

VOL.4 APR.
2025



The Lingua Spectrum

Cogito, ergo sum

ISSN: 3060-4958

"A special kind of beauty
exists which is born in
language, of language,
and for language."

**GASTON
BACHELARD**



MORE INFO:



www.lingvospektr.uz



t.me/thelinguaspectrum

Contextual disambiguation of polysemous words in English and Russian: a comparative corpus-based study

Karimbayeva Asal Shukhratovna
Trainee-teacher
Uzbekistan State World Languages University

Annotation: This article explores the mechanisms of contextual disambiguation of polysemous words in English and Russian through a comparative corpus-based approach. Using data from the Corpus of Contemporary American English (COCA) and the Russian National Corpus (RNC), the study analyzes the role of syntactic and lexical environment in determining word meaning. It identifies patterns that aid in meaning differentiation and discusses cross-linguistic similarities and differences. The findings contribute to better understanding of polysemy in natural language processing and bilingual lexicography.

Keywords: polysemy, context, disambiguation, corpus linguistics, English, Russian

Контекстуальная дизамбигуация полисемантических слов в английском и русском языках: сравнительное корпусное исследование

Каримбаева Асаль Шухратовна
Стажер-преподаватель
Узбекский Государственный Университет Мировых Языков

Аннотация: Данная статья посвящена исследованию механизмов контекстуальной дизамбигуации полисемичных слов в английском и русском языках на основе корпусного анализа. Используя материалы из корпуса современного американского английского (COCA) и Национального корпуса русского языка (НКРЯ), автор анализирует влияние синтаксической и лексической среды на выбор значений многозначных слов. Выявлены закономерности разрешения амбигуитета, а также сходства и различия в межъязыковом плане. Результаты могут быть полезны в области машинной обработки текста и билингвальной лексикографии.

Ключевые слова: полисемия, контекст, дизамбигуация, корпусная лингвистика, английский язык, русский язык

Ingliz va rus tillaridagi ko‘p ma’noli so‘zlarning kontekstual aniqlanishi: taqqoslovii korpusga asoslangan tadqiqot

Karimbayeva Asal Shuxrat qizi
Stajyor-o‘qituvchi
O‘zbekiston Davlat Jahon Tillari Universiteti

Annotatsiya: Ushbu maqolada ingliz va rus tillarida polisemyaga ega so‘zlarning kontekstual anglanishini taqqoslovii korpus asosida tahlil qilish ko‘zda tutilgan. COCA (Amerika ingliz tilining zamonaviy korpusi) va Rossiya milliy korpusidagi ma’lumotlar asosida sintaktik va leksik muhitning so‘z ma’nosini aniqlashdagi o‘rni tahlil qilinadi. Tadqiqot natijalari ma’no ajratishdagi umumiy qonuniyatlarni hamda tillararo o‘xshash va farqlarni ko‘rsatadi. Natijalar matnni avtomatik qayta ishlash va ikki tilli lug‘at tuzishda foydali bo‘lishi mumkin.

Kalit so‘zlar: polisemiya, kontekst, dizambiguatsiya, korpus lingvistikasi, ingliz tili, rus tili.

Introduction

Polysemy – the phenomenon where a single lexical item carries multiple related meanings – is one of the most complex and intriguing features of natural language. It allows languages to be efficient, enabling speakers to use a limited number of words to express a wide range of ideas. However, this economy of expression comes at the cost of increased interpretive complexity, particularly in tasks such as language comprehension, translation, language teaching, and natural language processing (NLP). Determining the intended meaning of a polysemous word requires careful attention to context, making disambiguation a critical challenge in both theoretical and applied linguistics. In both English and Russian, polysemy is not only widespread but also highly productive, meaning that new meanings often emerge through metaphor, metonymy, collocation, or grammatical variation. Learners, translators, and computational systems frequently struggle with such words, especially when dictionary definitions are ambiguous or when the same word takes on different roles depending on genre or syntactic structure. Despite the longstanding interest in polysemy in fields such as semantics, cognitive linguistics, and lexicography (e.g., Cruse, 2004; Geeraerts, 2010), the majority of existing studies focus on monolingual analysis. While such work is valuable, it often overlooks how different languages may employ distinct mechanisms for resolving ambiguity. This is particularly important in bilingual and cross-linguistic contexts, where polysemous words do not always align neatly across languages.

