Conceptualization of time in internet-texts of different emotional tonality

Anastasia Kolmogorova[1*], and *Alexander* Kalinin[1]

[1]Siberian Federal University, 79 Svobodny pr., 660041 Krasnoyarsk, Russian Federation

Abstract. The research is made in the frame of methodology of emotional text analysis – the computer technology allowing to classify texts according to the criterion of the emotion expressed in them. The paper focuses on specificity of verbal means used in texts of eight emotional classes to deal with the category of time. The data we used consist in 3900 Internet-texts from social network VKontakte assessed by 2000 informants on a crowdsource platform. We processed the raw data using the elements of mathematical modelling, and, a set of tools offered by Sketch Engine corpus manager platform. The hypothesis is that while experiencing an emotion, an individual feels the time differently and, consequently, speaks about it not in the same way. We paid a particular attention to the weight of the lexemes those semantics is connected with the idea of time in different emotional text classes.

1 Introduction

The research we conduct is focused on the problem of automatic text classification according the emotion expressed in it. Using the approach of supervised machine learning we need a training set. To collect it we retrieved 15000 posts from Russian social network VKontakte using the thematic hashtags containing key words of different emotions, later 3900 randomly taken texts from the collection were assessed by 2000 Russian informant on the one of crowd sourcing platform. In this way, we obtained a rather representative corpus of textual data available to the linguistic analysis. Using it, we can perform a number of research task which aim is to reveal some tendencies in using words and constructions while speaking about emotions or speaking while feeling an emotion.

In this perspective, the paper aims to summarize some results of our preliminary work motivated by the hypothesis that even if the concept of time is one of the most basic for human being, when we feel joy or distress our perception of time isn't the same. Thus, we speak differently about time when we feel different emotions.

In the next section of the paper (§ 2), we examine the main assumptions of such approach as emotional text analysis, the concept of time as it is perceived in modern studies on semantics and cognitive linguistics and we describe the background of the project in general. Then, in what follows (§ 3), we describe the results of conducted corpus analysis

---

* Corresponding author: nastiakol @mail.ru

of words and constructions with the semantics of time in 9 classes of texts while emphasizing the gaps between classes. After that (in (§ 4), we use the material to spell out the paper's theoretical implications. Finally, (§5) we summarize suggest how the obtained results can be further broadened.

2 Theoretical grounding

The approach, known since the pioneer work of B. Pang and L. Lee [1]   as sentiment-analysis, is specially focused on text classification taking into consideration the attitude of text author (positive / negative / neutral) towards the object of his discourse or the type of modality used to relate the facts of reality (subjective / objective). Now, a new approach – «emotional analysis of texts» – is actively developing. Its purpose is rather ambitious and consists in designing machine algorithms helping automatically recognize different emotions in texts. There are a number of research teams working on the problem within various paradigms [2; 3]. The most of models are developed on the data of European languages, but our group focuses mostly on creating a classifier for the Internet-texts in Russian.

2.1 Emotional analysis of texts: the problem of emotion classification

In the frame of the emotional analysis of texts the researchers face the problem of how many and what kind of emotional classes are sufficient for describing the main emotional states we are able to feel and to speak about.

The embarrassment felt by the scientists while answering this question, is due to the multiplicity of classifications of emotions in psychological literature. For instance, in [4] the classification is built up on 8 emotions thought as basic by C. Izard [5] (anger, disgust, fear, guilt, interest, joy, sadness, shame, and surprise)), in [6] they prefer five emotions from Ekman's classification. In our project, we assign the emotional labels to texts according to the model of emotions called as "Lövheim Cube" [7]. It includes 8 emotions: Shame / Humiliation, Fear / Terror, Enjoyment / Joy, Contempt / Disgust (each emotion is doubly named because its former nomination indicates the weakest degree of its manifestation, while the latter – the strongest), Distress / Anguish, Anger / Rage, Interest / Excitement and Surprise (or Startle). We also added to the list the ninth class – the neutral texts. The main advantage of Lövheim's view on the emotions consists in proposing a physiological grounding for explaining the nature of emotions and their number: the emotion we feel, as the scientist clarifies, depends on a specific combination of three monoamines in human blood (dopamine, serotonin and noradrenaline).

2.2 Concept of time in linguistic studies

As in this paper, we converge two domains – emotions and the concept of time – we need a brief review of approaches and findings concerning also the latter topic.

