

УДК 811.161.1'373:004(045)

*О.Н. Польщикова***СИНТАКСИЧЕСКИЙ СПОСОБ ОБРАЗОВАНИЯ ТЕРМИНОВ  
КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ**

Статья посвящена анализу особенностей синтаксического способа образования терминов компьютерной лингвистики. Обоснована актуальность изучения составных терминов профессиональной сферы компьютерной лингвистики. В исследуемой терминологии наблюдаются модели разнокомпонентных словосочетаний, основными из которых являются двухкомпонентные терминологические единицы с именем существительным в качестве главного слова. Среди выявленных словосочетаний имеются терминологические единицы с предлогами и без предлогов, функционируют достаточно сложные терминологические словосочетания с моделями согласования и управления. В составе двух-, трех-, четырех- и многокомпонентных терминов встречаются модели различных сочетаний частей речи. Компонентами терминологических словосочетаний компьютерной лингвистики могут выступать и термины, и общеупотребительные слова. Показано, что в исследуемой терминологии термины-словосочетания получили широкое распространение. Это объясняется тем, что такие номинации способны более точно, чем однословные термины, передавать конкретное название, образовывать родо-видовые связи и систему отраслевых понятий.

*Ключевые слова:* терминология, компьютерная лингвистика, синтаксический способ, термины-словосочетания, терминообразование.

DOI: 10.35634/2412-9534-2023-33-4-805-810

Синтаксический способ терминообразования является одним из самых продуктивных. По мнению В.П. Даниленко, этот факт вызван сложной внутренней соотносительностью «в кругу понятий современной науки и техники» [6, с. 103]. Исследовательница отметила, что термины-словосочетания, которые выражают «единые целостные понятия, обладая разной степенью смысловой разложимости, в целом более устойчивы по сравнению со свободными словосочетаниями общелитературного языка. По своей лексико-семантической организации их можно отнести к числу лексически ограниченных словосочетаний, характерной особенностью которых является то, что место одного из компонентов ... заполняется не любым словом соответствующей категории, а лишь некоторыми, образующими определенную семантическую группу» [6, с. 104].

Составным терминам отдается предпочтение вследствие того, что они «способны с наибольшей полнотой отразить необходимые отличительные признаки именуемого понятия. Терминами-словосочетаниями легче, чем другими словообразующими средствами (аффиксами, например), передать принадлежность классификационному ряду, основанному на родо-видовом соотношении понятий» [6, с. 132]. Не случайно, такие термины получили широкое распространение практически в каждой отраслевой терминологии. Этот факт подтверждается исследованиями терминологических словосочетаний, используемых для обозначений понятий в сферах нефтегазовой промышленности [2], биологии [3], кораблестроения [4], таможенного дела [5], спорта [7], дипломатии [8], биотехнологий [9], нанотехнологий [11], медицины [12], искусственного интеллекта [13], химии [16], общественных наук [17]. При этом остается недостаточно изученным синтаксический способ образования терминов компьютерной лингвистики. Эта профессиональная сфера возникла на стыке направлений лингвистики, информатики, искусственного интеллекта, в последние годы заметно прогрессирует и применяется для решения самых различных прикладных задач [1; 10; 14; 15; 18; 19]. Динамичное развитие средств и технологий автоматической обработки естественно-языковой информации приводит к необходимости номинации новых понятий в сфере компьютерной лингвистики. В таких условиях использование словосочетаний становится продуктивным способом образования терминов. В этой связи исследование особенностей синтаксического способа образования терминов компьютерной лингвистики представляется актуальным.

Цель статьи – раскрыть особенности синтаксического способа образования терминов компьютерной лингвистики.

Проведенный анализ фактического материала показал, что к числу терминов компьютерной лингвистики относятся, в основном, словосочетания. Таковую распространенность составных терминов можно объяснить тем, что словосочетания точнее, чем однословные термины, передают конкретное название, образуют родо-видовые связи, систему отраслевых понятий.

Способ номинации специальных понятий средствами синтаксического словообразования не удовлетворяет требованию краткости термина. Однако возможность максимального отражения в составном термине характерных признаков понятия компенсирует указанный недостаток [6, с. 107].

