of topics, channels of communication, frequency of mention, and connotations and effects of messages.

2 Theoretical Background

We assume that the generation of identity takes place through two simultaneous and mutually interdependent social categorization processes – belongingness and differentiation [3,4]. Our study undertakes to examine these two processes in action, constrained by two selection criteria: (1) Twitter messages only, and (2) messages relating directly to national politics. The homogeneity of format and topic draws attention to similarities and differences in content and in discourse strategies.

Twitter is a particularly fruitful resource for this type of analysis because its brevity constraint gives rise to an abundance of shortcut techniques including expressive lexis, the use of abbreviations and hyperlinks for proper names and keywords. Rigid information hierarchies reveal what users presume to be already known and/or shared by their in-group, and are a fertile soil for the investigation of presuppositions, cultural common ground, and cultural discrepancies [6,7,15]. This is especially prominent in Twitter discourse about politics, a topic where speakers generally exhibit willingness to report their opinions despite the fact that their perspectives are often conflicting. Although political opinions are usually articulated explicitly, belongingness to an identity group[1] may be partly implicit [19].

We focus on mechanisms of self-identification, formation and maintenance of in-groups and their differentiation from out-groups. The findings attempt to answer the following questions: 1) How are virtual political identities generated and maintained in a condensed public mode of communication? 2) What are the pragmatic instruments that help to achieve these processes?

Twitter can also help to understand implicit social categorization. Typically, research on social categorization is conducted using questionnaire or focus group methodologies, mainly addressing explicit political categorization. This study has incorporated some implicit factors of analysis, often crucial in political communication. Approaching human-generated digital content as empirical material for categorization analysis is not new (cp. [9]). Analysis of political messages on Twitter, although not directly focused on categorization, is also provided by several studies (cp. [22]). Several recent studies explore possible correlations between election outcomes and the level of Twitter activity of politicians (US Congress: [14], South Korea: [12]). This study, however, also analyzes political messages created by media organizations and other active users.

2.1 Collocations and concordance analysis

Co-occurrence statistics allow to quantitatively project some of a word's semantics grounded in users' categorization performance ([18]). Collocations show the relative most frequent (sometimes stereotypical, implicit) social categories in communication, but the research must be complimented with concordance analysis for semantic complexity. Of course, the output of such a combination of methods concerns the group (and not individual) patterns of social categorization, and pragmatic effects are related to statistical frequency of language used in communities and not to individual patterns of communication[2].

[1] We define identity as a continuous process where the sense of belongingness to a community interacts with the desire to be a unique individual. A community has an internal and an external structure (relationships within the group and relationships with other groups), and community identity can generate polarization effects.

[2] A pragmatic pattern is a typical way of using language in a linguistic community (e.g., in social media).

2.2 Political messages

Studies show that people frequently have difficulty explicitly articulating their ideology [19]. Thus self-report, focus groups, and questionnaires may often prove inadequate for analyzing political categorization. Ideological labels, moreover, may not correspond to subjective conceptions of beliefs, and undecided voters exhibit a much clearer opinion via implicit tasks than via explicit ones [19]. Political categories are distinguished above all by their extreme polarization (cp. [11,17]). On Twitter, initially informative messages are modified to become increasingly polarized [23].

2.3 The Latvian Parliamentary Election 2010

The *Saeima* (the parliament of Latvia) is elected using a proportional multi-partisan representation system for 100 seats. The 2010 election saw 13 competing political parties or their alliances. Candidates from 5 parties were elected: 33 seats for "Unity" (Unity), 29 seats for "Harmony Centre" (HC), 22 seats for the "Union of Greens and Farmers" (UGF), and 8 seats each for the National Association "All For Latvia!"-"TB/LNNK" (NA) and "For a Good Latvia" (FGL). The turnout for the 2010 elections was 63.12% or around 967 000 people.

3 Methodology and Design

The aims of this study are: (1) to build a feasible methodology using content and structural analysis of social media (in particular, Twitter) with respect to political communication; (2) to explore correlations between the election results and the representations of political parties and their candidates in Twitter communication; (3) to explore the identity generation of political actors in pre-election communication on Twitter.