The way of thinking the time in context of language has drastically changed during the last two decades. In structural linguistics, the idea of time was mostly associated with a grammar tense [8]. However, due to the developing of anthropocentric paradigm by the end of XX century, linguists became interested in lexical representation of time in discourse. While accepting that human worldview emerging through lifelong cognitive experience is biased, the researchers focus more on how *time* is perceived, survived, felt in discourse. The concept of *lived time* arises: "if the language of tense description is abstract, the

concepts which we use to describe the lived time refer to the human being and represent a kind of naïve philosophy of time" [9:85]. In their well-known work, G. Lakoff and M. Jonhson give a perfect illustration of such naïve time conceptualization by using the colloquialism "Time flows" [10]. They suggest that there is a more general conceptual model TIME PASSING IS MOTION that underlies many utterances of this type. To look "inside" the processes of time conceptualization linguists need to observe how lexemes containing the sema of time function in different contexts and situations. In the next sections, by using the tools and the methodology of corpus linguistics, we will analyze the differences in perceiving lived time in Internet-texts of 9 emotional classes.

2.3 Data and research methodology

As a source of data for our training data set we used 15000 posts from three publics in Russian social network VKontakte: Overheard, Caramel and the Room №6. Usually, the assessments are made by two or three persons whose evaluations are harmonized [11], but our concept was different: we elaborated a special interface allowing a non-discrete assessment procedure [12] and invited 2000 respondents to use it for assigning to each of them not a concrete emotion, but a position on a slider on 4 scales with two opposite emotions on the ends. Using mathematical modelling we succeed to identify the emotional load of each text. In this way we obtained a corpus of labeled text fragments for each emotion and for neutral texts. As the latter is rather little (only 14000 tokens) its use for research purposes has some limitations, however, we took it too into consideration as a kind of so called tertium comparationis.

   To conduct corpus analysis of data we use the platform of corpus manager Sketch Engine offer a large choice of tools for text statistical analysis [13].

3 Corpus analysis and its results

We began our analysis by looking at the functioning of the lexeme *время* (time) in our 9 subcorpora.

   Even if the absolute frequency of the lexeme varies considerably from one emotional class to another (e.g., from 63 items in Distress to 390 in Fear subcorpora), the values of normalized frequency diverge less significaly (the maximum of 1 605.32 items per million in Interest and the minimum of 1 183.67 in Distress subcorpora). Therefore, it is interesting to note that three emotional classes of texts with the most considerable values of normalized frequency of lexeme *время* are Interest, Shame and Startle and three of the less important frequency values – Distress, Anger and Neutral.

   However, more interesting differences in *время* lexical profile were revealed when observing its syntactical combinatorics in 9 subcorpora. We choose 4 type of collocations: 1) adjectives which modify the noun *время* (Adj. on the Figure 1); 2) verbs having *время* in syntactical position of object (e.g. *тянуть время*; V – on the Figure 1); 3) nouns which are used as homogenous members coordinated with *время* (*время и деньги*; Equ (Equivalents) – on the Figure 1.); 4) nouns in Accusative fulfilling the position of object of *время* and introduced by the preposition *на* (*время на стирку*).

   As quantitative analysis shows, the emotional classes of texts with the maximal number of collocates-adjectives and collocates-verbs are Anger and Startle; with the minimal number of collocates-adjectives – Distress, Shame and Neutral and for collocates-verbs – Interest, Disgust and Neutral. It means that in "surprised" and "angry" texts *time* is considered as an easily usable object with many facets. If we compare the "usability" of time in mentioned two classes of texts, we could see some gaps: in angry texts time is

mostly treated as resource (*тратить время*/ spend time), as an object made of flexible material (*тянуть время*/ literally "pull the time") or a point on time scale (*указывать время* /to indicate time). The collocate *тратить* is identified by LogDice statistical measure as a third among the most relevant collocates for lexeme *время*. In Startle subcorpus, time is taken mostly as something being in motion (*проводить время* /to pass the time) and being measurable (*считать время*/ to count the time). The former has eighth position in collocates' rating according to LogDice measure.

It is worth noticing, that in Anger and Shame suncorpora we see the most of collocates-homogenous terms, and on the contrary – in Fear and Neutral there are no collocates of this type. Due to this type of collocates we can see a more general category within which time is conceptualized in an emotional class of texts. For instance, in angry texts time categorical neighbors are *бессмысленность* (lack of sense), *нервы* (nerves), *безденежье* (lack of money), *силы* (physical forces), *деньги* (money); in ashamed texts – *анализы* (analysis), *мест*о (place), *деньги или деньга* (money), *нервы* (nerves). Thus, time in Anger subcorpus is likely included in a rather abstract and subjective category of negatively perceived lack of resources. On the contrary, in Shame subcorpus time is incorporated in more practically oriented model of situation of getting medical care.