Термины-словосочетания образуются на основе подчинительной связи и состоят из главных и подчиненных слов. Например, с главным компонентом *корпус* созданы десятки словосочетаний (*Брауновский корпус, выровненный корпус, диалектный корпус, диахронический корпус, лингвистический корпус, многоязычный корпус, мониторинговый корпус, морфологический корпус, национальный корпус, открытый корпус, параллельный корпус, речевой корпус, семантический корпус, синхронический корпус* и т.п.), в которых зависимые компоненты конкретизируют признаки главного понятия.

Примерами словосочетаний с главным компонентом *перевод* являются следующие термины: *автоматический перевод, машинный перевод, машинный перевод на основе правил, машинный перевод с помощью языка-посредника*. Многие терминологические словосочетания образованы на основе главного слова анализ, например: *авторегрессионный анализ, автоматический синтаксический анализ, анализ тональности текста, интеллектуальный анализ естественно-языковых данных*. Чем больше слов входит в состав словосочетания, тем точнее обозначены понятия, например, *система перевода, система машинного перевода, многоязычная система машинного перевода*.

Среди словосочетаний имеются терминологические единицы с предлогами (*генерация текста на естественном языке, машинный перевод с участием человека*) и без предлогов (*автоматическое реферирование, дерево парсинга, извлечение мнений, частота термина, коммуникативная структура высказывания*). Функционируют также примеры достаточно сложных терминологических словосочетаний с моделями согласования и управления, например, *алгоритмы иерархической кластеризации, распознавание и синтез звучащей речи, метод обратного распространения ошибки, машинный перевод с помощью языка-посредника, приложения по автоматической обработке и анализу текстов на естественном языке*.

По структуре терминологические словосочетания делятся на двух-, трех-, четырех- и многокомпонентные (по количеству полнозначных слов). В терминологии компьютерной лингвистики преобладают двухкомпонентные специальные названия. В состав наибольшего количества словосочетаний в терминологии компьютерной лингвистики входят существительное, как главное слово, и прилагательное, указывающее на признак понятия, названного существительным (*валентная структура, компьютерная семантика, лингвистический процессор, поисковый запрос*). Менее продуктивной является структура, которую составляют существительное и причастие (*анализирующий фильтр, выровненный корпус, уточняющие настройки*). Достаточно распространенным вариантом двухкомпонентного терминологического словосочетания является структура, построенная из двух существительных, первое из которых является главным (*значение слова, источник звука, контекст употребления, рубрицирование документа*). К двухкомпонентным терминам телекоммуникационной отрасли также следует отнести словосочетания, в которых между существительными используют предлог (*выравнивание по словам, ответ на запрос, ресурсы по лингвистике*). В состав некоторых двухкомпонентных телекоммуникационных терминов входят антропонимические названия (*закон Ципфа, преобразование Фурье, цепь Маркова*).

Двухкомпонентные словосочетания в терминологии компьютерной лингвистики могут состоять из двух терминов (*акустический спектр, синтаксический анализатор, фонетическая транскрипция, широкополосная спектрограмма*). Однако достаточно распространенными являются соединения, в которых один компонент является термином, а другой – общеупотребительным словом, перешедшим в разряд терминов (*анализ речи, кластеризация текстов, семантическая роль, формантная картина*). В терминологии компьютерной лингвистики функционируют также двухкомпонентные словосочетания, состоящие из общеупотребительных слов, перешедших в разряд терминов (*голосовой помощник, контролируемый язык, полнота поиска, словарная статья, уровень доверия*).

На основе двухкомпонентных словосочетаний образуется немало трехкомпонентных терминологических единиц (*именованные сущности – извлечение именованных сущностей, информационный поиск – базовый информационный поиск, разметка текста – морфологическая разметка текста, синтез речи – автоматический синтез речи*).

В терминологии компьютерной лингвистики используются следующие структурные виды трехкомпонентных словосочетаний:

1) 'существительное + существительное + существительное' (*анализ тональности текста, дерево принятия решений, система управления терминологией, средство автоматизации перевода*);

2) 'существительное + прилагательное (причастие) + существительное' (*метод скользящего окна, извлечение именованных сущностей, коэффициент взаимной информации, разрешение лексической многозначности, система автоматического перевода, ширина анализирующего фильтра, язык семантической разметки*);

3) 'прилагательное (причастие) + существительное + существительное' (*артикуляторный синтез речи, коммуникативная структура высказывания, критическая полоса слуха, цифровая обработка сигнала*); 'прилагательное + существительное + предлог + существительное' (*машинный перевод с трансфером*);

4) 'прилагательное (причастие) + прилагательное (причастие) + существительное' (*базовый информационный поиск, выровненный двуязычный корпус, искусственная нейронная сеть, лингвистические информационные ресурсы, статистический машинный перевод, теоретическая компьютерная лингвистика, широкая фонетическая транскрипция*).

