

RU

Контекстуальные признаки лингвистических многозначных терминов для их идентификации в китайском, английском и русском языках

Хуан Шанлун

Аннотация. Цель исследования – разработать эффективные методы построения корпуса, предназначенного для идентификации лингвистических значений терминов в китайском, английском и русском языках на основе закономерностей образования терминосочетаний. В статье для изучения способов образования лингвистических многозначных терминов, значимых для выявления их контекстуальных особенностей, были проанализированы словарные описания таких терминов, полученные методом сплошной выборки. На материале отобранных корпусов китайского, английского и русского языков были исследованы узкий и широкий контексты терминов, указывающие на их лингвистические значения. Научная новизна исследования заключается в получении принципиально новой методики создания корпуса, предназначенного для распознавания лингвистических терминов в разных языках, аналогичные исследования ранее не проводились. Полученные результаты показали, что в лингвистике выделяются различные модели образования терминов в трех языках, рассмотренных в данном исследовании, но наиболее типичной структурой является Adj. + N. В результате анализа контекстуальных особенностей были выявлены два варианта: для одного класса лингвистических многозначных терминов анализ их устойчивых сочетаний может способствовать прямому распознаванию значений терминов, для другого класса таких терминов при идентификации их значений следует использовать, наряду с устойчивыми сочетаниями, также широкий контекст их употребления.

EN

Contextual features of linguistic polysemous terms for their identification in Chinese, English and Russian

Shanglong Huang

Abstract. The main aim was to develop effective methods for building a corpus designed to identify the linguistic meanings of terms in Chinese, English and Russian based on the patterns of term phrases formation. In the research, we analyzed the dictionary descriptions of linguistic polysemous terms by using a continuous sampling method to study the ways of forming such terms that are significant for identifying their contextual features. The narrow and wide contexts of terms indicating their linguistic meanings were studied on the basis of the selected corpus of the Chinese, English and Russian languages. The novelty of the study lies in obtaining a fundamentally new methodology for creating a corpus designed to recognize linguistic terms in multiple languages; similar studies have not been conducted before. As the results show, there are various patterns of phrase formation in linguistics across the three languages studied, but the most typical structure for them is “Adj. + N.”. As for contextual features, two situations were observed – the analysis of the fixed phrases of some linguistic polysemous terms can assist in recognizing the meanings of the terms immediately, while some term phrases should be used for identification, combined with the terms’ features in wide context.

Introduction

Polysemy is defined as *the fact of having several meanings or the possession of multiple meanings, senses, or connotations* in the Oxford English Dictionary (<https://www.oed.com/search/dictionary/?scope=Entries&q=polysemy>). This phenomenon exists in almost every language. Certainly, a word with multiple meanings can create challenges in understanding and identification, regardless of the language.

Polysemous words in different languages are likely to have similarities or differences, because languages are organized with similar or different logics and structures. Chinese is ideographic, and its writing system uses Chinese characters, but English and Russian are epigraphic, and letters are used in their written situations. The uniqueness of using Chinese characters in the ideographic language is predicted to cause Chinese polysemous terms to differ

from the other two languages. Russian as an inflectional language has a rich variety of word endings, which can also differentiate it from the other two languages in terms composition.

Compared to homonyms that are words having different meanings with the same pronunciation, polysemous words possess senses that are interconnected. Thus, in dictionaries they are typically listed under a single entry with their various meanings grouped together, reflecting their semantic connections, while homonyms may have two or more lexical entries. Polysemous terms are more difficult to identify than homonymous terms, because the closer term meanings are to each other, the more challenging it becomes for us to distinguish between them (Huang, 2023).

The relevance of the research is that multilingualism and interdisciplinary methods play an essential role in linguistic study. In the field of terminology, there are more articles about the analysis of linguistic terms across two languages than over two languages, for example, *On the Comparability of Linguistic Terminology in Russian and English* (Инюткина, Большакова, 2021). Multilingualism enriches the materials and perspectives of terminology research and provides references for other research in multiple languages. Besides, the study can be seen as interdisciplinary research, encompassing the theories of linguistics, corpus linguistics and, potentially, informatics. Interdisciplinary research in terminology helps to eliminate limitations of traditional linguistic research. The study's findings endorse the automatic processes of thematic classification, content summarization and information retrieval during the analysis of scientific publications.