As for the "purposeful use of time" marked by noun-object in Accusative, four classes of texts don't contain any items of this kind: Enjoyment, Fear, Disgust and Neutral. However, in interested, angry, surprised, sad and ashamed texts (in order to decrease) there is a number of such occurrences.
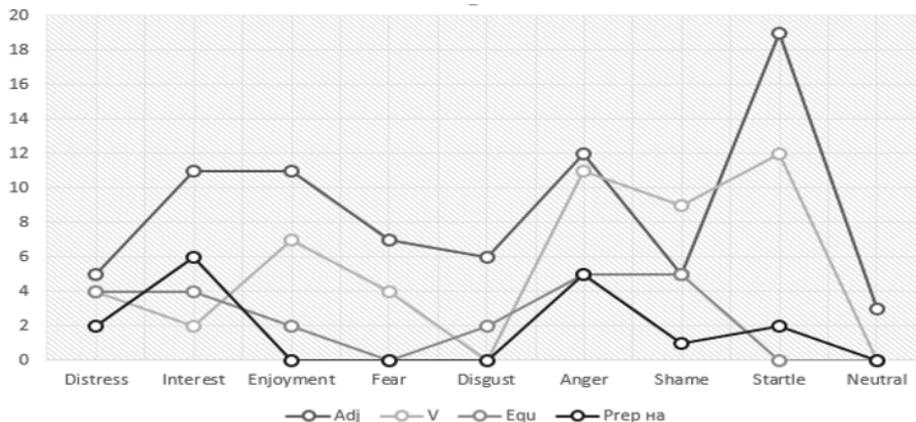

Fig. 1. The number of collocates of the lexeme *время* in four syntactical positions in 9 subcorpora of emotional texts.

The comparison of target-domains for spending time (*время на что-то*) in different subcorpora shows (Table 1), that if in angry and surprised texts people spend time to follow every day routines, interested texts use time idea to focus on social activities, when distressed individual prefer speak about time in human perspective, and when feeling shame, we link the idea of time with our visual perception.

Table 1. Target-domains for spending time in five subcorpora.

Subcorpus	Target-domain and examples
Interest	Social activities: *общение, подготовка, занятие, пара, работа, прием*
Anger	Every day routines: *покупка, нытье, стояние, готовка, свадьба*
Startle	Every day routines: *уборка, плановый осмотр*
Distress	Human being: *одиночество, человек*
Shame	Visual perception: *просмотр*

The next aspect which interested us was the frequency of terms denoting short time period (*минута*, *миг*, *мгновение*, *секунда*) and long time period (*год, десятилетие, век, вечность*).

Table 2. Frequency of terms denoting short and long time period.

Subcorpus	Normalized frequency of short time period terms (ipm)	Normalized frequency of long time period terms (ipm)
Interest	580.97	5295.22
Anger	516.66	**2331.12**
Startle	628.96	4196.11
Distress	**358.9**	5380.47
Shame	**682.06**	5175.45
Disgust	**364.35**	**2417.94**
Enjoyment	580.97	**7119.13**
Neutral	216.73	260.08
Fear	**1016.43**	**6135.11**

The frequencies displayed in the Table 2 demonstrate that the subcorpora the most sensible to the short periods are those of Shame and Fear. The latter seems to be very penetrated by the idea of time, because the summarized frequency of lexemes denoting long periods is one of the highest too. The less temporally marked subcorpus, apart of the neutral one, is Disgust – its frequency values for both lexical series are low. As for the lexemes denoting long periods of time, the highest score is shown by Enjoyment subcorpus and the lowest – by Anger.

The specificity of the most frequent bigrams containing word *время* (time) is also worth to be mentioned. In Fear texts the most specific and relevant n-grams are linked with the observation of she/he/we agents acting in a situation: *время он*, *время она*, *время мы*. In Anger texts the idea of time possession predominates: *свое время*; Disgust and Startle texts focus on concrete life situations concerning human body and having some duration: *время месячных* and *время секса*, respectively; feeling distressed one prefers speak in the perspective of his own personal time − *время я*; being enjoyed or interested people see time as discrete entity: *некоторое время*, *долгое время*, *какое-то время*; for ashamed texts the idea of time is deduced to the concrete situation when something happens − *это время*.