We collected a dataset of tweets covering the election week, performed careful manual extraction work and numerous statistical comparisons. We also created custom tools for analyzing Latvian Twitter content including a concordance tool. We believe that this makes our results, in several respects, even more precise than, e.g., [22] who automatically translated their corpus of empirical data (German tweets) into English and only then processed it with LIWC (Linguistic Inquiry and Word Count).

3.1 Dataset

The dataset consists of one week of Twitter messages (from 28-Sep-2010 to 04-Oct-2010) from a subset of Latvian Twitter users, including 4 days before the election, the day of the general election (October 2) and 2 days following the election. The total size is 50'032 messages, consisting of: 50% regular tweets; 18% retweets; and 32% replies. There are no publicly available official data about the total number of Twitter participants in Latvia. According to local media experts, the estimate is approximately 40'000 users (November 2010).

In order to choose a topically relevant set of Twitter accounts, we started with a manually selected set that included (1) accounts of political parties and their candidates to the Parliament (Saeima); (2) accounts of media organizations, political analysts, and other individuals who write about politics and the election; and (3) accounts of individuals most active in the Latvian Twitter-sphere. This formed an initial set of 179 accounts to follow. We enlarged the set of accounts by (1) retrieving tweets from the current set of accounts; (2) identifying new accounts mentioned in the tweets collected; (3) filtering out accounts not related to Latvia; and (4) repeating this process. The result is a total of 1'377 user accounts to collect tweets from.

We did not choose a random sample to avoid large amounts of redundant data consisting of ordinary discussions unrelated to our research interests - politics, identity generation, and the media. This intentionally selected dataset allows for a more precise analysis of the above research topics.

3.2 Tweet Processing and Analysis

Collected tweets are processed using the NLTK library [1]. The processing of tweets consists of: cleaning the dataset; saving the full tweet data for structure analysis; tokenizing tweets; replacing keywords, where we consolidate the various ways to write the same word or expression and replace it with a single keyword identifier.

Latvian is an inflected language in which the same word may appear in many forms. In the keyword replacement step, we collapse these forms into one keyword. We also replace different ways of writing the same expression (e.g. abbreviations and full names of party names). Since there was no stemming or lemmatizing software for Latvian that we could use, we created our own keyword replacement map for keywords related to elections.

Having processed the tweets, we performed: (1) content analysis in which we examined the text content of Twitter messages; and (2) structure analysis, in which we examine the metadata in tweets and associated with tweets. The main types of text processing performed in the content analysis phase are concordance lookup, word frequency analysis, and collocation (bigram) analysis. For collocation ranking, we used the Pointwise Mutual Information (PMI) metric [16].

4 Content Analysis

4.1. Representations of the candidates on Twitter

We made a list of all 1234 candidates competing for seats in the parliament, exploring their representations in selected tweets during the 4 days leading up to the election. Since only a small part of all candidates were represented in Twitter communication (in our dataset) four days before the election, we wished to compare our findings with publicity coverage of the candidates in other media in Latvia.

We identified 79 family names of the candidates occurring in collocations in the Twitter dataset, and 170 family names of the candidates occurring in the media monitoring dataset. We distinguish four groups of candidates: (1) those represented both in Twitter and print media and news agencies (44 candidates or 3.56% of all the candidates); (2) those who are represented mostly in Twitter (6.40%); (3) those who are represented mostly in print media and news agencies (7.37%); and (4) those who are mostly not represented in the media we studied (82.67% of all the candidates).

Further, we listed how many personal tweets, collocations and publications occur with every of the family names in various time periods (the average number of collocations of every family name of the candidates four days before the election is 4.68; later, we included only those (9) family names that are statistically significant with respect to their number of collocations ($n \geq 4.68$)). Almost all of these candidates (except one) were elected[1]. They also represent 4 out of the 5 parties elected to the parliament. We analyzed the split of the 100 elected candidates between four previously distinguished groups of candidates. Our calculations show that 32% of elected candidates correspond to the first group (represented in Twitter, print media, and news agencies); 5% correspond to the second group (mostly represented in Twitter); 42% correspond to the third group (mostly represented in print media and news agencies); and 16% correspond to the fourth group (mostly not represented in the media we studied). Based on all of the above, we have formulated a working hypothesis: (1) the more thematically varied and (2) the more frequent the communication, and (3) the more communication channels are used to mention a candidate, the higher the probability that he or she will be elected to parliament.