Четырехкомпонентные словосочетания в исследуемой терминологии образуются в соответствии со следующими видами структур:

1) 'существительное + существительное + существительное + существительное' (*ширина полосы пропускания фильтра*);

2) 'существительное + прилагательное (причастие) + существительное + существительное' (*система автоматического чтения текста*);

3) 'существительное + существительное + прилагательное + существительное' (*автоматическая обработка естественного языка, программа выравнивания параллельного текста*); 'существительное + существительное + предлог + прилагательное + существительное' (*генерация текста на естественном языке*);

4) 'прилагательное (причастие) + существительное + прилагательное (причастие) + существительное' (*двуязычная система машинного перевода, многоязычная система машинного перевода*); 'прилагательное (причастие) + существительное + прилагательное (причастие) + предлог + существительное' (*машинный перевод, основанный на знаниях*);

5) 'прилагательное (причастие) + существительное + существительное + существительное' (*стартовая точка окна данных, шаговый сдвиг окна данных*); 'прилагательное (причастие) + существительное + предлог + существительное + существительное' (*машинный перевод на основе правил, машинный перевод с помощью языка-посредника, машинный перевод с участием человека*);

6) 'прилагательное + прилагательное + прилагательное + существительное' (*сложная периодическая звуковая волна*);

7) 'существительное + существительное + числительное + существительное' (*цепь Маркова первого порядка*).

К терминам компьютерной лингвистики также относятся словосочетания с более сложным составом, а именно пятикомпонентные (*интеллектуальный анализ данных на естественном языке, обработка речевого сигнала методом окна, технология интерактивной визуализации тематических моделей, уточняющие настройки системы обработки речи*), шестикомпонентные (*извлечение информации из текстов на базе лексико-синтаксических шаблонов, программа автоматической обработки текстов на естественном языке*), семикомпонентные (*аддитивная регуляризация тематических моделей коллекций текстовых документов, семейство систем извлечения информации из мультязычных коллекций документов*) и даже восьмикомпонентные структуры (*лексический ресурс для анализа тональности социально-политических текстов на русском языке*).

Тенденцию функционирования многокомпонентных словосочетаний в отраслевых терминологиях можно объяснить стремлением передать необходимые и достаточные признаки обозначаемых понятий, то есть попыткой добиться максимальной мотивированности терминов.

Таким образом, проведенное исследование позволило раскрыть особенности синтаксического способа образования терминов компьютерной лингвистики. Чрезвычайная распространенность терминов-словосочетаний компьютерной лингвистики обусловлена их способностью точнее, чем однословные термины, передать конкретное название, образовывать родо-видовые связи и систему отраслевых понятий. Моделями синтаксического способа терминообразования в вышеуказанной сфере служат разнородные словосочетания, основными из которых являются двухкомпонентные терминологические единицы с главным словом существительным. Наибольшими по объему обнаружены восьмикомпонентные термины-словосочетания. Незначительное количество многокомпонентных терминов обусловлено узкой направленностью, конкретизацией.