4 Discussion

To open a discussion, we need to return to our base research question: do we speak differently about time when we feel different emotions?

The results of conducted corpus analysis allow us to give a weak positive answer: we can see some biases in lexical profiles of time vocabulary in different emotional classes of texts.

The mentioned above peculiarities give us some prompts about how time is perceived by people feeling different emotions.

Disgust texts: low sensitiveness towards emotions on lexical level.

Interest texts: the idea of time is modeled through the social activities representations.

Shame texts: focus on the short and discrete period of time when something happens.

Startle texts: the concept of time is perceived as a measurable value linked with every day life routines.

Distress texts: time is conceptualized diffusely and in first person perspective.

Anger texts: time is represented as a usable resource, which is necessary for actions.

Fear texts: the distinction between short and long periods is mostly important; the concept is modeled through the "not myself perspective".

Enjoyment texts: time is view in long-term perspective as something having a long duration.

5 Conclusion

In our preliminary research made in the frame of more general project, we formulated a range of suggestions about the dependence of verbally manifested time conceptualization in Internet-texts in Russian. Our weak hypothesis is confirmed: there are some observable and detectible gaps in thinking and speaking about time in different emotional classes of texts. The features proper to each subcorpus could be predetermined by many factors indirectly related with the emotion felt by a text author: traditional patterns used in culture to narrate about emotion; type of denotative situation linked with feeling an emotion state etc.

However, the problem statement is also important, because it has two paths for further development: 1) the research path, which could be able to conduct us to new and explanatory descriptions in semantics and construction grammar; 2) the technological path, which will assure the implementation of revealed features in the practice of machine learning.

Acknowledgments

References

1. B. Pang, & L. Lee, Opinion Mining and Sentiment Analysis, *Foundations and Trends in Information Retrieval*. *2*. 1–2. 1–135 (2008)

2. Zh.Peng, Q. Hu, & J. Dang Multi-kernel SVM Based Depression Recognition Using Social Media Data, *International Journal of Machine Learning and Cybernetics*, US.10 (1). 43−57 (2019)

3. G. Yang, H. He, & Q. Chen, Emotion-Semantic-Enhanced Neural Network. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 27 (1). 531–543. (2019). https://doi.org/10.1109/TASLP.2018.2885775

4. A. Neviarouskaya, H. Prendinger, & M. Ishizuka @AM: Textual Attitude Analysis Model. *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*.80−88 (2010)

5. C.E. Izard, Basic emotions, relations among the emotions and emotion-cognition relation. *Psychological Review*. *99* (3). 561–565. doi: 10.1037/0033-295x.99.3.561 (1992)

6. S. Aman & S. Szpakowicz, Identifying Expressions of Emotion in Text. *Lecture Notes in Computer Science*. 4629. 196−205 (2007)

7. H. Lövheim, A New Three-Dimensional Model for Emotions and Monoamine Neurotransmitters. *Medical Hypotheses. 78*. 341–348 (2012)

8. E.V. Paducheva *Semanticheskie issledovaniya: Semantika vremeni i vida v russkom yazyke; Semantika narrative* [Semantic studies: Tense and aspect semantics in Russian; Semantics of the narrative]. Moscow: Yazyki slavyanskoi kul'tury (2010) (In Russ.)

9. E.S. Yakovleva, *Fragmenty russkoi yazykovoi kartiny mira (modeli prostranstva, vremeni i vospriyatiya)* [Fragments of Russian linguistic world view (model of space, time and perception)]. Moscow: Izdatel'stvo «Gnozis» (1994) (In Russ)

10. G. Lakoff, The contemporary theory of metaphor A. Ortony (Ed.), *Metaphor and thought* (2nd ed.) (p. 202−251). Cambridge: Cambridge University Press. (1993)

11. S. Chaffar, D. Inkpen, Using a Heterogeneous Dataset for Emotion Analysis in Text. In Butz, C., Lingras, P. (eds.) *Advances in Artificial Intelligence*. *Canadian AI*. *Lecture Notes in Computer Science*. 6657. 62–67. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21043-3_8 (2011)

12. A. Kalinin, A. Kolmogorova, A. Malikova, Tool for Non-Discrete Emotional Annotation of Text. *Web application* (2020) http://levheim-survey.verbalab.ru/text.

13. A. Kilgarriff, V. Baisa, J.Bušta, M.Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, V. Suchomel, *Lexicography: Journal of ASIALEX Volume* 1. 7—36. (2015)