

Text Analysis in the Paradigm of Cognitive Research Based on Computer Linguistics

*Barotova Mubashira Barotovna*¹

Abstract: Theoretical approaches were considered and examples were given that had been obtained in the course of our experiments on the study of texts, before total experiments with informants, the main emphasis was given on the study of the text.

Key words: Super-phrasal unity, external, internal, native speaker, lexical and semantic compatibility, perception, understanding, interpreting and etc.

The text is the main constructive unit of the language and, as already mentioned, the basic linguistic context in which units of lower levels are realized (word, collocation, syntagma, utterance (phrase), super-phrasal unity and compositional fragment). Constructiveness and basic of the text seems obvious, but once again we will refer to a brief and authoritative V.B. Kasevich's formulation: "being an integral unit, the text reveals in relation to its structural components (super-phrasal units / paragraphs, statements, especially words) the property of non-addition: characteristics the text cannot be fully deduced from the features of its constituents. First of all, the meaning conveyed by the text is irreducible to the sum of the values of the components" [1].

Separate form implies, on the one hand, the presence of signals of the beginning and end, and on the other hand, the idea of frames: the knowledge of native speakers about the structure of texts of different functional styles (textual and communicative competence) [2]. Allocate "external" and "internal" (semantic) connectivity.

I. Bellert defines a connected text as "such a sequence of statements S_1, \dots, S_n , in which the semantic interpretation of the utterance S_i (for $2 < i < n$) depends on the interpretation of the utterances in the sequence S_1, \dots, S_{i-1} ". We can say that the coherence and integrity of the text is based on the interconnectedness and interdependence of its structural components. Connectivity is implemented as spatial (contact located structural components), "logical" and associative. The integrity and coherence of the text are important, but difficult to formalize text characteristics. Wholeness is usually defined as having a single theme (subject area, set of situations). The property of coherence (coherence) refers to the structural organization of the text. At the same time, semantic (thematic) and syntactic coherence are distinguished. (Among the formalized means of semantic connection, for example, linking words (conjunctions, words with temporal and causal meanings) and mechanisms of reference and coreference (repeating words in the text, other types of renomination). The syntactic coherence of the text - and statements as structural components of the text - is expressed, first of all, through the semantic-syntactic structure of these units. Researchers of text coherence use different terminology.

Recent studies increasingly separate cohesion and coherence. Cohesion is the connection of text elements, in which the interpretation of some elements depends on others [3]. Coherence is correlated with pragmatic side, it takes us beyond the limits of the text into a communicative situation and spirals into the knowledge base of the addressee. Coherence is most connected with the presumption of meaningfulness and realization of (semantic) expectations of the addressee. However, in real models of text comprehension by a native speaker, it is clear to distinguish between these two different types of connectivity is impossible.

In the procedures of speech activity, integrity and coherence are realized through the mechanisms of contextual predictability. It is natural to assume that if we take an arbitrary point within the text that corresponds to the boundary between certain linguistic units, then the characteristics of its right immediate "neighbor" will be far from random. Apparently, in addition to other structural characteristics, the text can be described as the interaction of metaphorically understood "curves of forces of connections between words" - or between more complex units of text, where some positions will have a strong impact on what may appear on the right, and others will predict their immediate "neighbors" rather weakly. The multiplicity of such curves is determined by a set of features and parameters by which the binding is carried out. The nature of these connections/predictability can be of different origin: (1) associated with lexical and semantic compatibility/incompatibility, (2) determined by the rules of syntax, (3) correlated with informational significance, (4) given by communicative situation in general and the task of communication in particular. Predictability can be more complex, when positions are predicted not by the characteristics of the immediate "neighbor" (preceding element), but on the basis of the listener's knowledge of the semantic coherence and / or integrity (topic, sense of the text). The strength of connections between words (rarely more complex units of analysis) is well described and predicted in mathematical network models. However, these models still have a natural limitation in the form of the already mentioned

¹ Bukhara engineering-technological institute

set of connections of different types in their linguistic nature, most of which are still poorly understood. We would like to hope that in the near future the possibility of such modeling will be significantly expanded - with varying types of units and contexts - taking into account various features and parameters. Such work, apparently, can be carried out by connecting specially selected and linguistically balanced collections, when each task has its own collection (or set of collections).