This article aims to address this gap by conducting a comparative corpus-based study of contextual disambiguation in English and Russian. The study is grounded in the belief that context – encompassing both the lexical and syntactic environment – plays a crucial role in determining meaning. By analyzing real usage data from two large and balanced corpora (COCA and the Russian National Corpus), the study seeks to uncover how contextual cues help resolve semantic ambiguity and whether these cues operate similarly or differently across typologically distinct languages. The main research goal is to examine how contextual factors disambiguate polysemous lexemes in both languages and to explore the extent to which these mechanisms are universal or language-specific. More specifically, the study sets out to: (1) analyze contextual cues such as collocations and syntactic patterns that guide the interpretation of polysemous words; (2) compare the semantic behavior of selected high-frequency polysemous lexemes in English and Russian corpora; (3) identify syntactic and lexical strategies used to differentiate meanings; and (4) discuss the implications of these findings for translation, second language acquisition, and computational linguistics. We hypothesize that while the broad contextual cues used for disambiguation – such as collocation, syntax, and genre – are similar across languages, their specific linguistic manifestations and frequencies differ due to typological, grammatical, and cultural factors. For example, Russian's rich morphological system may provide different cues than English's relatively analytic structure. By situating this study within a comparative framework and grounding it in empirical corpus data, we aim to contribute to a more nuanced understanding of polysemy and its contextual resolution. The findings are expected to offer practical value for educators, translators, lexicographers, and developers of NLP systems who work with English and Russian.

Methods

This study adopts a comparative corpus-based methodology to explore how context helps disambiguate the meanings of polysemous words in English and Russian. The choice of a corpus-linguistic approach is motivated by its ability to provide authentic, real-world examples of language use, allowing for empirical observation of disambiguation strategies in naturally occurring discourse. Two large, balanced corpora were used as primary data sources. The Corpus of Contemporary American English (COCA) contains over one billion words across genres such as fiction, newspapers, magazines, academic writing, and spoken language, covering texts from 1990 to 2020 (Davies, 2008–2020). It provides a rich database for observing the contextual behavior of English polysemous words. The Russian National Corpus (RNC), on the other hand, offers a balanced representation of the Russian language, containing more than 300 million words drawn from both written and spoken

registers. It includes diverse genres such as literary prose, news media, legal documents, and everyday conversations. The selection of lexical items was based on several stringent criteria. First, only high-frequency polysemous lexemes were considered to ensure that there would be sufficient data across contexts and registers. Second, each selected word had to demonstrate at least three clearly distinguishable senses, as documented in established dictionaries and prior linguistic studies. Third, the lexemes had to appear in both COCA and RNC in sufficient frequency to allow meaningful cross-linguistic comparison. The English words selected for analysis were *run*, *charge*, *set*, and *light*, all of which are among the most polysemous in the language. For Russian, the words *идти* (to go), *ключ* (key), *провод* (wire/conductor), and *нос* (nose) were chosen based on their semantic diversity and frequency. These words cover a range of domains, including physical motion, technology, metaphor, and abstract meaning, providing a robust sample for comparison. For each lexeme, between 50 and 100 concordance lines were extracted using random sampling from both corpora. Each instance was then subjected to manual annotation. The following variables were coded: (1) the intended sense of the word, based on dictionary definitions and contextual interpretation; (2) the collocational environment, including adjacent nouns, verbs, and modifiers; (3) the syntactic function of the word (e.g., subject, object, modifier); and (4) the genre and register in which the word occurred. The qualitative data analysis involved identifying patterns of semantic differentiation, focusing on how the meaning of a polysemous word is guided by its lexical and grammatical context. Where appropriate, quantitative measures such as frequency counts and co-occurrence patterns were used to support qualitative observations. The analysis also sought to highlight contrasts and parallels in how English and Russian speakers rely on context to resolve ambiguity. This mixed-methods approach – blending qualitative insight with empirical, quantitative backing – aims to provide a well-rounded picture of contextual disambiguation and to draw linguistically meaningful comparisons between two typologically different languages.

Results

The analysis yielded several significant findings that reveal both commonalities and contrasts in how context facilitates disambiguation of polysemous words in English and Russian.

First, lexical collocations emerged as the most consistent and powerful cues for determining word meaning in both languages. In English, for instance, the word *charge* was shown to activate different senses depending on the collocates: *charge a battery* denotes a technical process; *charge a fee* refers to a financial transaction; *lead a charge* has a military connotation; and *face a charge* refers to legal proceedings. Russian exhibits a similar mechanism. The noun *ключ* (key), for example, shifts its meaning across contexts: *ключ к успеху* (“key to success”) is metaphorical, *водный ключ* refers to a natural spring, and *ключ от квартиры* clearly denotes a physical object.