## СПИСОК ИСТОЧНИКОВ И ЛИТЕРАТУРЫ

1. Агузумцян Р.В., Великанова А.С., Польщиков К.А., Игитян Е.В., Лихошерстов Р.В. О применении интеллектуальных технологий обработки естественного языка и средств виртуальной реальности для поддержки принятия решений при подборе исполнителей проектов // Экономика. Информатика. 2021. № 48 (2). С 392–404.
2. Александрова Е.В. Классификация технических терминов нефтегазовой отрасли // Вестник Волгоградского государственного университета. Серия 9: Исследования молодых ученых. 2020. № 18. С. 11–13.
3. Базалина Е.Н. К проблеме лингвистических классификаций специальных терминов в области инженерной биологии // Язык и культура. 2014. № 11. С. 92–96.
4. Беспамятная Л.В., Черная Т.Н., Свистунов А.С. Термины-словосочетания как особый вид терминологической лексики (на материале субстантивных терминов кораблестроения) // Modern Science. 2022. № 2-1. С. 363–366.
5. Гросс В.О., Попова М.И. Многокомпонентные термины в терминосистеме таможенного дела // Молодежный вестник ИрГТУ. 2015. № 2. С. 39.
6. Даниленко В.П. Русская терминология. Опыт лингвистического описания. М.: Наука, 1977. 246 с.
7. Дронякина Н.В., Старых Д.А. Словообразовательная номинация в англоязычном терминологическом образовании смешанных единоборств // Вестник Кемеровского государственного университета. 2020. № 1(81). С. 226–233.
8. Криворот В.В. О некоторых особенностях дипломатической терминологии в английском языке // Актуальные научные исследования в современном мире. 2017. № 1-5(21). С. 74–77.
9. Кудинова Т.А. Способы компрессии словообразовательной структуры терминологических словосочетаний (на материале английского подъязыка биотехнологий) // Вестник Омского государственного педагогического университета. Гуманитарные исследования. 2018. № 1(18). С. 63–66.
10. Махди Т.Н., Игитян Е.В., Польщиков К.А., Корсунов Н.И. Оценка эффективности функционирования диалоговой системы на основе применения нечеткого вывода с нейросетевой настройкой // Экономика. Информатика. 2022. № 2. С. 356–374.
11. Милуд М.Р. Структурные особенности профессионально-коммуникативных подъязыков компьютерной терминологии и нанотехнологии // Коммуникативные исследования. 2019. № 4. С. 1072–1091.
12. Молокова Э.Ю. Категория ограничения в концептуальной структуре медицинского термина-словосочетания // Бюллетень Северного государственного медицинского университета. 2020. № 1(44). С. 300–301.
13. Мусаева А.С. Общезыковой субстрат терминологии искусственного интеллекта (на примере терминов-словосочетаний) // Актуальные вопросы современной филологии и журналистики. 2022. № 3(46). С. 12–19.
14. Польщиков К.А., Лазарев С.А., Константинов И.С., Польщикова О.Н., Свойкина Л.Ф., Игитян Е.В., Балакшин М.С. Модель для оценки эффективности выполнения робототехнической системой коммуникативных функций // СТИН. 2020. № 6. С. 4–7.
15. Польщиков К.А., Польщикова О.Н., Игитян Е.В., Балакшин М.С. Алгоритм поддержки принятия решений по выбору средств обработки больших массивов естественно-языковых данных // Научные ведомости БелГУ. Серия: Экономика. Информатика. 2019. № 3. С. 553–562.
16. Салимова И.М. Структурно-языковые средства выражения химических понятий // Профессиональный проект: идеи, технологии, результаты. 2013. № 4(13). С. 140–142.
17. Юдина И.Г., Лаврик О.Л. Развитие русскоязычных и англоязычных терминов и устойчивых словосочетаний в области взаимоотношения общества и науки // Социология науки и технологий. 2022. № 3. С. 89–117.
18. Polshchikov K.A., Velikanova A.S., Igityan E.V. Neural network natural language processing tools for identifying personal priorities in the project performers selection in the field of smart agriculture // IOP Conference Series: Earth and Environmental Science. 2022. Vol. 1069. pp. 012012.
19. Velikanova A.S., Polshchikov K.A., Likhosherstov R.V., Polshchikova A.K. The use of virtual reality and fuzzy neural network tools to identify the focus on achieving project results // Journal of Physics: Conference Series. 2021. Vol. 2060. pp. 173707.