Naturally, during a communicative act, a person continuously plans (programs) his speech or his perception, realizing necessary adjustments, switching, etc. From this point of view, each next unit must be somehow "verified" and consistent with what has already been said (or perceived) by the current moment. The accuracy of the forecast is evaluated in the applied direction, which until now has only English the name "readability" (which corresponds not so much to "readability" as to "understandability" of the text, i.e. the correct extraction of meaning even with fluent reading or the presence of distortions).

Apparently, the minimum "reconciliation window" ("analysis window") is equal to one unit (for example, one statement or one word); minimum the necessary forecasting is at the same time as if typical, statistically predominant, while the maximum forecasting is determined by the text and the communicative situation as a whole.

The founder of this approach, Marvin Minsky, defines a frame as a data structure designed to represent a certain typical situation [5]. For example, there are frames of household, business and scientific communicative situations that allow predicting the development of events in these situations (in particular, the generation and perception of texts of different functional styles). Knowledge by the addressee (hearer) of the corresponding frame, apparently, correlates with the knowledge of the addressee of the meaning (integrity) and semantic coherence of the text, where the text acts as an implementation of this frame.

It is essential to oppose the following types of goals and, accordingly, research procedures for studying texts: o understanding and interpreting a text by a person, which is done in line with the traditional and / or cognitivist approaches (see, for example, the works of M.B. Bergelson [6], as well as the works foreign authors (partially discussed below); o in the spirit of applied tasks - automatic understanding of the text (or, for example, automatic extraction of information from text, machine translation tasks, automatic summarization. The difference of such approaches suggests placing different native speakers at the center of the study. In the case of applied research as "artificial native speaker" is an automaton. natural consequence. Such a difference is the degree of involvement of what can be called a "knowledge base", which allows predicting the development of events based on knowledge of the types of communicative situations (non-linguistic data). It is obvious that the automaton "experiences difficulties" in the formation of some the macrostructure of the text, which is the result of the functioning in the procedures of perception (understanding, interpretation) of not only the structural components of the text, but also the so-called background and inference knowledge. The degree of involvement of background and inference knowledge seems to depend on the type of frame and on the knowledge communicant of this frame.

Let us briefly dwell on the most fruitful positions of modern perception and understanding of the text. The starting point is that "A coherent text is more than a language in itself and much more than a sequence of single sentences." The procedures for perception and understanding of the text are traditionally treated as multilevel. However, questions such as the number and nature of levels, the interaction of these levels, etc., require investigation. A.S. Stern identified three levels of perception: sensory, perceptual and semantic. These three levels are distinguished mainly on the basis of psycho-physiological criteria perception and processing of information, but not language criteria; in particular, the sensory ("lower") level is not linguistic. In the work devoted, however, to the understanding of the written text, the following five levels are distinguished: surface structure, the propositional text-base, the referential situation model, the communication context, and the discussion genre. The first three levels are traditionally accepted by most psycholinguists, starting with work . At the level of the surface structure, the addressee works with such units as words (probably even word forms) and the surface structure of the clause (the structural component of the text, characterized by semantic, syntactic and prosodic integrity, but not exceeding some critical volume. The base text is usually structured set (system) of propositions" probably, while we are talking only about explicitly expressed propositions. The situation model relates the addressee to the meaning of the text; both the text itself (explicit text, text in the narrow sense) and background knowledge of the addressee take part in its construction. The communicative level correlates with the pragmatic components of the communicative situation. The level of functional style and / or speech genre (text genre) in foreign studies is correlated with various classes and subclasses, largely corresponding to those identified by a number of researchers. Some provisions and terms in Grasser's scheme are not indisputable and require clarification. In particular, the structure called "text base" (text base) in various works is understood in two ways:

- 1) as a system of propositions, as a rule, corresponding to individual statements of the text;
- 2) as a macrostructure, for the emergence of which it is important (even necessary) to use background and inference knowledge;

Individual propositions enter into this macrostructure as members, entering into certain hierarchical relations [7]. Background knowledge fills in the semantic gaps that are inevitable in almost any text, while inferential knowledge draws consequences from propositions and their combinations and introduces an element of order introduced by the addressee. Thus, the second understanding of the "base text" is closer to the situation model, and the degree of proximity is

determined by the degree of involvement of the addressee's background and output knowledge in the perception and understanding of the text.