Second, syntactic environment and grammatical markers played a substantial role in sense differentiation. In English, the transitivity and grammatical role of the verb *run* influence its interpretation: *run a business* (transitive, abstract sense) contrasts with *run quickly* (intransitive, physical motion). Russian verbs, particularly *идти* (to go), showed even greater sensitivity to grammatical and morphological features. For example, *идёт дождь* (impersonal subject) refers to weather; *идти по улице* involves literal movement; and *время идёт* implies the abstract passage of time. Aspectual variation (perfective vs. imperfective) further contributed to semantic nuance in Russian, a phenomenon largely absent in English.

Third, genre and register had a marked influence on word meaning. Technical and professional genres favored specialized or abstract senses. In COCA, the word *set* in academic texts often referred to mathematical or logical groupings, while in spoken English it was more likely to appear in idiomatic or phrasal contexts (e.g., *set up*, *set off*). In RNC, metaphorical uses of *ключ* were particularly prevalent in political speeches, psychological literature, and social commentary. This genre sensitivity underscores the need for context-aware interpretation, particularly in tasks such as translation or corpus-based teaching material design.

Fourth, the study observed frequency imbalances and asymmetries across languages, particularly regarding the semantic richness of certain lexemes. The English word set has over 400 documented meanings in the Oxford English Dictionary, while its closest Russian equivalent *набор* does not exhibit the same degree of polysemy. Similarly, many Russian verbs derive semantic variation from grammatical aspect or prefixation, mechanisms that lack a direct parallel in English. This suggests that while both languages rely on context, the means by which meaning is encoded and retrieved differ substantially.

In addition to these findings, the study also noted cross-linguistic gaps in lexical alignment. Some senses of a word in one language have no direct equivalent in the other, creating challenges for translation and bilingual lexicography. These gaps highlight the necessity for flexible, context-sensitive approaches when handling polysemous terms in multilingual applications.

Discussion

The comparative analysis between English and Russian reveals both universal and language-specific mechanisms employed for contextual disambiguation of polysemous words. Universally, lexical collocations and syntactic structures function as primary disambiguators, reinforcing cognitive-linguistic theories which argue that meaning is shaped through usage patterns, mental schemas, and contextual associations (Langacker, 2008). These mechanisms operate in both languages, reflecting shared cognitive processes in language understanding. However, the study also highlights significant language-specific strategies that influence the interpretation of polysemous lexemes. Russian, for example, makes extensive use of grammatical tools such as aspect, case inflections, and morphologically rich verb forms. These grammatical elements often provide additional cues that help to distinguish between meanings, especially when the lexical context is ambiguous. In contrast, English relies more heavily on prepositions, phrasal verbs, and the flexibility of syntactic positioning to express meaning variation. Another noteworthy difference lies in the structural diversity of collocational patterns. English exhibits a wider variety of collocations due to its analytic nature and high frequency of idiomatic expressions. Russian, while also rich in collocations, tends to favor metaphorical extensions and aspectual alternations that are grammatically encoded. For example, the Russian verb *идти* (to go) demonstrates nuanced shifts in meaning based on aspectual pairs and syntactic roles, which are less common in English. Moreover, the analysis underscores the importance of genre in shaping word meaning. Different text genres activate different semantic frames: academic and journalistic writing, for example, favor technical and formal senses, while fiction and conversational texts tend to evoke more figurative or colloquial meanings. These genre-based preferences not only affect frequency of use but also influence the interpretive strategies readers employ. Importantly, the findings reveal limitations in traditional lexicographic approaches that present isolated dictionary definitions without sufficient contextualization. Dictionaries often fail to capture the subtle interaction between collocation, syntax, and genre, leading to potential misinterpretation in language learning or translation contexts. This observation supports a growing pedagogical trend that advocates for corpus-informed language instruction and lexicography. In applied contexts such as machine translation, natural language processing (NLP), and bilingual education, these findings are particularly relevant. Disambiguating polysemous words accurately requires models and materials that integrate contextual features rather than rely solely on isolated word senses. Future technologies and teaching tools must therefore incorporate contextual intelligence – such as semantic roles, syntactic cues, and genre indicators – to improve accuracy and effectiveness.

