JOURNAL OF INTELLECTUAL PROPERTY AND HUMAN RIGHTS

Volume: 4 Issue: 3 | Mar–2025 ISSN: 2720-6882
http://journals.academiczone.net/index.php/jiphr

The Issues of Giving Meronyms in the Corpus

Mustafaeva Sojida Ulashevna
Termez State Engineering and Agrotechnologies University,
Department of Uzbek language and literature teacher
sojidamustafayeva557@gmail.com

Annotation: This thesis about meronyms, which represent part-whole relationships and its several challenges when are given in a linguistic corpus. One major issue is ambiguity in part-whole relations, where the same part may belong to multiple wholes depending on the context (e.g., "wheel" as a part of both a car and a bicycle). Additionally, the paradox of inconsistencies arise when meronyms exist at different hierarchical levels, making it difficult to determine the appropriate level of annotation (e.g., "leaf" as a part of "branch" vs. "tree"). Another challenge is context-dependent meronyms, where a word functions as a meronym only within specific domains (e.g., "petal" is a part of "flower" only in botanical contexts). Furthermore, lexical variation and synonymy complicate the notion, as different terms may refer to the same part in different linguistic or dialectical differences (e.g., "tire" vs. "wheel" for a car component). Moreover, multiword and compound expressions present difficulties in corpora, as some meronyms appear in phrases rather than single words (e.g., "front door" as a part of a house). Lastly, cross-linguistic variation makes lexemes cognitively complicated, as languages differ in how they express part-whole relationships (e.g., some languages use a single term for both "hand" and "arm) such as in Uzbek language. Identifying these challenges requires clear explanations, hierarchical structuring of meronyms, and clarify concrete strategies to ensure consistency in corpus analysis.

Key words: meronyms, part-whole relations, corpus, computer corpus, linguistic corpus, semantic structures, word levels.

Introduction:

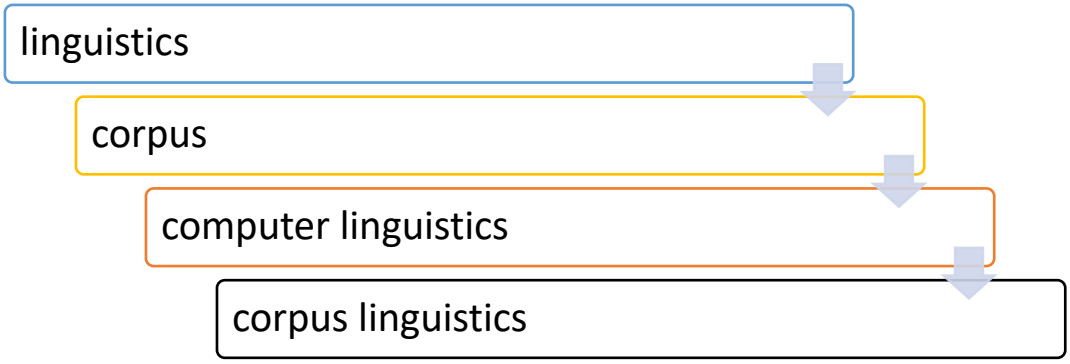
General information about linguistic corpus

At a time when world computer linguistics is rapidly developing as an integrated field, it serves as an important tool in solving language problems through a number of methods and methods. A number of scientific achievements are being achieved through the positive influence of computer technologies on language development and, conversely, language technology on computer technologies, and many scientists in the world are consistently conducting research in various fields of computer linguistics [N. Abdurakhimova., Corpus Lingvistikasi.-Darslik., Tashkent - 2023., p.5]. **Corpus** – [kɔ:pəs] – (noun/noun): 1. "a collection of written texts, especially the entire works of a particular author or a body of writing on a particular subject" "the Darwinian corpus" 2. In anatomy: "the main body or mass of structure" [https://languages.oup.com/google-dictionary-en] – the word corpus comes from the English word corpus, which is a set of linguistic units that form a set of texts collected for a specific purpose, a collection of texts, written or spoken in natural language, stored electronically, and placed in a computerized search engine based on software, on-line or off-line. The term corpus is currently widely used in all fields, especially in the world of linguistics, where the concept of "corpus" is developing as a comprehensive modern technolinguistic framework. If we approach this term more broadly, we will encounter the terms computational linguistics and corpus linguistics, here we will explain both terms: Corpus linguistics is an empirical method of studying language through text corpuses [Meyer, Charles F. (2023). English Corpus Linguistics (2nd ed.). Cambridge: Cambridge University Press. p. 4./ https://en.wikipedia.org/wiki/Main_Page]. "Corpus linguistics is a component