According to F. Johnson-Laird, the structure of propositions is only one of the types "semantic record" that a person uses when perceiving a text and for storing the results of this process in memory. The other two kinds are "mental models" and images. "The status of the latter is the least clear, although it is claimed that there is a mapping of propositional structures onto mental models, and the latter onto images. As for mental models, such a model is understood as a direct reflection of the situation described by perceived text. The concepts used in the Johnson-Laird hypothesis correlate with linguistic, psycholinguistic and psychological notions. Macrostructure of text semantics as a hierarchy of propositions plus background and inferential knowledge is comparable to an extended view of the framing system valid for a given text. A mental model is a system of self-cognitive frames, a percept, corresponding to the semantics of the text. <...> "image" in the described system can be understood as meaning - or a certain system of meanings - in the spirit of L.S. Vygotsky and A.N. Leontiev".

Semantic representations of different levels (of different cognitive depths), as a rule, coexist, representing different stages of processing the same text. However, depending on the communicative goals (text style and personal attitudes), a person can set the level of understanding Johnson-Laird writes that in the case of statements that reflect stereotypical (determinate) situations, the depth is much higher: the subjects remember better the meaning of statements than their linguistic form. In the opposite situation - for statements that reflect non-stereotypical, somewhat unusual situations, the cognitive depth is much less: the linguistic form, the surface structure of the utterance (up to the literal composition) is better remembered. This kind of experimental data demonstrates "the absence of a forced set of procedures and operations in the processes of speech perception (and, respectively, storing information in memory and retrieving it from memory). The addressee determines the strategy of perception and the depth of understanding of the text. Distinguishing "superficial perception (and understanding)" from "perception, understanding and interpreting the text" is consistent with attempts to highlight the differences between the "text base" and the "situation model" in the procedures of text perception. In another work - on the material of written texts - it is shown that the relationship between the levels - the level of the situation model and more superficial levels (surface structures and the base of the text) - in the procedures of understanding can depend on the functional style of the text (journalistic vs. literary and artistic).

In conclusion of a brief review of cognitive and speech activity ideas about the procedures for understanding the text, we present psycho-physiological substantiation of the possibility of functioning of the simulated procedures. It is assumed that short-term memory is involved at the level of the surface structure. About two sentences can be held in working memory (to a very large approximation); is activated there and the most important information functions, which can be correlated with sufficiently high levels of text presentation. In experimental work with native speakers, the manifestation of psycho-physiological limitations - in particular, the work of various types of human memory - is significantly determined by conditions of communication or (in the terminology described above) communicative context and functional style of the text.

References:

1. Iordanskaja, L., Paperno, S. A Russian-English Collocational Dictionary of the Human Body, Columbus/Ohio. 1996.
2. Johnson-Laird P. Mental models: Towards a cognitive science of language, inference and consciousness – Cambridge, 1983.
3. Whitney P., Budd D., Bramucci R. S. Crane RS. On babies, bathwater, and schemata: a reconsideration of top-down processes in comprehension // *Discourse Process*. 1995, 20
4. Хоккет Ч. Проблема языковых универсалий // *Новое в лингвистике*. Вып. V – М.: Иностр. лит., 1970
5. Yagunova E., Savina A. The Experimental Study of Terminology Collocations: Calculations and Experiments with Informants // *Proceedings of the workshop on creation, harmonization and application of terminology resources: CHAT 2011. NEALT Proceedings Series*. May 11, 2011. / Eds. Tatiana Gornostay, Andrejs Vasiljevs and Inguna Skadiņa
6. Zwaan R. A. Effects of genre expectations on text comprehension // *J. Exp. Psychol. Learn. Mem. Cogn.* 1994, 20.
7. Mubashira Barotova. (2021). SOME FEATURES OF TRANSLATING ORIGINAL LITERARY TEXT. *JournalNX - A Multidisciplinary Peer Reviewed Journal*, 7(03), 369–373. Retrieved from <https://repo.journalnx.com/index.php/nx/article/view/2788>
8. Mubashira Barotova. EditorJournals and Conferences. (2021, October 28). SOME SPECIFIC FEATURES AND DIFFICULTIES OF THE TRANSLATION OF THE NOVEL "JANE AIR" BY CHARLOTTE BRONTË. <https://doi.org/10.17605/OSF.IO/H4EWG>
9. Barotova, M. (2020). THE PROBLEMS OF RECREATING WRITERS STYLE IN TRANSLATION. *Theoretical & Applied Science*, (2), 189-192. <https://www.elibrary.ru/item.asp?id=42658997>
10. Barotovna, B. M. (2021). DEVELOPMENT OF LISTENING SKILLS AT THE SENIOR STAGE OF TEACHING ENGLISH. *Web of Scientist: International Scientific Research Journal*, 2(10), 119-127. <https://wos.academiascience.org/index.php/wos/article/view/398>