Conclusion

This study has investigated the contextual disambiguation of polysemous words in English and Russian through a comparative corpus-based methodology. The findings confirm that contextual information – especially collocational patterns and syntactic environment – is essential for interpreting polysemous words accurately. While both languages share fundamental disambiguation strategies, each exhibits unique linguistic features that shape how meaning is constructed and

understood. The research underscores the value of corpus-based methods for studying polysemy, offering empirical insights that go beyond theoretical semantics or dictionary-based analysis. It also highlights the pedagogical benefits of integrating authentic, context-rich examples into second language instruction. Teachers and learners can benefit from exposure to real corpus data that illustrates how polysemous words function across genres and syntactic frames. Moreover, the study has practical implications for the fields of bilingual lexicography, automatic translation, and computational linguistics. Systems designed to process natural language must be sensitive not only to lexical meaning but also to contextual variation across languages. Incorporating grammatical and genre-specific information can significantly enhance the accuracy of language technologies and the quality of cross-linguistic communication. Future research could expand the scope of this analysis by including additional corpora – particularly spoken language data – and by applying computational modeling to detect subtle disambiguation patterns. Furthermore, comparing English and Russian with typologically distant languages such as Chinese, Turkish, or Finnish could provide deeper insight into whether disambiguation strategies are truly universal or heavily language-dependent. Ultimately, understanding how polysemous words are interpreted in context enhances our comprehension of language complexity, cognitive processing, and intercultural communication. This knowledge contributes not only to theoretical advancements but also to more effective applications in education, translation, and language technology.

References

1. Cruse, D. A. (2004). *Meaning in language: An introduction to semantics and pragmatics* (2nd ed.). Oxford University Press.
2. Davies, M. (2008–2020). *The Corpus of Contemporary American English (COCA)*. <https://www.english-corpora.org/coca/>
3. Firth, J. R. (1957). *Papers in linguistics 1934–1951*. Oxford University Press.
4. Geeraerts, D. (2010). *Theories of lexical semantics*. Oxford University Press.
5. Janda, L. A., & Ljashevskaya, O. N. (2011). Aspectual clusters of Russian verbs. In R. D. Binnick (Ed.), *The Oxford handbook of tense and aspect* (pp. 598–624). Oxford University Press.
6. Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford University Press.
7. Oxford University Press. (2023). *Oxford English Dictionary*. <https://www.oed.com>
8. Russian National Corpus. (n.d.). *Российский национальный корпус*. <https://www.ruscorpora.ru>

Iskandarova N.I. Caryl Churchill: a revolutionary force in modern theater.....	1118
Abdiyeva M.A. 1917-1941- yillarda O‘zbekistonda an‘anaviy hunarmandchilik va sun‘iy sanoatlashtirishning ijtimoiy oqibatlari tarixshunosligi.....	1124
Jurayev A.I. Mustaqillik yillarida Jizzax viloyatida sog‘liqni saqlash borasidagi muhim islohotlar samarasi (Onalik va bolalikni muhofaza qilish misolida).....	1130
Nizamatdinov Ye.K. Badruddin Muhammad Kardariy yashagan davrda siyosiy holat	1134
Otaxanov A. Temuriylar sulolasi hukumronligi davrida tibbiyot.....	1138
Rasulov G.P. O‘zbekistonda madaniy merosni muhofaza qilishning huquqiy va tashkiliy asoslari	1143
Raxmatova D.J. O‘zbekiston SSR SSSRdagi asosiy paxta ishlab chiqarish Respublikasi.....	1147
Shodmonov X.X. O‘zbekiston muzeylarining ta‘limiy faoliyatining rivojlanishi.....	1150
Xudoyberganova G. Markaziy Osiyoda zohidlik maktablari shakllanish jarayonlari.....	1154
G‘ulomova R. Qarg‘ish so‘zlarning ijtimoiy-madaniy kelib chiqish tarixi.....	1159
Сейтумбетова А.П. Фразеологические единства в английском, русском и каракалпакском языках.....	1165
Сейтумбетова А.П. Фразеологические единицы как предмет исследования лингвокультурологии	1171
Karimbaeva A.Sh. Professional development of EFL teachers in Uzbekistan: from traditional methods to global competency standards.....	1176
Karimbaeva A.Sh. Contextual disambiguation of polysemous words in English and Russian: a comparative corpus-based study	1180
Далиева М.Х. Засухоустойчивые растения с доказанными фильтрующими свойствами по очистке воздуха Ташкента от PM2.5.....	1185
Satibaldiev E.K. Urban Biodiversity Corridors as a Climate-Adaptation Strategy in Contemporary Uzbekistan.....	1191
Satibaldieva N.A. Landscape-Scale Wetland Restoration to Mitigate Dust Storms in Contemporary Uzbekistan.....	1198