Table 1: Mentions of political parties (collocations on Twitter, publications in print media and news agencies) and the number of seats in the parliament.

Party	Collocations		Publications[2]		Seats
	28.09-01.10	28.09-04.10	27.09-01.10	1 year	
Unity	34	73	276	2799	33 c[3]
UGF	5	20	223	4837	22 c[6]
HC	10	31	232	4674	29
FGL	41	77	230	1647	8
NA	12	32	168	713	8
LP	3	30	0	0	0
FHRUL	3	7	118	1452	0
R	0	6	0	0	0
OPR	2	2	0	0	0
ML	0	1	0	0	0
DL	0	0	0	144	0
LCDU	0	0	0	0	0
PC	0	0	0	0	0

DL = “Daugava for Latvia”; FGL = “For a Good Latvia”; FHRUL = Union "For Human Rights in a United Latvia"; HC = “Harmony Centre”; LCDU = “Latvian Christian Democratic Union”; LP = “The Last Party”; ML = “Made in Latvia”; NA = National Association “All For Latvia!” – “TB/LNNK”; OPR = “For a Presidential Republic”; PC = “People's Control”; R = Social Democratic Alliance “Responsibility”; UGF = “Union of Greens and Farmers”; Unity = Union “Unity”.

[1] *Election of the 10th Parliament of the Republic of Latvia, October 2, 2010: list and statistics of the candidates*. *The website of the Central election committee*. Retrieved January 4, 2011 from http://www.cvk.lv/cgi-bin/wdbcgiw/base/komisijas2010.cvkand10.sak

[2] Publications in print media and news agencies for (1) the election week (27-Sep – 03-Oct); (2) one year (28-Sep-2009 - 03.10.2010). Dates differ from those in tweet collocations due to the source of press data.

[3] c = Formed the ruling coalition.

4.2 Representations of the parties on Twitter, in print media, and by news agencies prior to the election

For names of political parties (Table 1) we listed: (1) how many collocations occur with each name; (2) how many publications from print media and news agencies mention each name; and (3) the results each party has achieved in the election. Every party with an above-average number of collocations in Twitter communication before the election (8.46) is elected to the parliament. An exception is UGF, which was elected despite a below-average number of collocations. We assume that the latter was compensated in the long term by the highest number of publications in print media and news agencies. However, with the high ranking of mention on Twitter before the election (41 collocations), FGL obtained significantly fewer parliament places than "Unity" or other political parties with a lower ranking of mention on Twitter. Initially, it can be assumed that FGL was affected by relatively lower publicity rates in print media and news agencies; but in fact, FGL had conducted a more extensive advertising campaign than any other political party). Further investigation points to an important qualitative factor. A review of collocations of FGL and "Unity" in a detailed concordance analysis leads to the observation that the "Unity" collocations feature more positive connotations than the FGL collocations. This allows us to emphasize and modify our above hypothesis regarding the candidates: (1) the more thematically varied and (2) the more frequent the communication, and (3) the more communication channels a political party is mentioned in *positively*, the higher the probability that it will be elected to the parliament.

4.3 Identity-generation processes for political parties and individuals in Twitter communication

Two political parties – FGL and "Unity" - have significantly higher rankings of mention than other parties. Moreover, their candidates for the post of Prime Minister (Ainārs Šlesers (FGL) and Valdis Dombrovskis (Unity)) have similar rankings of mention. In spite of these similarities, the two have strikingly different election results ("Unity" won the election and got 33 seats in the parliament, with Valdis Dombrovskis approved as the prime minister, while FGL got only 8 seats in the parliament). This led us to investigate more closely the identity generation of these individuals and organizations through political categorization in pre-election tweets. First, we identified 10 collocations of significantly high ratings for the four name keywords. Secondly, we used concordance analysis to examine the semantics in each collocation.