11. Baratovna, B. M. (2018). The concept of pedagogical technology in educational practice. *Научные исследования*, (6 (26)), 56-57. <https://cyberleninka.ru/article/n/the-concept-of-pedagogical-technology-in-educational-practice>
12. Barotovna, B. M. (2017). Typology of English and Russian phonological systems. *Вопросы науки и образования*, (5 (6)), 85-87. <https://cyberleninka.ru/article/n/typology-of-english-and-russian-phonological-systems>
13. Baratovna, B. M. (2018). Modern integrations and their benefits in teaching process. *Научные исследования*, (3 (22)), 53-54. <https://cyberleninka.ru/article/n/modern-integrations-and-their-benefits-in-teaching-process>
14. Khudoyberdievna, S. Z. (2022). Modern Methods of Translating Phraseological Units. *Eurasian Research Bulletin*, 4, 153-158. <https://geniusjournals.org/index.php/erb/article/view/516>.
15. Khudoyberdievna, S. Z. (2017). Didactic games as framework of students in cooperation. *Научный журнал*, (3 (16)), 48-50. <https://cyberleninka.ru/article/n/didactic-games-as-framework-of-students-in-cooperation>.
16. Z Saidova., ADVANTAGES AND DISADVANTAGES OF MODULAR OBJECT-ORIENTED DYNAMIC LEARNING ENVIRONMENT (MOODLE) IN THE SYSTEM OF EDUCATION. https://journal.buxdu.uz/index.php/journals_buxdu/article/view/435.
17. Saidova, Z.(2022). ИЗУЧЕНИЕ ФРАЗЕОЛОГИИ И СРАВНИТЕЛЬНЫЙ АНАЛИЗ ФРАЗЕОЛОГИЧЕСКИХ ЕДИНИЦ, ВЫРАЖАЮЩИХ ПСИХИЧЕСКОЕ СОСТОЯНИЕ ЧЕЛОВЕКА. ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu. Uz), 8 (8). http://journal.buxdu.uz/index.php/journals_buxdu/article/view/4354.
18. Saidova, Z.(2022). ИЗУЧЕНИЕ ФРАЗЕОЛОГИИ И СРАВНИТЕЛЬНЫЙ АНАЛИЗ ФРАЗЕОЛОГИЧЕСКИХ ЕДИНИЦ, ВЫРАЖАЮЩИХ ПСИХИЧЕСКОЕ СОСТОЯНИЕ ЧЕЛОВЕКА. ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu. Uz), 8 (8). http://journal.buxdu.uz/index.php/journals_buxdu/article/view/4354.
19. Khudoyberdievna, S. Z. (2021, November). Language Expressing Psychoemotional State of Human. In *INTERNATIONAL CONFERENCE ON MULTIDISCIPLINARY RESEARCH AND INNOVATIVE TECHNOLOGIES* (Vol. 2, pp. 108-113). <https://mrit.academiascience.org/index.php/mrit/article/view/102>.
20. Khudoyberdievna, S. Z. (2022). Classification of verbal phraseological units denoting the emotional state of a person. *Integration of Pragmalinguistics, Functional Translation Studies and Language Teaching Processes*, 90-93.
21. Khudoyberdievna, S. Z. (2018). Implementation of some techniques in developing reading skills in English classes. *Достижения науки и образования*, (5 (27)), 59-60. <https://cyberleninka.ru/article/n/implementation-of-some-techniques-in-developing-reading-skills-in-english-classes>