The article focused on analyzing polysemous terms in linguistics across Chinese, English and Russian. The research tasks were as follows:

- 1) to set up principles for selecting research material and to select suitable terms in the three languages from the dictionaries for the study;
- 2) to examine the regularities in how polysemous terms are formed in these languages by analyzing all the terms in the dictionaries;
- 3) to identify the contextual features within each language that aid in recognizing these polysemous terms which name linguistic concepts.

The research methods included comparison and continuous sampling, which were used for choosing the research material from a big number of terms, grammatical analysis, which was applied to study the patterns of term composition, corpus analysis and description, which were utilized for collecting and demonstrating the contextual markers for identification.

The research material was selected from the following dictionaries:

- Ахманова О. С. Словарь лингвистических терминов. М.: URSS, 2021;
- Crystal D. A Dictionary of Linguistics and Phonetics. Hoboken: John Wiley & Sons, 2011;
- 戴维 克里斯特尔. 现代语言学词典 / 沈家煊, 译. 北京, 2011 (Crystal D. Dictionary of Modern Linguistics //

Shen J. trans. Beijing: The Commercial Press, 2011).

The material of the following corpora was also used:

- Национальный корпус русского языка (Russian National Corpus). <https://ruscorpora.ru>;
- КиберЛенинка (CyberLeninka). <https://cyberleninka.ru>;
- Google Ngram Viewer. <https://books.google.com/ngrams>;
- Corpus of Contemporary American English (COCA). <https://www.english-corpora.org/coca>;
- Wiley Online Library. <https://onlinelibrary.wiley.com>;
- BCC (BLCU Corpus Center). <https://bcc.blcu.edu.cn>;
- 术语在线 (Term Online). <https://www.termonline.cn>;
- 中国知网 (CNKI – China National Knowledge Infrastructure). <https://chn.oversea.cnki.net>.

The theoretical background is reflected by the articles of J. Festman (2021), E. V. Krapivnik and A. O. Nosenko (Крапивник, Носенко, 2021), P. V. Gibkiy and N. V. Suprunchuk (Гибкий, Супрунчук, 2021), S. Kim, L. Liu and F. Cao (2017), K. Gortych-Michalak and J. Grzybek (2013), which were dedicated to the research about the relations or mutual effects between two or more languages. Also, it is seen in the interdisciplinary research of T. Hamon and N. Grabar (2010), Z. Liu, Y. Chen, B. Tang, X. Wang, Q. Chen, H. Li, J. Wang, Q. Deng and S. Zhu (2015), E. Laparra, A. Mascio, S. Velupilla and T. Miller (2021), W. D. Yuniarti, S. Hartati, S. Priyanta and H. D. Surjono (2022), D. Wijayasekara, O. Linda, M. Manic and C. Rieger (2014), M. Pikhart and B. Klímová (2020), contributing to the interaction of linguistics and informatics or electronics, such as medical information extraction, adaptable models for health record based on NLP, data processing for the assessment system, electronic communication management and so on.

The practical value shows that the research materials can be used in universities at courses and seminars on theoretical, practical, comparative terminology, especially useful for Russian native speakers, or for students who are Chinese native speakers and use English and Russian in study. The results obtained can be applied in the process of educational activities, such as writing textbooks and reference books for teaching. Besides, in future research of applied linguistics these results will be beneficial to setting up an identification system for recognizing the linguistic senses of polysemous terms by means of a to-be-built corpus containing numerous contextual markers that indicate the linguistic meanings.

Discussion and results

Selection of terms

As the first step, term selection from the dictionaries mentioned above was completed. When looking through these dictionaries, we observed that linguistic terms could be categorized into specialized terms which merely function

in the field of linguistics, terms shared by multiple disciplines, terms that are the same as the general vocabulary, and others. It is understandable that linguistic polysemous terms fall into the categories of terms used across multiple disciplines and those that are consistent with ordinary vocabulary. In other words, other lexical meanings of the chosen linguistic polysemous terms appear in everyday language or in other fields of study.