We have listed in Table 2 what percentage of the topics bear positive, neautral or negative connotations and how many topics are covered by each of the keywords. As Table 2 demonstrates, the individual and the organization are categorized similarly in the case of Šlesers and his political party FGL: both are more related to negative topics than positive ones. The case of Valdis Dombrovskis and his political party "Unity" is different: the individual is mostly categorized in positive or neutral topics, while the political party is categorized in negative or neutral ones. This shows that the generation of identity of an organization and that of its individual members may involve significant discrepancies in terms of cognitive versus physical distances[4]. In this case, the cognitive distance between Dombrovskis and "Unity" is bigger than the 'physical' one. This may be in part due to the fact that the "Unity" election campaign focused exclusively on Dombrovskis, promoting him as the principal benefit to the voters. Thus the individual became more cognitively important than the whole (an organization).

This allows us to expand our hypothesis regarding politicians and political parties as follows: (1) the more thematically varied and (2) the more frequent the communication, and (3) the more communication channels are used to mention a member of an organization (in this case, a politician) *positively*[5], the higher the probability that he or she will become cognitively more important than the organization (in this case, the political party) and cause a shift in the perception of the significance of the organization.

Table 2: Connotations of keyword topics.

	Positive	Neutral	Negative	No of topics
Politician (party)				
Dombrovskis (Unity)	27.59%	68.97%	3.45%	29
Šlesers (FGL)	25.00%	41.67%	33.33%	12
Political party				
Unity	0.00%	54.55%	45.45%	11
FGL	21.43%	21.43%	57.14%	14

5 Structural Analysis

In this section, we analyze Twitter messages by examining implicit and explicit metadata and structural information contained in tweets.

5.1 Hashtag Analysis

Hashtags were used in 2'238 tweets (4.47% of all tweets). In total, 750 different hashtags were used 2'668 times. Most hashtags were used just once. 29.06% of hashtags (218) were used more than once and 2.26% (17) were used at least 20 times.

The most popular hashtag was #velesanas ("election"), used in 459 tweets (17.2% of tweets containing hashtags). Other election-related hashtags that were used at least 20 times include #nobalsoju ("i voted"), #politsports (political sport), #pietiek

[4] In the Spreading-Activation Theory, assuming a correlation between the collocational structure of the corpus and the mental models of its users, collocational structure reflects the cognitive distance between conceptual entities such as political parties and individuals. Indirectly connected nodes are more distant than directly connected ones.

[5] Using manual concordance analysis, connotations are determined and generalized according to three categories (positive, neutral, and negative), determined individually for each tweet. Examples include: "Friends, tomorrow I shall vote for Dombrovskis, because I trust his professionalism …" (positive); "Šlesers doubts the objectivity of social media …" (neutral); "Dombrovskis: a protégé of corruption or a racketeer?" (negative).

("enough!"), #vēlēšanas (#velesanas with Latvian diacritics), #cieti ("solid" – a slogan of FGL), #twibbon (twibbons were used to show party support).

For the purposes of this paper, we limited Table 3 to hashtags related to politics. Most of the top 10 hashtags on election day were related to politics (9 out of 10) and appear in the table. Other days had less election-related tags, but also a lower hashtag usage activity in general. The #velesanas ("election") hashtag appeared the day before the election and had a remarkable spike in its usage on election day and the day following it, receding back to background level the day after that.

Table 3: Dynamics of top hashtags related to politics (30-Sep-2011 – 04-Oct-2011).