Поступила в редакцию 12.02.2023

Польщикова Ольга Николаевна, кандидат филологических наук,  
доцент кафедры русского языка, профессионально-речевой и межкультурной коммуникации  
ФГАОУ ВО «Белгородский государственный национальный исследовательский университет»  
308015, Россия, г. Белгород, ул. Победы, 85  
E-mail: polshchikova@bsu.edu.ru

*O.N. Polishchikova***SYNTAXIC METHOD OF COMPUTER LINGUISTICS TERMS FORMATION**

DOI: 10.35634/2412-9534-2023-33-4-805-810

The article is devoted to the analysis of the features of the syntactic method of forming terms in computational linguistics. The relevance of the study of compound terms in the professional field of computational linguistics is substantiated. In the terminology under study, there are models of multi-component phrases, the main of which are two-component terminological units with a noun as the main word. Among the identified phrases, there are terminological units with and without prepositions, rather complex terminological phrases with models of coordination and control function. As part of two-, three-, four- and multi-component terms, there are models of various combinations of parts of speech. The components of terminological phrases in computational linguistics can be both terms and commonly used words. It is shown that in the terminology under study, the terms-phrases are widely used. This is explained by the fact that such nominations are capable of conveying a specific name more accurately than single-word terms, forming genus-species relationships and a system of industry concepts.

*Keywords:* terminology, computational linguistics, syntactic method, terms-phrases, term formation.

## REFERENCES

1. Aguzumcyan R.V., Velikanova A.S., Pol'shchikov K.A., Igityan E.V., Lihosherstov R.V. O primeneniі intel'ktual'nyh tekhnologij obrabotki estestvennogo yazyka i sredstv virtual'noj real'nosti dlya podderzhki prinyatiya reshenij pri podbore ispolnitelej proektov [Application of intellectual technologies of natural language processing and virtual reality means to support decision-making when selecting project executors] // *Ekonomika/ Informatika* [Economics. Information technologies]. 2021. № 48 (2). pp. 392–404. (In Russian).
2. Aleksandrova E.V. Klassifikaciya tekhnicheskikh terminov neftegazovoj otrasli [Classification of technical terms of the oil and gas industry] // *Vestnik Volgogradskogo gosudarstvennogo universiteta* [Bulletin of the Volgograd State University]. Series 9: Young Scientists Research. 2020. № 18. pp. 11–13. (In Russian).
3. Bazalina E.N. K probleme lingvisticheskikh klassifikacij special'nyh terminov v oblasti inzhenernoj biologii [To the problem of linguistic classifications of special terms in the field of engineering biology] // *Yazik i kultura* [Language and Culture]. 2014. № 11. P. 92–96. (In Russian).
4. Bespamyatnaya L.V., Chernaya T.N., Svistunov A.S. Terminy-slovosochetaniya kak osobyj vid terminologicheskoy leksiki (na materiale substantivnyh terminov korablestroeniya) [Terms-phrases as a special type of terminological vocabulary (based on substantive shipbuilding terms)] // *Modern Science*. 2022. № 2-1. pp. 363–366. (In Russian).
5. Gross V.O., Popova M.I. Mnogokomponentnye terminy v terminosisteme tamozhennogo dela [Multicomponent terms in the terminological system of customs affairs] // *Molodezhnyy vestnik ISTU* [Youth Bulletin of ISTU]. 2015. № 2. P. 39. (In Russian).
6. Danilenko V.P. Russkaya terminologiya. Opyt lingvisticheskogo opisaniya [Russian terminology. Experience of linguistic description]. M.: Nauka, 1977. 246 p. (In Russian).
7. Dronyakina N.V., Staryh D.A. Slovoobrazovatel'naya nominaciya v angloyazychnom terminoobrazovanii smeshannyh edinoborstv [Word-building nomination in the English-language term formation of mixed martial arts] // *Vestnik Kemerovskogo gosudarstvennogo universiteta* [Bulletin of the Kemerovo State University]. 2020. № 1(81). pp. 226–233. (In Russian).
8. Krivorot V.V. O nekotoryh osobennostyah diplomaticheskoy terminologii v anglijskom yazyke [On some features of diplomatic terminology in English] // *Aktualniye nauchnye issledovaniya v sovremennom mire* [Actual scientific research in the modern world]. 2017. № 1-5(21). pp. 74–77. (In Russian).
9. Kudinova T.A. Sposoby kompressii slovoobrazovatel'noj struktury terminologicheskikh slovosochetaniy (na materiale anglijskogo pod'yazyka biotekhnologij) [Ways of compression of the word-formation structure of terminological phrases (on the material of the English sublanguage of biotechnology)] // *Vestnik Omskogo gosudarstvennogo pedagogicheskogo universiteta. Gumanitarnye nauki* [Bulletin of the Omsk State Pedagogical University. Humanitarian research]. 2018. № 1(18). pp. 63–66. (In Russian).
10. Mahdi T.N., Igityan E.V., Pol'shchikov K.A., Korsunov N.I. Ocenivanie effektivnosti funkcionirovaniya dialogovoj sistemy na osnove primeneniya nechetkogo vyvoda s nejrosetevoj nastrojkoj [Evaluation of the effectiveness of the functioning of the dialogue system based on the use of fuzzy inference with neural network tuning] // *Ekonomika/ Informatika* [Economics. Information technologies]. 2022. № 2. pp. 356–374. (In Russian).
11. Milud M.R. Ctrukturnye osobennosti professional'no-kommunikativnyh pod'yazykov komp'yuternoj terminologii i nanotekhnologii [Structural features of professional communicative sublanguages of computer terminology and nanotechnology] // *Kommunikativniye issledovaniya* [Communicative Research]. 2019. № 4. pp. 1072–1091. (In Russian).

