PROBLEMS OF MODELING PHRASEOLOGICAL UNITS IN A SEMANTIC ANALYZER

Nurboyeva Maftuna Vahobjon qizi
Doctor of Philosophy (PhD) in Philology,
Lecturer at Millat Umidi University

Abstract. The article discusses the specific features of morphological, syntactic, and semantic analyzers used in natural language processing systems and their significance in the automatic analysis of linguistic units. Approaches to the development and improvement of semantic analyzers in foreign and Uzbek linguistics are analyzed, with particular attention paid to the role of phraseological units in semantic analysis and the problems of their linguistic modeling. The study also examines the possibilities of improving the efficiency of automatic semantic analysis through modeling the semantic features of phraseological units.

Keywords: semantic analyzer, morphological analyzer, syntactic analyzer, semantic analysis, natural language processing, linguistic modeling, phraseological unit, phraseological model, automatic analysis.

When speaking of the semantic analyzer, it is first necessary to clarify the concept of the analyzer. As can be seen, the word “analyzer” derives from the word “analysis” and is directly related to this concept. “Derived from the Greek ‘analysis’ (‘decomposition’), this word denotes a method of investigation that people use to learn about things unknown to them; unlike synthesis, which examines an object as a whole, analysis means dividing a concept into its component parts and studying them separately. In reaching a final judgment, the knowledge obtained through analysis and synthesis is combined. In this respect, using these two methods separately is of no use, since analysis cannot produce a holistic judgment without synthesis, and synthesis cannot give deep knowledge about a thing or phenomenon without analysis” [1].

Initially used in physiology and, at the same time, in biology, chemistry, physics, and psychology, this term, "with the development of corpus linguistics, has become widespread in linguistics as well and denotes a tool that examines a word from its morphological, semantic, and syntactic aspects in the process of natural language processing in automatic search systems" [2]. In linguistics, there are the following types of linguistic analyzers: semantic, morphological, and syntactic analyzers. A set of algorithms that determine the grammatical features of words constitutes a morphological analyzer, while a syntactic analyzer studies the syntactic features of words and their combinability. In the semantic analyzer, whose basis is semantic analysis, the process of identifying and eliminating lexical ambiguity is carried out. The aim is to remove lexical and grammatical homonymy in the text and to facilitate the syntactic analysis process. Among Russian corpus linguistics projects, the DikTUM-1 system is of particular importance; it performs the task of lexicographic processing of text and includes the following analysis processes: Morphoanalyzer, SegmMorf (performs morphemic analysis of word forms), Lintez (performs semantic analysis of context units), PsiAn (psychological and semantic analysis of context), and Janr (stylistic and statistical analysis) [3]. It is evident that the role of the semantic analyzer is unparalleled in fully forming the Uzbek National Corpus system.

A number of studies on semantic analyzers have been conducted in world linguistics and Turkology. In particular, A.K. Asiryan, in his research, tested several methods for the morphological analysis of text, studied separately the positive and weak aspects of each method in carrying out the analysis, and emphasized that the characteristics of the text and the purpose of the analysis must also be taken into account when choosing a suitable method for morphological analysis of text [4]. A.V. Dybo and A.V. Sheymovich, in turn, developed the theoretical and practical foundations of a general automatic morphoanalyzer for Turkic languages such as Tatar, Kazakh, Shor, Tuvan, Kyrgyz, Karakalpak, Uzbek, and Uyghur, and carried out the

analysis on the basis of the analysis algorithm developed by F. Krylov within the StarLing system [5]. There are also other studies devoted to the morphological analysis of Turkic languages in which special algorithms and models were developed to automate morphological analysis, taking into account the agglutinative nature of the Turkish language [6]. T.V. Batura, in turn, conducted research on the semantic analysis and representation of text in computational linguistics and on creating a semantic model of context [7]. In addition, researchers such as M.G. Malkovsky, A.S. Starostin, E.A. Kanevsky, N.V. Kolpakova, D.H. Younger, N.A. Israilova, P.S. Bakasova, E. Klein, E. Loper, M. Korobov, A.D. Moskvina, D. Orlova, O.A. Mitrofanova, P.V. Panicheva, O.N. Lyashevskaya, I. Astafyeva, A. Gareyshina, and A. Bonch-Osmolovskaya have also conducted research in this field [8–10].