Hashtag	30-Sep	01-Oct	02-Oct	03-Oct	04-Oct
#ir	27	34	21	32	
#pietiek	7	10	16	10	5
#pll	5				
#politika	5				
#politsports	5				
#velesanas		12	346	94	5
#cieti		8	16		
#fail		9	5		9
#sleptareklama		5			
#nobalsoju			71		
#twibbon			60		
#vēlēšanas			35	11	
#velesanas2010			7		

Hashtags that retained popularity for at least 4 days in this 5 day period were the journalism tags #pietiek and #ir. Both refer to publications seen by top Twitter users as prestigeous and integral organizations for investigative journalism. The hashtag #sleptareklama ("hidden advertising") coincided with the appearance of controversial hockey-related advertisements that were suspected of containing hidden political advertising. A creative usage of a hashtag is its syntactic integration into a sentence: a notable example is using #ir, the magazine whose name literally means "is", as a verb: e.g., "There #is still time to form a new coalition".

Apart from the obvious purpose of attracting attention to major topics, hashtags carry the connotation of familiarity with the object of the tag, be it a topic, an individual, or an organization – at the very least, one must know what is worth tagging. Tags help to define group identity in two recursive ways: by highlighting issues considered important by the group, and by presenting the group as the kind of community where such issues are considered important.

5.2 Analysis of Retweeting

We considered a retweet any Twitter message that contains the string "RT @nickname" (17.68% of the selected dataset). Most retweets start with "RT @nickname", i.e. are marked as such and point to the original message. These results shows more uniformity of retweet formats than reported in [2], possibly a result of more officialized retweet functionality. For further analysis, we used retweets which contained information about the original tweet (i.e. 90.46% of all retweets). An analysis of the top 20 most retweeted posts reveals that 70% of these posts are directly related to the elections; 10% are loosely related; 20% are unrelated.

There were 14 election-related messages among the 20 most retweeted messages. Of these, the majority (8 out of 14) were satirical tweets criticising a political party or a politician. Seven refer to FGL or its prominent members Ainārs Šlesers and Andris Šķēle. Other parties mentioned in these retweets were HC and FHRUL (one tweet each). The two most retweeted messages are related to the election.

5.3 Opinion leaders and in-group demarcation mechanisms

The content of top retweets and hashtags reveals that the opinion leaders in the Latvian Twitter-sphere, the in-group that enjoys the highest popularity and prestige, can be vaguely defined as a group of centrists who see themselves as positioned between two perceived polarities. The cognitive space, as regarded by the in-group, can be characterized thus: to the left are *krievi* ("the Russians"), the parties and their supporters commonly perceived as pro-Muscovite and representing the interests of the Russian-speaking population (HC, FHRUL). To the right are *nēģi* ("the parasites" – an imprecise translation of the word taken from a popular tweet criticising this group), the nationalist alliance (FGL, NA) that the Twitter opinion leaders see as outdated and highly corrupt, exploiting their privilege for personal gain. The in-group supports the political alliance "Unity" and particularly its leader, Valdis Dombrovskis, who was subsequently elected Prime Minister.

The fact that the in-group appears to take a centrist position is significant: their output is less polarizing than could be expected of a highly politicized group. Still, there is a clear demarcation of the in-group from both out-groups described above. This is achieved by the opinion leaders of the in-group through several group-identity-generating mechanisms and strengthened by the heightened emphasis on the social self [4], typical of both online communities and political discourse.

Manipulating cognitive distances is relatively easy in the dematerialized virtual space, which facilitates impressions of togetherness and mutual identification within the in-group, on the one hand, but also the distancing of the in-group from out-groups. Perhaps surprisingly, the brevity constraint of Twitter messaging, rather than complicating political categorization, can facilitate it: the format is well suited to the in-group's simplified tripartite view of the political space. Thus, through repeated tweeting of negative content containing the letters "PLL" or "PCTVL" (acronyms of the names of political parties on the two sides of the perceived spectrum), it is soon enough to write "PLL" or "PCTVL" to evoke a cognitive frame [8] associated with negative content. Clearly, the details of this content will be unique for each user; but as long as there is a basic understanding of a commonality of reference – in this case, of the negativity of the referents – a mention of a party acronym will effectively serve as an invitation to 'fill in the gaps' with each reader's own meaning [13].