For selecting the terms for our study four principles were set up: 1) the terms are also found in everyday vocabulary or other fields of study; 2) the chosen terms point to the same object in all three languages; 3) single terms are selected instead of terminology phrases; and 4) the terms belong to different parts of speech.

As a result, we chose two sets of terms, which were 同化, *assimilation*, *ассимиляция* and 自主, *autonomous*, *автономный*.

Patterns of term phrases formation

An overall analysis of the general features of forming linguistic terminological phrases in the three languages by comparing terms in dictionaries shows clearly whether different languages make great differences in term phrase composition.

For **Chinese** linguistic terminological phrases, one of the most typical structures is compound nouns in which two nouns are combined (N. + N.). In this structure the first noun is a modifier specifying the second noun. For example, 行为|主义 (behaviourism, the first two characters of the Chinese phrase means “behaviour”, the last two can be translated into “principle” or “doctrine”), 工具|格 (instrumental case), 音律|单位 (phonematic unit), 词汇|系统 (lexical system), etc.

The next structure is compound nouns which are formed by an adjective and a noun (Adj. + N.). Obviously, the adjective modifies the noun, giving a specific meaning to the noun phrase. For example, 自主|句法 (autonomous syntax), 基本|形式 (basic form), 间接|宾语 (indirect object), etc.

Then follows the structure with a verb and a noun (V. + N.). The verb used in the structure can be considered as an infinitive, because no word's form changes in Chinese, which is not an inflecting language or agglutinative language. Take *running machine* for example, “to run” is an infinitive, but when we want to modify the word *machine*, we should use the gerund of *run* or the so-called “-ing” form of it. However, in Chinese it is called “跑步|机”, and “跑步” literally means “to run”. The examples of this structure are 指称|意义 (referential meaning), 替换|框架 (substitution frame), 描写|语法 (descriptive grammar), etc.

English linguistic terminological phrases are typically seen in the form of compound nouns with an adjective as a modifier (Adj. + N.). For example, *semantic components*, *cyclic rules*, *double negative*, *regional dialects*, etc. Indeed, in this structure, the noun can often be a gerund, such as *compensatory lengthening* and *horizontal splitting*.

The second common structure is compound nouns consisting of two nouns (N. + N.). For example, *word grammar*, *address system*, *government phonology*, *future tense*, etc. In this structure, the first noun or second can often be a gerund too, for example, *level skipping*, *working hypothesis*, *key lowering*, *bracketing paradox*, etc.

There is another structure including two nouns with the possessive case (N. + of + N.). The grammatical meaning of this structure should be compared to that of the structure N. + N. – the former indicates a possessive relation between the nouns, and the latter implies a noun modifying the other. For example, *manner of discourse*, *context of utterance*, *theory of markedness*, etc.

For **Russian** terminological phrases used in linguistics, the structure Adj. + N. is also typical. The difference lies in the fact that Russian adjectives must agree with nouns in term of genders (masculine, feminine and neutral), numbers (singular and plural) and cases (nominative, genitive, dative, accusative, instrumental and prepositional). Examples are *безличное предложение* (*impersonal sentence*), *дифференциальный модус* (*differential mode*), *множественное число* (*plural*), etc.

Another structure for Russian linguistic terminological phrases is, in appearance, like the structure N. + N., but actually, in grammatical meaning, the same as the structure N. + of + N. That is the structure N₁ + N₂ (N₁ means a noun with the nominative case, while N₂ represents a noun with the genitive case). This is not a simple juxtaposition of two nouns, but a noun in the genitive case, playing a role as a modifier, is placed after the other noun. For examples, *теория моногенеза* (*theory of monogenesis*).

As is discussed above, for each language involved in the research the most common structure of composing linguistic terminological phrases is Adj. + N. In practice, it is exactly the typical type of phrases that we use in colloquial language most of the time, such as *good quality*, *high school*, *red book* and so on. Therefore, it is no surprise that this structure is often found in linguistic terminology phrases.