In Uzbek linguistics as well, research has been conducted on methods of creating a morphological analyzer, on ways of using the NLTK package of the Python language in it, and on creating and applying program code [11–12]. In particular, Sh. Hamroyeva has developed research and teaching manuals on creating linguistic support for morphological analysis, through which the analysis process can be carried out correctly and accurately [13]. Sh. Gulyamova, in turn, has carried out research on modeling homonymous, polysemous, and polyfunctional words and applying them to the semantic analyzer [2]. X. Axmedova has also conducted research on the semantic analysis of sentences in Uzbek and on developing algorithms and mathematical models for the automatic identification of homonymous, polysemous, and polyfunctional words within sentences [14]. It can be observed that the semantic analyzer is a relatively understudied area in the Uzbek national corpus system, and there are many issues in this direction awaiting research. This is because, for the semantic analyzer, morphological and syntactic analysis in the system must first be improved; moreover, while applying morphological and syntactic analyzers to the corpus through special

formulas involves relatively little complexity, the semantic analyzer requires “reading” meaning from the system, which also explains why it is rather difficult to implement.

References

1. ЎзМЭ. 1-жилд. – Тошкент, 2000. – 879 с.
2. Гулямова Ш. Ўзбек тили семантик анализаторининг лингвистик асослари: Филол. фан. д-ри ... дисс. – Фарғона, 2022. – 281 б.
3. Амиева А.М., Филимонов В.В., Сергеев А.П., Тарасов Д.А. Инструменты корпусной лингвистики // https://elar.urfu.ru/bitstream/10995/36257/1/ittsm-2016-34.pdf (murojaat sanasi: 09.04.2025).
4. Асирян А.К. Сравнение инструментов морфологической разметки / Intellectual potential of the XXI century 2017 / http://www.sworld.education (murojaat sanasi: 10.04.2025).
5. Дыбо А.В., Шеймович А.В. Автоматический морфологический анализ для корпусов тюркских языков // Филология и культура. Philology and culture. – 2014. №2(36). – С. 20–26.
6. Гатиатуллин А.Р., Баширов А.М. Морфологический анализатор тюркских словоформ на базе структурнофункциональной модели тюркской морфемы // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Казань, 2017. – С. 319–327.
7. Батура Т.В. Семантический анализ и способы представления смысла текста в компьютерной лингвистике. Программные продукты и системы / Software & Systems 4 (29) 2016. – С. 45–56.
8. Мальковский М.Г., Старостин А.С. Модель синтаксиса в системе морфо синтаксического анализа «TREETON» //Тр. междунар. Конф. Диалог 2006. – М., 2006. – С. 481–492
9. Исраилова Н.А., Бакасова П.С. Морфологический анализатор кыргызского

языка // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Труды конференции. В 2-х томах. Т 2. – Казань: Издательство Академии наук Республики Татарстан, 2017. – 327 с.

10. Ляшевская О.Н., Астафьева И., Бонч-Осмоловская А., Гарейшина А., Гришина Ю., Дьячков В., Ионов М., Королева А., Кудринский М., Литягина А., Лучина Е., Сидорова Е., Толдова С., Савчук С., Коваль С. Оценка методов автоматического анализа текста: морфологические парсеры русского языка // Компьютерная лингвистика и интеллектуальные технологии. – 2010. № 9 (16). – С. 318–326.
11. Elov B., Hamroyeva Sh., Elova D. Morfologik analizatorni yaratish usullari. O‘zbekiston: til va madaniyat (Amaliy filologiya), 2022, 5(1). – B. 67–87
12. Elov B., Hamroyeva Sh., Xusainova Z. NLP (tabiiy tilga ishlov berish)ning vazifalari va zamonaviy yondashuvlar. TerDU, Filologik tadqiqotlar: til, adabiyot, ta’lim. 2022. – B. 5-6.
13. Ҳамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. – GlobeEdit, 2020. – 268 б.
14. Axmedova X.I. O‘zbek tilidagi gaplarni semantik analiz qilishning modeli, algoritmlari va axborot tizimini ishlab chiqish: Filol. fan. d-ri ... diss. av-ti. – Toshkent, 2023. – 52 b.