Political jokes[6], abundant in top retweets, work in a similar manner. Provided that the humourous effect is usually achieved by inviting the audience to *frame-shift* through an unexpected element [7], political jokes on Twitter are doubly rewarding because they give the audience the feeling of belongingness through having understood the frame shift without surrounding linguistic context and through a very limited number of signs. Similarly to a hashtag, a retweet works recursively by simultaneously flaunting an individual's understanding (and hence his belonging to the in-group) and helping to define his individual identity through the content of what is understood and retweeted.

Our corpus shows that power and control are very much the preoccupation of Twitter users, and the independently formed, 'grass-roots' community of top tweeters quickly forms their own behaviour canons. This is typical of online communities, where a myriad of rules and expectations underlie seemingly free, chaotic communication [10]. A popular political message on Twitter is at once an expression of individual and group identity, an invitation to the in-group members to share the opinion expressed, and a warning about the consequence of deviating from the group's norms. By way of illustration, a message retweeted 15 times reads: "I heard that Šlesers won't vote for PLL *either*, because they're said to be thiefs" (our emphasis). In addition to cleverly poking fun at the politician by suggesting he will not vote for his own party, the message succeeds in conveying that the author will not vote for Šlesers, that he assumes that his in-group members will not do so, and that anyone who does vote for Šlesers will be seen as voting for a thief and undermining his or her in-group membership. In short, Twitter conformity mechanisms are just as compact as the medium itself.

Yet without a conforming audience, such successful guidance toward a rigidified, formal categorisation would not be possible (we may well judge the above message as successful, since it is on the list of top retweets). The tension between individual opinion and in-group identification (the personal vs. the interpersonal/social self) is resolved through a balance of stereotyping processes: just as the political parties and actors are stereotyped to fit into one of the few cognitive categories carved out for the occasion of the election, so the individual members engage in a certain degree of *self-stereotyping* [20]. Members will be more willing to overlook differences of opinion and concentrate on their commonalities (real or imagined) when membership is seen as beneficial, and particularly if the group is seen as working toward a common goal of some sort – in this case, victory in the parliamentary elections [4]. Because intra-group attraction on Twitter in the run-up to parliamentary election is ideational rather than interpersonal, the in-group *achieves* a high degree of political cohesion in part simply through *perceiving itself* as a cohesive unit.

[6] An example that is comparatively demure and reproducible in an academic paper refers to the leader of the party perceived as being in the "parasites" group: "A little boy falls. Šlesers helps him up. 'So, I guess now you will vote for me?' 'I only hurt my foot, not my head!'".

6 Results and Conclusion

We have formulated a correlation according to which three factors contribute to the efficiency of political messages in the electoral discourse – in particular, *for a given collocation bigram*: (a) the variety of thematic contexts of occurrence, (b) the frequency of mention, (c) positive connotations. (While there are other factors determining efficiency, this study has focused on popularity-oriented facets.) We have therefore extended the results stated by [12, 14] regarding the correlation between minority parties, Twitter activity, and election results. The dynamics of Twitter users' interest in the event (the election) can be observed through hashtag usage and the most retweeted messages. Top retweets, in turn, convey user sentiment toward political parties and individuals.

We have noted instances of discrepancy between attitudes toward individual politicians as opposed to attitudes toward political groups, and observed that frequent positive mention of individuals can lead to a heightened cognitive significance of this individual, causing the perception of the significance of the relevant organization to recede into the background.

We envision possible applications of this work in analysis tools correlating Twitter dynamics with the structure generated from the parameters: (a) the variety of occurrence contexts, (b) the frequency of mention, (c) positive connotations (generated semi-automatically). The items which fit into the highest ranking of such analysis results can be further analyzed manually and a variety of pragmatic effects (stereotyping, presupposition generation a.o.) might be observed.

Finally, we can hypothesize that the user of a microblogging resource such as Twitter extends the sphere of his or her cognitive processing by involving additional interactive structures of communication. Thus, if we assume that the social categorization in a community consists of (a) self-categorization as the most crucial and basic level of identity building, (b) interpersonal communities of individuals, and (c) large-scale social communities (e.g., national identity communities) including sub-communities [4], we could argue that self-categorization involves a substantial amount of extended cognitive processing offloaded onto the digital environment (in our case, Twitter). In this sense, the results provided by our study can complement research on the extented mind [5, 21]. A more detailed analysis of the extended self and offloading effects in cognitive processing is a topic for another study.