Volume: 4 Issue: 3 | Mar–2025 ISSN: 2720-6882
http://journals.academiczone.net/index.php/jiphr

of computational linguistics, a section that deals with the development of general principles for the construction and use of linguistic corpora (text corpora) based on computer technologies" [Zakharov V.P. Corpus linguistics. Textbook. - St. Petersburg, 2005. -3.c]. So, "corpus linguistics is an independent branch of computer linguistics, which deals with the development of principles for the creation and use of linguistic corpora (text corpora) using computer technologies" [Zaharov V.P. Corpus linguistics. Textbook. - St. Petersburg, 2005. - 48.c] In both concepts, it can be understood that a new branch of the traditional field of linguistics is computer linguistics and its branch is corpus linguistics. So, "corpus linguistics is an independent branch of computer linguistics, which deals with the development of principles for the creation and use of linguistic corpora (text corpora) using computer technologies" [Zaharov V.P. Corpus linguistics. Textbook. - St. Petersburg, 2005. - 48.c] In both concepts, it can be understood that a new branch of the traditional field of linguistics is computer linguistics and its branch is corpus linguistics.

3.1 Picture. the hierarchical distribution of the corpus:



Based on the above ideas, corpus linguistics is a branch of linguistics that studies and analyzes natural languages using a collection of texts (corpus). Its development dates back to the first half of the 20th century, when Bloomfield, Fritz, and Bonders began their work in the 1940s; in the 1960s, the publication of "Computational Analysis of Present-Day American English" (1967)[Francis, W. Nelson; Kučera, Henry (1 June 1967) in the English-speaking world was a milestone in the development of modern corpus linguistics; Henry Kučera and W. Nelson Francis published "Brown Copus" [Francis, W. Nelson; Kučera, Henry (1 June 1967)].

Computational Analysis of Present-Day American English. Providence: Brown UniversityPress.ISBN 978-0870571053./ https://en.wikipedia.org/wiki/Main_Page] is a structured and balanced corpus of American English, which in 1961 contained more than 1 million words, including more than two thousand text samples from various genres. Later, on the basis of this corpus, the "American Heritage Dictionary" was founded in 1969 ["Corpusnaia lingvistika" (A.B. Kutuzov) License Creative commons Attribution Share-Alike 3.0 Unported (Electronic resource) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf] and the Lancaster-Oslo/Bergen (LOB) corpus; Several corpora have been created, such as the London-Lund Corpus (LLC) (1975; contains orthographic transcription, phonetic, prosodic (stressed and unstressed, long and short syllable pronunciation) tags, covering 500,000 word usage cases). In Uzbek linguistics, the peak of the formation of corpora began in 2018 with theoretical research, and in 2021 with the creation of the Uzbek language educational corpus [http://uzschoolcorpara.uz/], and today we can see it in the context of new areas such as Uzbek computer linguistics, natural language processing (NLP), machine learning, and data mining. Just as the progress of the era requires progress and development in all areas, the world of linguistics is unimaginable today without computer technologies, which means that a computer is a machine that can collect and analyze the comprehensive properties of words existing in all languages, and with its help, it allows us to study language units through corpus statistical analysis. Thus, corpus linguistics is expressed as a part of computer linguistics and is explained by the effective use of its effective

Volume: 4 Issue: 3 | Mar–2025 ISSN: 2720-6882
http://journals.academiczone.net/index.php/jiphr

technocomputers on the basis of computer technologies. Corpus linguistics has been a rapidly developing area of world computer linguistics since the 90s of the 20th century, and quite noticeable work has been done in this regard. Since the middle of the 20th century, corpus linguistics has been taught as a discipline in higher education institutions around the world, including its theoretical and practical aspects, features, and aspects such as programming. In this regard, I.Yu. Shemyakin: "The strength of computer linguistics is determined by the ability to manage linguistic signs "saturated" with a certain set of meanings; there is a direct principle of "reciprocity" between the computer and the language; it is also reflected in the concept of computer - program - knowledge - language/language" [Shemyakin Yu.I. The beginning of computer linguistics: Textbook. − M.: Publishing House of MGOU, A/O "Rosvuznauka", 1992.] Thus, "the subject of corpus linguistics is the language corpus.