However, from the point of view of translating terms, a given structure in one language is not always equivalent to the same structure in another language. Take English term *language environment* for example, it is with the structure N. + N. in English. What about Chinese and Russian? It is also N. + N. in Chinese (语言环境), but Adj. + N. in Russian (языковая среда). Such a correspondence can be expressed as “en: N. + N. = chin: N. + N. = russ: Adj. + N.”. Another example can be *syntactic function* / 句法功能 / синтаксическая функция. The relationship between them at this point is expressed as “en: Adj. + N. = chin: N. + N. = russ: Adj. + N.”.

Context in the corpora

As a matter of fact, the ways terminological phrases are constructed reflect the *narrow context* of the terms, which refers to not only phrases, but also sentences (Верещагин, Костомаров, 1980). After searching for the selected terms in the corpora, we got information about their context.

For Chinese polysemous linguistic terms, we used “术语在线” (Term Online. <https://www.termonline.cn>), on which the fields where a given term is used and the term definitions in the corresponding fields are presented, and BCC (<https://bcc.blcu.edu.cn>), which serves to confirm or uncover linguistic phenomena and to facilitate information extraction and automated linguistic analysis (荀恩东, 饶高琦, 肖晓悦 [Xun, Rao, Xiao] et al., 2016). In BCC, we searched about the words with a specific part of speech beside a given term in context, for example, if we needed to collect all the possible adjectives on the left of the word 成分 (*element*), we input “a成分”.

With the help of the corpus “术语在线” (Term Online), we know that the selected terms in Chinese appear in the following fields of study. 同化 (assimilation): Behavioral Medicine, Zoology, Psychology, Linguistics, Chemical Engineering, Physiology and Botany; 自主 (autonomous): Management, Science and Technology.

According to the formation patterns presented above, the noun “同化” may be the first or the second noun in the structure **N. + N.** or the second noun in the structures **Adj. + N.** and **V. + N.** Thus, we input four forms of the term, “n同化”, “同化n”, “a同化” and “v同化”, in BCC to check its narrow context. The top 5 most frequent narrow contexts of each form are presented as follows.

n同化: 产品同化 (assimilation of products), 文化同化 (cultural assimilation), 民族同化 (assimilation of nations), 结构同化 (structural assimilation), 社会同化 (social assimilation); 同化n: 同化现象 (phenomenon of assimilation), 同化过程 (process of assimilation), 同化问题 (problem of assimilation), 同化能力 (ability of assimilation), 同化产品 (assimilated product); a同化: – (there are not enough examples for this form); v同化: 是同化 (to be assimilation), 吸收同化 (absorption and assimilation), 不能同化 (cannot assimilate), 产生同化 (produce assimilation), 无法同化 (to be unable to assimilate).

“自主” is an adjective, so it is related to the structure **Adj. + N.** We input “自主n” in the corpus, and the top 5 most frequent narrow contexts are demonstrated as follows.

自主n: 自主知识产权 (Independent Intellectual Property), 自主意识 (sense of autonomy), 自主能力 (independence ability), 自主技术 (autonomous technology), 自主决策 (self-determination).

To analyze polysemous linguistic terms in English, we took advantage of COCA (<https://www.english-corpora.org/coca>), in which we found all words with specific parts of speech beside a term in context, for example, when looking for all the adjectives on the left of the word *element*, we input “ADJ element”.

Based on the terminological phrase composition patterns, the noun *assimilation* can be in the structures **Adj. + N.**, **N. + N.** and **N. + of + N.** Thus, in COCA we input five forms of the term, “ADJ assimilation”, “NOUN assimilation”, “assimilation NOUN”, “NOUN of assimilation” and “assimilation of NOUN”. The top 5 most frequent contexts of each form are shown below.

ADJ assimilation: cultural, forced, segmented, social, biased; NOUN assimilation: data, carbon, iron, nutrient, technology; assimilation NOUN: process, effect, theory, lamp, policies; NOUN of assimilation: process, processes, policy, forces, model; assimilation of NOUN: immigrants, Indians, nutrients, women, outsiders.