Acknowledgments

This work has been supported by the European Social Fund project «Support for Doctoral Studies at the University of Latvia».

References

1. Bird, S., Klein E., & Loper , E. (2009), Natural Language Processing with Python. O'Reilly.
2. Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. HICSS-43. IEEE: Kauai, HI, January 6.
3. Brewer, M. B., (1991). The social self: On being the same and different at the same time. Personality and Social Psychology Bulletin, 17, 475-482.
4. Brewer, M. B., & Gardner, W. (1996). Who is this "we"? Levels of collective identity and self representations. Journal of Personality and Social Psychology, 71, 83-93.
5. Clark, A. (2006). Pressing the flesh: Exploring a tension in the study of the embodied, embedded mind. Philosophy and Phenomenological Research.
6. Clark, H.H. et al. (1983). Common ground and the understanding of demonstrative reference. Journal of Verbal Learning and Verbal Behaviour, 22, 245-258.
7. Coulson, S. et al. (2006). Looking back: Joke comprehension and the space structuring model. Humor, 19, 229-250.
8. Fillmore, C. J. 1976. Frame semantics and the nature of language. Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech, Vol. 280, pp. 20-32
9. Glushko, R.J., Maglio, P.P., Matlock, T., & Barsalou, L.W. (2008). Categorization in the wild. Trends in Cognitive Sciences, 12, 129-135.
10. Haythornthwaite, C. (2007). Social networks and online community. In A. Joinson, K. McKenna, T. Postmes, & U.-D. Reips (Eds.). The Oxford Handbook of Internet Psychology (pp. 121-137). Oxford: Oxford University Press.
11. Heit, E., & Nicholson, S.P. (2010). The opposite of republican: polarization and political categorization. Cognitive Science, 34, 1503-1516.
12. Hsu, C.-L., & Park, H.W. (2010). Sociology of hyperlink networks of Web 1.0, Web 2.0, and Twitter: a case study of South Korea. Social Science Computer Review, 0894439310382517, first published on September 21, 2010.
13. Langacker, R.W. (2000). Grammar and Conceptualization. Berlin/New York: Mouton de Gruyter.
14. Lassen, D.S., & Brown, A.S. (2010). Twitter: the electoral connection? Social Science Computer Review, 0894439310382749, first published on September 23, 2010.
15. Laurent, J.P. et al. (2006). On understanding idiomatic language: the salience hypothesis assessed by ERP's. Brain Research, 1068, 151-160
16. Manning, C.D., & Schutze, H. (2003). Foundations of Statistical Natural Language Processing. The MIT Press, Cambridge, Massachusetts, 2003.
17. Medin, D. L., Lynch, E. B., & Solomon, K. E. (2000). Are there kinds of concepts? Annual Review of Psychology, 51, 121-147.
18. Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. Cognitive Science, 34, 1388-1429.
19. Nosek, B.A., Graham, J., & Hawkins, C.B. (2010). Implicit political cognition. In B. Gawronski, & B.K. Payne (Eds.). Handbook of Implicit Social Cognition: Measurement, Theory, and Applications (pp. 548-564). New York: The Guilford Press.
20. Smith, E. R., & Henry, S. (1996). An in-group becomes part of the self:Response time evidence. Personality and Social Psychology Bulletin 22, 635-642.
21. Spivey, M., Richardson, D. & Fitneva, S. (2004). Thinking outside the brain: Spatial indices to linguistic and visual information. In J. Henderson & F. Ferreira (Eds.), The Interface of Vision, Language, and Action (161-189). New York: Psychology Press.
22. Tumasjan, A., Sprenger, T.O., Sandner, P.G., & Welpe, I.M. (2010). Predicting elections with twitter: what 140 charcters reveal about political sentiment. Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. 178-185.
23. Yardi, S., & Boyd, D. (2010). Dynamic debates: an analysis of group polarization over time on Twitter. Bulletin of Science, Technology & Society, 30(5), 316-327.