

# Systematic Semantic Analysis of Texts

Farzona Sh. Nasreddinova<sup>a</sup> and Farangiz A. Khamrakulova<sup>b</sup>  
*Samarkand State Institute of Foreign Languages, Samarkand, Uzbekistan*

**Keywords:** English Language, Analysis, Semantic Analysis, Systematic Text Analysis, Classification Methods.

**Abstract:** The scientific article is devoted to the consideration of ways to solve the problems of systematic text analysis and semantic text analysis. Systematic-analytical texts require interesting data search and analysis. Researchers in the field of automatic text processing are systematically moving from the simplest analysis methods to more complex methods, gradually approaching a semantic representation of the text that corresponds to human perception, but it seems that linguistics is completely imitating reality.

## 1 INTRODUCTION

In modern linguistics, any algorithmic systematic analysis model of a language makes greater or lesser assumptions when it can only be partial or complete. Partial analyses reveal only a part of the language, i.e., its mechanisms. Partial modelling usually reveals an ideal representation of the text. Analysis of modelling does not take into account errors in expression, therefore, combining models into a single complete system that simultaneously models all the mechanisms of language requires a separate systematic analysis.

The problems of semantic text analysis include many areas such as searching, sorting, and classifying documents in local and global networks. The starting point for analysis is usually large arrays of unstructured or semi-structured natural language text. In this case, it contains correspondences between some key objects and documents used for analysis. The simplest form can be the subject of phrases and statements or whole phrases or sentences in the analysis. In addition, similarities can be found.

## 2 LITERATURE REVIEW


Currently, there are many ways to express the systematic analysis of the text, but none of them is perfect. Many scholars have worked on interlinking


the analysis of the text. Thus, I.A. Melchuk introduced the concept of lexical function, developed the concepts of semantic analysis, and considered them in the context of an explanatory dictionary, which is a language model. In the process of analysis, he shows that the meanings of words are not directly related to the surrounding reality, but to the native speaker's ideas about this reality. V. Sh. Rubashkin and D. G. Lahuti introduced a hierarchy of systematic relations for a more efficient operation of semantic analysis. The famous linguist E. V. Paducheva suggests considering the thematic classes of words, in particular verbs, because they carry the main semantic load.

In this approach, the idea of dividing language concepts into certain semantic groups is important, taking into account that these concepts have some insignificant common semantic component. Elements of such groups will have the same related concepts.

Analysing universal studies should be convenient for discovering new knowledge, that is, it is necessary to model the text to analyse the accuracy of the texts. This is where systematic analysis comes in handy.

The semantic analysis proposed by V.A. Tuzov contains formalisms of predicate logic in languages, functions on these concepts, and conclusions that can describe new concepts. In the future, scientific thinking may develop in the direction of creating such semantic languages.

<sup>a</sup>  <https://orcid.org/0000-0003-3961-9700>

<sup>b</sup>  <https://orcid.org/0009-0003-7116-4934>

### 3 RESEARCH METHODOLOGY

Despite the apparent simplicity of the task of classifying and systematizing the analysed texts, and identifying text topics, it is very difficult to implement them. The problem cannot be solved satisfactorily by relying only on keywords or the syntactic structure of simple phrases. The use of general semantic analysis alone does not fundamentally change anything. Existing systematic analyses provide classification accuracy about assessment: predefined predictive analysis - about 5%, using predefined analysis and adjusting the topic of texts - up to 80%.

Text synthesis. In a narrow sense, text synthesis refers to the construction of natural language phrases and sentences from formal language records. Structured phrases may or may not be subject to the requirement of stylistic correctness, but in any case, they should not contain semantic and grammatical errors.

Checking the correctness of texts. This is due to the need to fully analyse the sentences, with the help of which you can check the grammatical correctness of the texts.

In the process of systematic analysis, the degree of automation is that all definitions can be automatically checked for the consistency of the collected definitions. An alternative approach could be one in which definitions of concepts are created from existing texts with such descriptions and then revised as necessary in the process of communication with an expert. To implement this approach, it is necessary to be able to analyse the semantics of texts in detail.

The essence of explaining the terms in the text is to form a brief description of the main analysis of the text. There are two different comment options. In the first case, a small number of sentences in the text are identified and analysed, which fully reflect the main themes of the text. In the second case, the main themes of the text are identified as meanings, and these meanings are expressed through new sentences and text. The last option is preferable, but it is also more difficult. All modern abstract annotation systems are based on the first option.

It is called classification and categorization of documents, identification of document topics, and automatic abstracting and annotation. This is a relatively young field, and most of the important results have been obtained in recent years. First of all, this is due to the emergence of very large volumes of textual data available to everyone and the emergence of computing power corresponding to such volumes. Text analysis systems operate on a set of documents whose words are considered features. In addition, the

size of such documents can be very large, and the total vocabulary for all documents can reach several hundred thousand words.

### 4 RESEARCH FINDINGS

After testing such an analytical system, we immediately see that the most frequent words are compound adverbs that have almost no effect simultaneously. Such words are called stop words and are removed from documents before being converted into a vector model. In addition to the general vocabulary of words, it is useful to compile your vocabulary for each specific task. Another preprocessing method besides removing words is to highlight the important part of the word.

The following algorithm of systematic analysis is used and used in everyday life. By creating an electronic text rule, we indirectly control the decision rules that use many systematic analyses. The non-rule nodes contain the type of "questions" to the document, while the leaves contain the answers in the form of the resulting category. "Questions" can be asked by the user himself, as in the example above, or calculated based on a training sample, in which case they usually take the following form: "Do such words exist in the text?"

The simplicity of the analysis is offset by the complexity of building such a tree of questions from a set of systems. In addition to classification, structured decisions can be used to analyse the structure of documents and categories, where rules can be valuable. In practice, the decision is mainly used for this purpose, because in terms of classification quality, they are much lower than systematic models, which will be discussed later.

### 5 CONCLUSION

The article discussed the problems and methods of semantic text analysis, but how does one evaluate how correct the result of a particular method is? A text with certain categories is divided into two parts: one is taught and the other is analysed. It is assumed that the documents to be systematically classified will be like the documents in the test sample. Of course, this may not be the case at all, but, unfortunately, there is no other way to evaluate the quality of the classification. The generally accepted characteristics of classification quality are accuracy and completeness. Accuracy is calculated as the ratio of

the number of correct positive predictions by the classifier (when a document is given a category) to the total number of positive predictions. Recovery is the ratio of the number of correct positive predictions to the number of documents to be assigned a category. Classifiers with high precision usually have low recall and vice versa.

## REFERENCES

- D.G.Lahuti, Rubashkin V.Sh. Semantic dictionary for information technologies // Scientific and technical information. 2000.S. 1-9.
- E.V. Paducheva Dynamic models in the semantics of vocabulary. M.: Languages of Slavic culture, 2004. 608 p.
- Semantic analysis of natural language texts URL: [http://studopedia.su/10\\_45700\\_semanticheskiy-analiz-estestvenno-yazikovih-tekstov.html](http://studopedia.su/10_45700_semanticheskiy-analiz-estestvenno-yazikovih-tekstov.html) (access date: 12/21/2016).
- V.A.Tuzov Computer semantics of the Russian language. SPb.: Publishing house St. Petersburg. Univ., 2004. 400 p.
- Is'haqov. M.M, Jumayev G.I. Representation of diplomatic and embassy relations in the works" Habib us-siyar" and" Matla us-sa'dayn". Sharqshunoslik. Востоковедение. Oriental Studies, – Tashkent: – B 29-31.