COGNITIVE CRITERIA FOR TEXT ENTRY ARRANGEMENT IN RUSSIAN-LANGUAGE THREAT-CORPUS

Prof. Dr. Elena Pozdnyakova¹.²
Prof. Dr. Vera Zabotkina ²

¹ Moscow State Institute of International Relations (University), Russian Federation
² Russian State University for the Humanities, Russian Federation

ABSTRACT
The study makes a case for purposely elaborated text entry arrangement criteria in THREAT-corpus. THREAT-corpus (Russian language, 10.4 m words) was built in 2018 to investigate “sociocultural challenges/threats” in fiction, non-fiction and internet news texts. Various types of sociocultural threats and challenges, spread around the Internet and literature, strongly influence modern culture and ideology. The aim of the study is to substantiate the approach to text entry arrangement in the corpus. For the purpose linguists analyzed the conceptual domain “sociocultural challenges/threats” and built a frame of the domain. So, frame “sociocultural threats” was by design implemented as a cognitive foundation for the corpus.

The set of methods used included conceptual analysis, framing, representative sampling, cognitive modeling. The methods applied made it possible to collect texts, detect and describe sociocultural threats in diverse types of discourse and, consequently, to analyze cognitive aspects of frame-based corpus elaboration.

With the objective to optimise analytical processing, all text entries were grouped by the following criteria: 1) discursive characteristics and strategies; 2) types of threats represented; 3) types of cognitive mechanisms for text arrangement.

The first criterion groups texts as representative samples of different types of discourse. The second criterion allows for categorizing texts by content. The third criterion creates a cognitive foundation for text entry arrangements.

Combining corpus elaboration with linguistic frame analysis resulted in more precise configurations of cognitive models in the conceptual domain “sociocultural threats/challenges”. Word frequency for lexemes, manifesting sociocultural challenges appeared to be an important factor of conceptual structures representation.

Keywords: corpus, sociocultural threats, frame, representative sampling, cognitive criteria

INTRODUCTION
Interdisciplinarity has become the motto for cognitive scientists. Vertically, the following levels of research integration can be distinguished: local integration in one knowledge area (e.g. linguistics); cluster integration (e.g. the humanities); integration of clusters (i.e. synthesis of the humanities and sciences) [1]

Interdisciplinary approach within linguistics has resulted in fruitful collaborations, Cognitive Corpus Linguistics being one of them.
“Within cognitive linguistics, there is an increasing awareness that the study of linguistic phenomena needs to be grounded in usage. Ideally, research in cognitive linguistics should be based on authentic language use, its results should be replicable, and its claims falsifiable. Consequently, more and more studies now turn to corpora as a source of data in what one might call Cognitive Corpus Linguistics. We use this term to refer to research that formulates questions about human cognition in such a way that statistical analysis based on corpus data can yield answers to these questions. As we see it, Cognitive Corpus Linguistics is aligned theoretically with other cognitive approaches to language. Shared issues include the commitment to engage in interdisciplinary dialogue with the other cognitive sciences, the assumption that language structure emerges from language use, and the hypothesis that grammatical knowledge is non-modular”[2].

The research, presented in the article, is in full agreement with the program of cognitive linguists and corpus linguists collaboration. However, following principles of interdisciplinary approach on cluster integration level, the group of researchers have gone further and worked out a program for a cluster of disciplines: cognitive linguistics, literary studies, history, narratology, discourse analysis, corpus linguistics. The discipline cluster has been set with a goal of the in-depth study of sociocultural challenges/threats in modern Russian society. Given that various types of sociocultural challenges are spread around the Internet and literature, they strongly influence modern culture, ideology, and, to a certain extent, affect national worldview. To make the representative sampling comprehensive fiction, opinion-based journalism, historic and historiographic texts, internet news compiled three main groups of texts under study.

**MATERIALS AND METHODS**

Cluster integration required a tool that could consolidate text entries from three groups. THREAT-corpus (Russian language, 10.4 m words) was purposely built in 2018 to investigate sociocultural challenges/threats in all types of texts [3].

At the moment the structure of the corpus looks like this:

**Table 1. Quantitative characteristics of corpus text entries**

<table>
<thead>
<tr>
<th>Number of fiction entries</th>
<th>Number of opinion-based journalism, historical and historiographic texts</th>
<th>Number of internet news texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>68</td>
<td>138</td>
<td>144</td>
</tr>
</tbody>
</table>

The material in the corpus was organized in three main entries in accordance with entry text discourse characteristics. The corpus was based on SketchEngine platform for text entry operating.

The set of methods used included conceptual analysis, framing, representative sampling, cognitive modeling. The methods applied made it possible to collect texts, detect and describe sociocultural threats in diverse types of discourse and, consequently, to analyze cognitive aspects of frame-based text entry arrangement.
RESULTS AND DISCUSSION

The aim of the study is to substantiate the approach to text entry arrangement in the corpus. With the objective to optimize analytical processing, all text entries were grouped by the following criteria: 1) discursive characteristics and strategies; 2) types of threats represented; 3) cognitive mechanisms for text arrangement.