The adjective *autonomous* is a part of the structure **Adj. + N.**, so we input “autonomous NOUN”. The top 5 contexts of it are the following.

autonomous NOUN: region, vehicles, university, vehicle, regions.

When dealing with Russian polysemous terms in linguistics, we got access to the Russian National Corpus (<https://ruscorpora.ru>). Besides, we referred to Google Ngram Viewer (<https://books.google.com/ngrams>) for the context of Russian linguistic polysemous terms as well.

The composition patterns suggest that the noun *ассимиляция* may be included in the structures **Adj. + N** and **N₁ + N₂**. Assisted by the function “Word at glance” (in Russ. “портрет слова” (word portrait)) in the Russian National Corpus, we search for the modifier of the term. For the contexts of the structure **N₁ + N₂**, we use the Russian corpus from 2010 to 2022 of Google Ngram Viewer by inputting two forms, respectively *ассимиляция* as the first noun of the phrase and as the second noun with genitive case, thus, that is “ассимиляция=>*_NOUN” and “*_NOUN=>ассимиляции”. The top 5 results of each of them are listed below.

Modifier: языковой (language), насильственный (forced), всеобщий (universal), постепенный (gradual), культурный (cultural); ассимиляция=>*_NOUN: диссимилиация (dissimilation), аккомодация (accommodation), интеграция (integration), народов (of nations), населения (of population); *_NOUN=>ассимиляции: процесс (process), процессы (processes), результат (result), политика (policy), интеграция (integration).

The adjective *автономный* only accords with the structure **Adj. + N.**, so we search for the nouns that it modifies using “Word at glance” (word portrait) in the Russian National Corpus. The top 5 nouns modified by the term are as follows.

Nouns defined by the adjective: окпуг (area), республика (republic), область (region), плавание (voyage), робот (robot).

As the results of contexts show, almost all the term phrases in the three languages are hardly conceptually connected to linguistics except the Russian phrase *языковая ассимиляция* (*language assimilation*), which could share a semantic link with linguistics. For the term *assimilation*, it is frequent to see the phrases *cultural assimilation*, *social assimilation*, *assimilation of nations* and *process of assimilation* in all three languages. For the term *autonomous*, we can frequently observe *autonomous region* in context in English and Russian.

Therefore, in order to identify the meaning of the terms right in linguistics, we need, exactly, the features of the contexts that are discovered in the field of linguistics. For there are few contexts of linguistics in general corpora, in other words, the contexts in linguistics are not common for the selected terms, it is required to study their contexts in the corpora within linguistics specifically to figure out the contextual features for their identification.

Contextual features for identifying linguistics meanings

Apart from the features of the terms as components in the phrases (narrow context), it is advantageous to analyze the features of the broader contexts within paragraphs, which is called *wide context*, a linguistic unit beyond a sentence, such as a paragraph or a chapter (Вережанин, Костомаров, 1980). We referred to not only the corpora mentioned earlier but also the dictionaries listed above to study polysemous terms' features presented in context.

In *A Dictionary of Linguistics and Phonetics* (Crystal, 2011), the linguistic term 同化 (assimilation) is narrowly defined as the process by which one segment exerts an influence on another, so that the two sounds become similar or identical. In Chinese, it can be broadly defined as a translation strategy that employs a conservative means of the original text to make it close to the target language and cater to local needs (Venuti, 1995). It is notable that in English this translation method is called "domestication" instead of "assimilation". So, this term in Chinese is polysemous even in the field of linguistics. There is a need to discuss the term with different meanings separately.

"同化" as assimilation is, precisely, a phonetic term, so it is related to syllable, letter, word, pronunciation, language, etc. Its contextual features are presented in Table 1.