Discussion and results

General analysis

We scientifically substantiate the issues of one of the word-meaning relationships, meronyms, that is, a word class category that represents a whole object and its elements - parts, and their description, definition, and analysis in linguistic corpora as follows, that is, this issue may include the following:

- identifying, constructing, and defining "whole-part" relationships between words in a corpus as part-whole relationships
- marking lexical and semantic features of words - meronyms - by analyzing their syntactic and semantic properties
- classification and tagging of meronyms belonging to the part-whole class - how meronyms are marked, annotated, and processed by automated systems in the corpus, which accelerates their further semantic clarity
- the frequency and contextual status of meronyms in the corpus helps to study in what combinations they occur and in which fields they are most often used.
- the mechanism for analyzing meronymic word elements in computational linguistics (NLP) is a current issue
- corpus plays a significant role in today's fields, especially in the current era of artificialization of language and its natural mechanisms, as a means of understanding words and revealing their semantic analysis.

Meronyms (part-whole relationships) are an important part of natural language processing (NLP) and semantic framework analysis in this field. Various approaches are used to represent them clearly and systematically in corpora:

1) lexical-semantic models – in databases such as WordNet, meronyms are classified as part-whole relations.

2) syntactic and morphological analysis - meronyms are distinguished in text corpora using units such as "part of...", "component of...."

3) ontological approaches - in databases such as DBpedia and ConceptNet, meronym relationships are defined as special attributes.

4) annotation and tagging – Automated analysis is facilitated by defining part-whole relationships using special tags.

5) accurate representation of meronyms in corpora is important for the development of semantic search systems, artificial intelligence, and automatic translation processes. Therefore, research is ongoing to identify and formalize meronym relationships.

Volume: 4 Issue: 3 | Mar–2025 ISSN: 2720-6882
http://journals.academiczone.net/index.php/jiphr

The issue of meronyms in the corpus is one of the important topics for linguistics and computational linguistics. Determining, defining and analyzing meronymic relations in corpora, studying their syntactic and semantic features enrich the theoretical and practical aspects of linguistics. This process is also important for artificial intelligence and natural language processing (NLP) systems, because it is necessary to identify meronymic relationships in automatic text understanding and semantic analysis. Therefore, the development of methods for recording and effective use of meronyms in the corpus can greatly contribute to the development of linguistic research and information technology in the future.

Conclusion:

Meronymy in a corpus can be analyzed by identifying part-whole relationships through lexical patterns, syntactic structures, and collocations. By using corpus linguistic tools like frequency analysis, concordance searches, and dependency parsing, we can systematically extract and study how meronyms appear in natural language. This approach helps in understanding the semantic structure of words and their relationships in both English and Uzbek

References

1. Francis, W. Nelson; Kučera, Henry (1 June 1967). Computational Analysis of Present-Day American English. Providence: Brown University Press. ISBN 978-6. 0870571053./ https://en.wikipedia.org/wiki/Main_Page
2. Francis, W. Nelson; Kučera, Henry (1 June 1967). Computational Analysis of Present-Day American English. Providence: Brown University Press. ISBN 978-0870571053./ https://en.wikipedia.org/wiki/Main_Page“Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf
3. Meyer, Charles F. (2023). English Corpus Linguistics (2nd ed.). Cambridge: Cambridge University Press. p. 4./https://en.wikipedia.org/wiki/Main_Page
4. N.Abduraximova.,Korpus Lingvistikasi.-Darslik., Toshkent - 2023.,b.5
5. V.Zaxarov., B.Mengliyev., Sh.Hamroyeva., Korpus Lingvistikasi: korpus tuzish va undan foydalanish., – O’quv qo’llanma; Toshkent. -2021., b.11
6. Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. –3.с
7. Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – 48.с
8. Кутузов А.Б. Корпусная лингвистика. − (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) Национальный корпус русского языка // Вопросы языкознания. – Москва: Наука, 2006. – С. 149-155.
9. Недошивина Е.В. Программа для работы с корпусами текстов: обзор основных корпусных менеджеров. Работа с системой DDC. // Языковая инженерия: в поиске смыслов/ (электронный ресурс).
10. Шемякин Ю.И. Начала компьютерной лингвистики: Учеб. пособие. − М.: Изд-во МГОУ, А/О "Росвузнаука", 1992.

Internet sites

1. https://languages.oup.com/google-dictionary-en

Volume: 4 Issue: 3 | Mar–2025 ISSN: 2720-6882
http://journals.academiczone.net/index.php/jiphr

2. http://uzschoolcorpara.uz/
3. http://wikipedia. org/wiki/ Корпусная лингвистика. Статья в интеренет-энциклопедии Википедия.