

A study of texts of an extremist forum “Kavkazchat” using Linguistic Inquiry and Word Count (LIWC)



Tatiana Litvinova, PhD, head of RusProfiling Lab, Voronezh State Pedagogical University, Lenina 86, 394043 Voronezh, Russian Federation, centr_rus_yaz@mail.ru
Olga Litvinova, PhD student, RusProfiling Lab, Voronezh State Pedagogical University, Lenina 86, 394043 Voronezh, Russian Federation, olga_litvinova_teacher@mail.ru

Abstract

Internet has turned into a powerful tool of manipulating the minds and behavior of young people. This gives rise to autonomy of countless youth extremist movements. Internet resources are also massive tools of promotion and propaganda for terrorist activities. It is thus of growing importance to analyze extremist and terrorist online language using modern linguistic tools to understand psychology of threats. The material for the current study are texts of a Russian Islamic extremist forum ‘Kavkazchat’ accessed via The Dark Web Portal (Qin et al., 2005). The texts were analyzed by means of the *Linguistic Inquiry and Word Count (LIWC)* software (Pennebaker, Francis, Booth, 2001) with the help of the included Russian dictionary (Kailer, Chung, 2011) as well as those designed by the authors. This software allows large bodies of texts to be analyzed along a whole range of linguistic and psychological categories. For the first time Russian texts of an extremist forum have been analyzed using LIWC along a wide range of linguistic and psychological categories. The obtained results were compared with the data for extremist texts in English.

Background

LIWC is commonly used to analyze extremist and terrorist texts, mostly in English, for example language of the ISIS, Al-Qaeda, etc. (Chung and Pennebaker, 2011; Pennebaker and Chung, 2005; Pennebaker and Chung, 2008). The software is easy to use to analyze such texts as hidden motives and interests of terrorist groups could be examined in their dynamics as well. Unlike traditional content-analysis software, it is also a unique tool of identification of latent psychological characteristics of the authors. It should be noted that most of the above study was performed using publicly available sources produced by individual authors (books, extremist media, etc.), however it is also important to analyze forums where people immediately involved in extremist ideology interact with one another.

- Examples from Kavkazchat:
 - *I really hope these were our people. Looking at those faces of kafirs, that should be true. Allah, let these special operations be of any help! And hopefully they (special operations) continue for the profit of Islam and Muslims! Allahu Akbar! I am happy that the female shahids attacked the Moscow subway. THIS IS FOR VOSHA AND DADA!!!! For all the Muslims who lost their lives! (translations ours). Inshallah must enter their homes and bring fear and terror so that they can understand what it is like.*
 - *There are no innocents. The country's population along with the government should be held accountable for their policies. If what the government does daily is to kill people in the Caucasus, they should expect unrests in Moscow.*
 - *Inshallah what happened in Moscow is just a beginning. You will be revenged for these atrocities in the Caucasus before you know it!*
 - *You will be blown up, killed and made use of as you don't deserve any better until you start respecting yourselves and put a stop to the continuing massacre in the Caucasus.*

Objectives

- This study is devoted to the analysis of posts from Russian-language forum of North Caucasian extremists KavazChat on one particular topic, namely The 2010 Moscow Metro bombings (29/03/2010). KavazChat is included on the RF Federal list of extremist materials and access to the site is blocked. We accessed “KavazChat” due to Dark Web Project (<http://www.azsecure-data.org/dark-web-forums.html>)
- Objectives:
 1. Analysis of posts from Russian Islamic extremist forum ‘Kavkazchat’ devoted to discussion of 2010 Moscow Metro bombings using Linguistic Inquiry and Word Count (LIWC).
 2. Analysis of comments of Internet users on the news about bombings using LIWC.
 3. Comparison of obtained results to identify the ideology expressed in Islamic extremist texts.

Methods

- **Text corpora**
- Text corpora comparable in terms of the number of texts they contained and dealt with the same topic (discussions of 2010 Moscow Metro bombings produced immediately following the event) were compiled to make it convenient to employ them in comparative analysis using LIWC: **extremist corpus** and **reference corpus**. (readers’ comments on news about bombings from 2 popular Russian internet media)

Corpus	N of authors	N of comments	Dates of posting	Corpus size
Extremist corpus (EC)	67	466	29/03/2010–17/04/2010	274209 characters (without spaces), or 50447 words
Reference corpus (RC)	143	1313	30/03/2010–06/04/2010	261874 characters without spaces, or 42878 words