Table 1. The contextual features of the term "同化" (assimilation)

No.	Narrow context	Wide context
1	语音同化 (phonetic assimilation), 邻接同化 (contiguous assimilation), 非邻接同化 (non-contiguous assimilation), 逆同化 (regressive assimilation), 顺同化 (progressive assimilation), 聚结同化 (coalescent assimilation), 相互同化 (reciprocal assimilation), 单一特征同化 (single-feature assimilation)	—
2	全部/完全同化 (total/complete assimilation), 部分/不完全同化 (partial/incomplete assimilation)	音节 (syllable), 字母 (letter), 词 (word)
3	同化过程 (process of assimilation), 同化现象 (phenomenon of assimilation)	音 (sound), 发音 (pronunciation), 语言 (language), X语 (a certain language), etc.

The linguistic term phrases in line 1 of Table 1 are specialized terms that are applied only in linguistics, so its narrow context is enough to identify its linguistics meanings in context. The phrases in line 2 are also linguistics terms, but in Chinese their meaning may be blurry, so we need the features in wide context to distinguish them. Those in line 3 are some phrases which can be used in linguistics context, thus, if we find any word of the wide context in the process of identification, we can confirm that these phrases have the meanings in linguistics (the wide contexts were collected from "中国知网" (CNKI. <https://chn.oversea.cnki.net>)).

"同化" as domestication is a term used in the field of translation, so it is related to language, translation, strategy, foreignization and so on. Its contextual features are demonstrated in Table 2.

Table 2. The contextual features of the term "同化" (domestication)

No.	Narrow context	Wide context
1	同化策略 (strategy of domestication), 同化理论 (theory of domestication), 同化和异化 (domestication and foreignization)	翻译 (translation), 策略 (strategy), 异化 (foreignization), etc.

In fact, whenever the word "translation" appears in the wide context, it is essentially concluded that the term has the linguistic definition, because the meaning of this Chinese term is almost unique in the context of translation studies.

With the term "自主" (autonomous) we can compose some fixed phrases, which are the basis for its identification. In Table 3, the linguistic term's fixed phrases are collected, which help to identify it directly, so there is no need for wide contexts.

Table 3. The contextual features of the term "自主" (autonomous)

No.	Narrow context
1	自主音位 (autonomous phoneme), 自主句法 (autonomous syntax), 自主语 (autonomous speech), 自主音段 (autosegment)

In English, *assimilation* is only a phonetic term, plus, it has the same meaning as the Chinese term "同化" when the latter means assimilation. So, it possesses the context details in Table 4 (the wide contexts were received from Wiley Online Library (<https://onlinelibrary.wiley.com>)).

Table 4. The contextual features of the term "assimilation"

No.	Narrow context	Wide context
1	phonetic assimilation, contiguous assimilation, non-contiguous assimilation, regressive assimilation, progressive assimilation, coalescent assimilation, reciprocal assimilation, single-feature assimilation	—
2	total/complete assimilation, partial/incomplete assimilation	syllable, letter, word, pronunciation, language, a certain language (such as English and Chinese), etc.
3	process of assimilation, phenomenon of assimilation	

The narrow contexts of the term *autonomous* are effective enough for recognizing its meaning of linguistics. Look at its contextual details in Table 5.

Table 5. The contextual features of the term “autonomous”

No.	Narrow context
1	autonomous phoneme , autonomous syntax , autonomous speech

The Russian term *ассимиляция* has two distinct meanings in linguistics according to *A Dictionary of Linguistic Terms* of O. S. Akhmanova (Ахманова, 2021). One is the adaptation of one sound to another, consisting of the adjustment of the recursion of the preceding sound to the excursion of the following one, the other one is the change in the form of number, gender, case, etc. under the influence of context. We obtained four groups of features presented in Table 6 (the phrase in line 4 is related to the term’s second definition).