The first criterion groups texts as representative samples of different types of discourse. The second criterion allows for categorizing texts by content. The third criterion creates a cognitive foundation for text entry arrangements.

Teubert wrote, “It is the discourse itself, and not a language-external taxonomy of linguistic entities, which will have to provide the categories and classifications that are needed to answer a given research question” [4]. To follow the quote, we focused on discourse characteristics that provided us with cognitive criteria for text entry arrangement.

Assuming that three groups of texts will demonstrate diverging discursive characteristics and strategies related to “sociocultural challenges/threats” representation, we stated the first cognitive criterion as follows: each text entry in the corpus represents some discursive characteristics and strategies that are intended to influence people’s cognitive systems. Consequently, each text entry can be studied as a representative sample, conceptualizing sociocultural challenges/threats.

The second cognitive criterion was established as follows: each text entry in the corpus represents various types of challenges/threats, so keywords that are verbalizations of concepts THREAT, CHALLENGE should be found. The “back-and-forth” conceptual analysis should be employed: from real-world challenges to texts and from texts to the initial list of threats and challenges.

The third cognitive criterion for text entry arrangement stated: conceptual domain “sociocultural challenges/threats” is structured from a cognitive point of view as a frame with slots, representing relevant conceptual structures.

The conceptual domain “sociocultural challenges/threats” was analyzed and a frame of the domain was built [5].

So, frame “sociocultural threats” was by design implemented as a cognitive foundation for the corpus. Figure 1 demonstrates its slots: social threats, cultural threats, communication threats, threats to human rights, ideological and epistemological threats, ethical and value-related threats. The central concept of the frame is a loss of identity. It is central because a loss of identity is crucial for a person and can result from all the other types of sociocultural threats.

Figure 2 shows the attributes of text entry marked fragment. A special corpus navigator was worked out to facilitate the search in the corpus for agents, causing factors, patients of threat, counteragents, and aspects of identity loss.

The method of conceptual analysis was also employed to study conceptualizations of agents of threat in modern Russian-language discourse. A study of threat-representations in the corpus led to identifying groups of agents that are able and intended to lay a scheme to threat and perform it or bring the idea of threat to the population of Russia.
Fig. 1 Frame “sociocultural threats”

Fig. 2 Attributes of text entry marked fragment (contributed by D.E. Orlov)
Window “Agent” in figure 2 offers a choice of external forces; internal forces; new social actors; cyberspace actors; forces, using innovative technologies to the detriment of humanity.

Window “Patient” in figure 2 demonstrates relations between agents and patients, undergoing threat. Patients, undergoing threat and experiencing fear are in the state of unrest, need of defence. Patients in modern Russia are children and teenagers, intellectuals, migrants, minorities, etc.

Some groups of patients demonstrate attempts to counter the threat, acting as counteragents.

This tool in the corpus allows downloading marked text fragments so that all text entries could be analyzed in accordance with the frame.

**CONCLUSION**

Combining corpus building with linguistic frame analysis resulted in more precise configurations of cognitive models in the conceptual domain “sociocultural threats/challenges”.

To sum up, below are three cognitive criteria that serve as a basis for text entry arrangement on the one hand and contribute to Cognitive Corpus Linguistics on the other.

*Each text entry can be studied as a representative sample, conceptualizing sociocultural challenges/threats.*

*Keywords that are verbalizations of concepts THREAT, CHALLENGE should be found.*

*The “back-and-forth” conceptual analysis should be employed: from real-world challenges to texts and from texts to the initial list of threats and challenges.*

*Conceptual domain “sociocultural challenges/threats” is organized from a cognitive point of view as a frame with slots, representing relevant conceptual structures.*

Going back to cluster integration we can conclude, that a joint effort of researchers in the humanities (cognitive linguistics, literary studies, history, narratology, discourse analysis, corpus linguistics) facilitated a breakthrough in such a complicated area of study as sociocultural challenges and threats in modern Russia.

Cognitive Corpus Linguistics was aligned theoretically with other approaches to the study of a complicated social phenomenon. As Stefan Th. Gries puts it: “It’s time to finally recognize this connection between corpus linguistics, cognitive linguistics and psycholinguistics” [6]. Moreover, J. Newman sees the prospects of cognitive linguistics as follows: “Corpora are a natural source of data for cognitive linguists, since corpora, more than any other source of data, reflect “usage” – a notion which is often claimed to be of critical importance to the field of cognitive linguistics. Corpora are relevant to all the main topics of interest in cognitive linguistics: metaphor, polysemy, synonymy, prototypes, and constructional analysis [7]. The article demonstrates that a corpus is a useful tool for the analysis of complicated social phenomena on the discourse level.
ACKNOWLEDGEMENTS

The article is prepared with financial support of Russian Science Foundation grant 17-78-30029

REFERENCES