12. Molokova E.Yu. Kategoriya ogranicheniya v konceptual'noj strukture medicinskogo termina-slovochetaniya [Restriction category in the conceptual structure of a medical term-phrase] // Byulleten Severnogo gosudarstvennogo medicinskogo universiteta [Bulletin of the Northern State Medical University]. 2020. № 1(44). pp. 300–301. (In Russian).
13. Musaeva A.S. Obshcheyazykovoj substrat terminologii iskusstvennogo intellekta (na primere terminov-slovochetanij) [General language substratum of terminology of artificial intelligence (on the example of terms-phrases)] // Aktualniye voprosi sivremennoy filologii i zurnalistiki [Topical issues of modern philology and journalism]. 2022. № 3(46). pp. 12–19. (In Russian).
14. Pol'shchikov K.A., Lazarev S.A., Konstantinov I.S., Pol'shchikova O.N., Svojkina L.F., Igityan E.V., Balakshin M.S. Model' dlya ocenki effektivnosti vypolneniya robototekhnicheskoy sistemoj kommunikativnyh funkciy [A model for evaluating the effectiveness of a robotic system performing communicative functions] // STIN. 2020. № 6. pp. 4–7. (In Russian).
15. Pol'shchikov K.A., Pol'shchikova O.N., Igityan E.V., Balakshin M.S. Algoritm podderzhki prinyatiya reshenij po vyboru sredstv obrabotki bol'shih massivov estestvenno-yazykovyh dannyh [Decision support algorithm for the choice of means for processing large arrays of natural language data] // Nauchnye Vedomosti BelSU. Seriya: Ekonomika. Informatika [Nauchnye Vedomosti BelSU. Series: Economy. Computer science]. 2019. № 3. pp. 553–562. (In Russian).
16. Salimova I.M. Strukturno-yazykovye sredstva vyrazheniya himicheskikh ponyatij [Structural and linguistic means of expressing chemical concepts] // Professionalniy proekt: idei, Tehnologii, rezultati [Professional project: ideas, technologies, results]. 2013. № 4(13). pp. 140–142. (In Russian).
17. Yudina I.G., Lavrik O.L. Razvitie russkoyazychnyh i angloyazychnyh terminov i ustojchivyh slovochetanij v oblasti vzaimootnosheniya obshchestva i nauki [Development of Russian and English terms and set phrases in the field of relations between society and science] // Sociologiya nauki i tehnologii [Sociology of science and technology]. 2022. № 3. pp. 89–117. (In Russian).
18. Polshchykov K.A., Velikanova A.S., Igityan E.V. Neural network natural language processing tools for identifying personal priorities in the project performers selection in the field of smart agriculture // IOP Conference Series: Earth and Environmental Science. 2022. Vol. 1069. pp. 012012. (In English).
19. Velikanova A.S., Polshchykov K.A., Likhosherstov R.V., Polshchykova A.K. The use of virtual reality and fuzzy neural network tools to identify the focus on achieving project results // Journal of Physics: Conference Series. 2021. Vol. 2060. pp. 173707. (In English).

Received 12.02.2023

Polshchikova O.N., Candidate of Philology, Associate Professor  
of the Department of Russian Language, Professional Speech and Intercultural Communication  
Belgorod State University  
Pobedy st., 85, Belgorod, Russia, 308015  
E-mail: polshchikova@bsu.edu.ru