Table 6. The contextual features of the term “ассимиляция” (assimilation)

No.	Narrow context	Wide context
1	взаимная ассимиляция (reciprocal assimilation), графическая ассимиляция (graphic assimilation), дистактная/гармоническая ассимиляция (incontiguous assimilation), контактная ассимиляция (juxtapositional/contact assimilation), прогрессивная ассимиляция (progressive assimilation), регрессивная ассимиляция (regressive assimilation), эпентетическая ассимиляция (epenthetic assimilation)	–
2	полная ассимиляция (total/complete assimilation), частичная/неполная ассимиляция (partial/incomplete assimilation)	слог (syllable), буква (letter), слово (word), произношение (pronunciation), язык (language), определенный язык (как английский, китайский и др.) (a certain language (such as English and Chinese), etc.)
3	процесс ассимиляции (process of assimilation), явление/феномен ассимиляции (phenomenon of assimilation)	
4	морфологическая ассимиляция (morphological analogy)	

Likewise, line 1 and line 4 of Table 6 are fixed phrases of the term, which can be applied to identify the term’s meaning in linguistics without any wide context. But the phrases in line 2 and line 3 are not concrete, that is why the wide contexts are necessary to assist the identification (the wide context were obtained from “CyberLeninka” (<https://cyberleninka.ru>)).

The term *автономный* (autonomous) can be used to form fixed phrases as well, as in Chinese and English. In this case, its narrow context is serviceable enough for the identification of the linguistic meaning. Look at the contextual features in Table 7.

Table 7. The contextual features of the term “автономный” (autonomous)

No.	Narrow context
1	автономные слова (autonomous words), автономный грамматический процесс (autonomous grammatical process)

Conclusion

The research findings of the article are presented as follows:

1. The principles for selecting the terms used in the study are essential, because different parts of speech and meanings of words will have different effects on the research process. Two groups of linguistic polysemous terms, 同化, assimilation, ассимиляция and 自主, autonomous, автономный, which respectively are a noun and an adjective, coincided with the principles and contributed to successfully carrying out the study.

2. Various properties of different languages affect the composition of their terminology phrases. In Chinese, the common structures of linguistic terminology phrases are **N. + N.**, **Adj. + N.** and **V. + N.**; in English, they are **Adj. + N.**, **N. + N.** and **N. + of + N.**; in Russian, the typical patterns are **Adj. + N.** and **N₁ + N₂**. For all three languages, the frequent composition pattern is the commonly used structure **Adj. + N.**

3. The meaning of a specific polysemous term in each language can be discerned through the context in which it appears (it was also proved by a research paper in which the authors claim that algorithms based on context can help to match polysemous terms with their interpretation (Tsinaraki, Velegakis, Kiyavitskaya et al., 2010)). When it comes to setting up corpora of contextual features of linguistic terms, in some languages a certain term is even polysemous in linguistics contexts, so its contextual features for identification should be collected separately to build the corpus. Furthermore, when terms are used in idiomatic phrases, these phrases serve as indicators of the terms’ linguistic meanings and can be directly utilized for establishing the corpus. However, when the phrases formed by the terms have ambiguous meanings, it is essential to combine the terms’ features within both wide and narrow contexts during the setting up of the corpus.

For further research perspectives, the study will be closely related to the interdisciplinary research, as literature reviews on the topic of information extraction (Allahyari, Pouriyeh, Assefi et al., 2017; Wang, Wang, Rastegar-Mojarad et al., 2018) proved that the combination of linguistics and informatics is a popular subject of study nowadays. The research findings of the article can be developed into establishing a complete corpus of contextual markers and creating a system for identifying linguistic terms derived from the corpus including valid contextual features.