- **Russian LIWC**
- **User dictionaries:**
 - ✓ a dictionary of discursive markers - *DiscourseMark* (14 features),
 - ✓ a dictionary of intensifiers and downtoners - *Intens* (2 features),
 - ✓ a dictionary of perception vocabulary - *PerceptLex* (1 feature),
 - ✓ a dictionary of pronouns and adverbs describing the speaker - *Ego* (I, my, in my opinion; 1 feature),
 - ✓ a dictionary of emotional words - *Emo* (negative and positive; 2 features)
- **Semantic word-class scoring (Mihalcea and Pulman, 2009).**

Results

- **Standard LIWC features:**
- **Function words**
- Extremists use more pronouns overall and personal pronouns in particular. Among personal pronouns the largest difference there are in subgroups of “we” “you”, “they”, with the highest scoring for ты “you singular” (2,76)
- Extremists use more negations
- **Emotions:**
- Extremists use more words describing positive emotions and less words describing negative emotions
- Internet users use more words describing anxiety and sadness, extremists use more words describing anger
- **Social processes words** are dominant in extremist corpus
- **Personal concerns:**
- Internet users talk more about work, leisure, home, extremists talk more about religion

Results

- **User dictionaries:**
- Extremist texts:
 - *Discursive markers*, denoting differences (2.07), consequences (1,8), aim (1,3), conditions (1,4)
 - *Intens* : downtoners (1,2)
 - *Emo*, positive words (1,2)
- The revealed pattern of extremist ideology related to polarized thinking and opposition *we – you (they)*, reference to religion is similar to that revealed by Prentice et al. It also should be noted that an elevated number of words describing *them* could be a sign of fixation warning behaviour which is defined as any behaviour indicating an increasingly pathological preoccupation with a person or a cause. The fixated person expresses a preoccupation with the group or person considered responsible for the subject’s grievance by allocating large amounts of time to discussing, theorising about, or studying the perceived enemy.
- Extremist express positive attitude to the bombings and make justifications of terrorist actions using different types of discursive markers. Hart and Lind (2011) have also found out that language of Islamic terrorists is more optimistic than that of a comparison non-violent groups.

Conclusions

- We presented a general overview of the Islamic extremist language that does not want to offer a deep understanding of the meaning that is hidden behind each of the LIWC dictionaries that we used. Further research should investigate more deeply the meaning of LIWC categories, using a comparative perspective to compare the language used by different terrorist groups.

Acknowledgment

Funding of the project N° MK-5718.2018.6 “Speech portrait of the extremist: corpus-statistical research (on the material of the extremist forum “KavkazChat”)” from RF President’s grant for young scientists is gratefully acknowledged.

References

- Chung, C. K., Pennebaker, J. W. (2011). *Using computerized text analysis to assess threatening communications and actual behavior*. In C. Chauvin (ed.), *Threatening communications and behavior: Perspectives in the pursuit of public figures* (pp. 3-32). Washington, D. C.: The National Academies Press.
- Mihalcea R., Pulman S. 2009. Linguistic ethnography: Identifying dominant word classes in text. In *Computational Linguistics and Intelligent Text Processing*, Springer, pages 594–602.
- Pennebaker, J. W. & Chung, C.K. (2005). *Tracking the social dynamics of responses to terrorism: Language, behavior, and the Internet*. In S. Wessely, V.N. Krasnov (eds.). *Psychological responses to the new terrorism: A NATO-Russia dialogue* (pp. 159-170). Amsterdam: ISO Press.
- Pennebaker, J. W., & Chung, C.K. (2008). *Computerized text analysis of Al-Qaeda transcripts*. In K. Krippendorff, M. Bock (eds.). *A content analysis reader* (pp. 453-465). Thousand Oaks, CA: Sage.
- Pennebaker, J. W., Francis, M. E., Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah: Lawrence Erlbaum Associates.
- Qin, J., Zhou, Y., Lai, G., Reid, E., Sageman, M., Chen, H. (2005). The Dark Web Portal Project: Collecting and Analyzing the Presence of Terrorist Groups on the Web. In P. Kantor, G. Muresan, F. Roberts, D. D. Zeng, F.-Y. Wang, H. Chen, & R. C. Merkle (eds.). *Lecture Notes in Computer Science*, vol. 3495, pp. 623-624.
- Prentice, Sh., Rayson, P., Taylor, P. J.: The language of Islamic extremism: Towards an automated identification of beliefs, motivations and justifications. *International Journal of Corpus Linguistics* 17(2), 259–286 (2012),
- Roderick P. Hart & Colene J. Lind (2011) The rhetoric of Islamic activism: A *DICTION* study, *Dynamics of Asymmetric Conflict*, 4:2, 113-125