Источники | References

1. Верещагин Е. М., Костомаров В. Г. Лингвострановедческая теория слова. М.: Русский язык, 1980.
2. Гибкий П. В., Супрунчук Н. В. Сохранение эквивалентности при передаче значения категории пассивного залога с русского языка на китайский (на материале сайтов минского тракторного завода и индустриального парка «Великий камень») // Вестник Омского государственного педагогического университета. Гуманитарные исследования. 2021. № 4 (33).
3. Инютина Л. А., Большакова И. Д. О сопоставимости лингвистической терминологии в русском и английском языке // Успехи гуманитарных наук. 2021. № 9.
4. Крапивник Е. В., Носенко А. О. Национально специфичные особенности метафоризации как способа терминообразования (на материале лингвистической терминологии русского и английского языков) // Гуманитарные исследования. 2021. № 1 (77).
5. Allahyari M., Pouriyeh S., Assefi M., Safaei S., Trippe E., Gutierrez J., Kochut K. A brief survey of text mining: classification, clustering and extraction techniques. <https://doi.org/10.48550/arXiv.1707.02919>.
6. Festman J. Learning and processing multiple languages: The more the easier? // Language Learning. 2021. Vol. 71 (1).
7. Gortych-Michalak K., Grzybek J. Polysemic Terms in Chinese, German, Greek and Polish Legal Language: A Comparative Study // Comparative Legal Linguistics: International Journal for Legal Communication. 2013. Vol. 15.
8. Hamon T., Grabar N. Linguistic approach for identification of medication names and related information in clinical narratives // Journal of the American Medical Informatics Association. 2010. Vol. 17 (5).
9. Huang S. Application of Informatics and Electronics on a Linguistic Level for Handling Consubstantial Terms in Automatic Processing of Scientific Publications // IEEE 24th International Conference of Young Professionals in Electron Devices and Materials (EDM). 2023. <https://doi.org/10.1109/EDM58354.2023.10225126>
10. Kim S., Liu L., Cao F. How does first language (L1) influence second language (L2) reading in the brain? Evidence from Korean-English and Chinese-English bilinguals // Brain and Language. 2017. Vol. 171.
11. Laparra E., Mascio A., Velupilla S., Miller T. A review of recent work in transfer learning and domain adaptation for natural language processing of electronic health records // Yearbook of Medical Informatics. 2021. Vol. 30 (1).
12. Liu Z., Chen Y., Tang B., Wang X., Chen Q., Li H., Wang J., Deng Q., Zhu S. Automatic de-identification of electronic medical records using token-level and character-level conditional random fields // Journal of Biomedical Informatics. 2015. Vol. 58.
13. Pikhart M., Klímová B. Managerial computer communication: implementation of applied linguistics approaches in managing electronic communication // Advances in Computer, Communication and Computational Sciences: Proceedings of IC4S 2019. Berlin: Springer, 2020. Vol. 1158.
14. Tsinaraki C., Velegrakis Y., Kiyavitskaya N., Mylopoulos J. A Context-Based Model for the Interpretation of Polysemous Terms // On the Move to Meaningful Internet Systems (OTM 2010). Berlin – Heidelberg: Springer, 2010. Vol. 6427. https://doi.org/10.1007/978-3-642-16949-6_20
15. Venuti L. The Translator's Invisibility: A History of Translation. L. – N. Y.: Routledge, 1995.
16. Wang Y., Wang L., Rastegar-Mojarad M., Moon S., Shen F., Afzal N., Liu S., Zeng Y., Mehrabi S., Sohn S., Liu H. Clinical information extraction applications: A literature review // Journal of Biomedical Informatics. 2018. Vol. 77.
17. Wijayasekara D., Linda O., Manic M., Rieger C. Mining building energy management system data using fuzzy anomaly detection and linguistic descriptions // IEEE Transactions on Industrial Informatics. 2014. Vol. 10 (3).
18. Yuniarti W. D., Hartati S., Priyanta S., Surjono H. D. Utilization of linguistic data for learner assessment on e-learning: instrument and processing // Seventh International Conference on Informatics and Computing (ICIC). 2022. <https://doi.org/10.1109/icic56845.2022.10006977>
19. 荀恩东, 饶高琦, 肖晓悦, 臧娇娇. 大数据背景下BCC语料库的研制 // 语料库语言学. 2016. № 1 (Xun E., Rao G., Xiao X., Zang J. The construction of the BCC Corpus in the age of Big Data // Corpus Linguistics. 2016. Vol. 3 (1)).

Информация об авторах | Author information

Хуан Шанлун¹¹ Новосибирский государственный университетShanglong Huang¹¹ Novosibirsk State University¹ s.khuan2@alumni.nsu.ru

Информация о статье | About this article

Дата поступления рукописи (received): 14.11.2024; опубликовано online (published online): 02.06.2025.

Ключевые слова (keywords): лингвистические многозначные термины; сравнение языков; закономерности образования терминов; контекстуальные признаки; идентификация терминов; linguistic polysemous terms; comparison of languages; patterns of term formation; contextual features; identification of terms.