



DOI: 10.22363/2313-2299-2022-13-1-80-95

УДК 811.111'42'276.5:62:004.912

Научная статья / Research article

Структурные модели англоязычных терминов для автоматической обработки корпусов научно-технических текстов

Ю.И. Бутенко¹✉, Н.С. Николаева², Е.Ю. Карцева¹

¹Российский университет дружбы народов,
117198, Российская Федерация, Москва, ул. Миклухо-Маклая, 6

²Московский государственный технический университет им. Н.Э. Баумана,
105005, Российская Федерация, Москва, 2-я Бауманская, д. 5, стр. 1

✉ iubutenko@bmstu.ru

Аннотация. В статье рассмотрены структурные модели английских многокомпонентных терминов из предметной области «Welding types» как основа для разметки корпуса научно-технических текстов. Обозначено место корпусов научно-технических текстов в корпусной лингвистике и перспективы дальнейших научных исследований на их основе. Актуальность исследования обусловлена необходимостью создания корпусов научно-технических текстов в целом и средств автоматической разметки терминов в частности. Обосновано, что основной проблемой при создании корпусов научно-технических текстов является автоматическая разметка терминологических словосочетаний. Проведен анализ современного состояния терминосистемы предметной области «Welding types». Рассмотрена формальная структура элементов терминосистемы «Welding types». Представлены результаты анализа двух-, трех-, четырехкомпонентных английских терминологических словосочетаний предметной области «Welding types», а также их структурные модели. Все структурные модели английских терминологических словосочетаний проиллюстрированы примерами. Выделены наиболее продуктивные модели английских терминологических словосочетаний. Показано, что наиболее продуктивная модель — сочетание ядерного элемента с именем существительным или прилагательным в функции препозитивного определения — прослеживается у двухкомпонентных словосочетаний, но анализ более сложных образований показывает, что модель «левое определение, присоединенное к ядру термина» присутствует и в них, демонстрируя родовые признаки. Обоснована необходимость перечисления всех возможных структурных моделей терминологических сочетаний предметной области «Welding types». Новизна исследования видится в формировании базы данных структурных моделей терминологических словосочетаний как основы надкорпусной базы данных о структуре терминов для повышения качества автоматической разметки корпусов научно-технических текстов и обработки терминов-кандидатов при проведении корпусных исследований.

Ключевые слова: термин, терминологическое словосочетание, структурная модель, разметка, корпус научно-технических текстов, научно-технический дискурс

© Бутенко Ю.И., Николаева Н.С., Карцева Е.Ю., 2022



This work is licensed under a Creative Commons Attribution 4.0 International License

<https://creativecommons.org/licenses/by/4.0/>

История статьи:

Дата поступления: 01.03.2021

Дата приема в печать: 15.12.2021

Для цитирования:

Бутенко Ю.И., Николаева Н.С., Карцева Е.Ю. Структурные модели англоязычных терминов для автоматической обработки корпусов научно-технических текстов // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2022. Т. 13. № 1. С. 80—95. <https://doi.org/10.22363/2313-2299-2022-13-1-80-95>

Вклад авторов:

Бутенко Ю.И. — концепция и дизайн исследования использования структурных моделей терминологических словосочетаний для автоматической разметки корпусов научно-технических текстов, Николаева Н.С. — обработка иллюстративного материала, описание моделей терминологических словосочетаний, Карцева Е.Ю. — сбор иллюстративного материала по предметной области, написание текста.

UDC 811.111'42'276.5:62:004.912

Structural Models of English Terms of Automated Processing of Scientific and Technical Texts Corpora

Iuliia I. Butenko¹, Natalia S. Nikolaeva², Elena Yu. Kartseva¹

¹Peoples' Friendship University of Russia (RUDN University),
6, Mikluho-Maklaya str., Moscow, Russian Federation, 117198

²Bauman Moscow State Technical University,
5/1, 2-nd Baymanskaya str., Moscow, Russian Federation, 105005

 iubutenko@bmatu.ru

Abstract. The article is devoted to the structural models of English multi-component terms from the subject area “Welding types” as a basis for marking the corpora of scientific and technical texts. The place of corpora of scientific and technical texts in corpus linguistics and prospects of further scientific research based on them are marked. Relevance of the research is conditioned by the necessity to create the corpus of scientific and technical texts, in general, and means of automatic marking of terms, in particular. It has been substantiated that the main problem in creating the corpus of scientific and technical texts is automatic marking of terminological word combinations. The analysis of the current state of the terminology system of the subject area “Welding types” has been carried out. The formal structure of elements of the “Welding types” terminology system is considered. The results of the analysis of two, three, four-component English terminological word combinations of the “Welding types” subject area and their structural models are presented. All structural models of English terminology combinations are illustrated with examples. The most productive models of English terms word combinations are highlighted. It is shown that the most productive model — the combination of a nucleus element with a noun or an adjective in the function of the prepositional definition — can be traced in two-component word combinations, but the analysis of more complex formations shows that the model of “left definition attached to the term kernel” is also present in them, demonstrating generic features. The necessity of enumerating all possible structural models of terminological combinations in the subject area “Welding types” has been substantiated. The novelty of the study is seen in the formation of a database of structural models of terminological combinations as the basis of a superstructure database on the structure of

terms to improve the quality of automatic marking of the bodies of scientific and technical texts and processing of terms-candidates in the conduct of body studies.

Key words: term, terminological word combination, structural model, markup, scientific and technical texts corpora, scientific and technical discourse

Article history:

Received: 01.03.2021

Accepted: 15.12.2021

For citation:

Butenko, Iu.I., Nikolaeva, N.S. & Kartseva, E.Yu. (2022). Structural Models of English Terms of Automated Processing of Scientific and Technical Texts Corpora. *RUDN Journal of Language Studies, Semiotics and Semantics*, 13(1), 80—95. <https://doi.org/10.22363/2313-2299-2022-13-1-80-95>

Authors' contribution:

Iuliia I. Butenko — the concept and design of the study on the use of structural models of terminological phrases for automatic marking of scientific and technical text corpora, Natalia S. Nikolaeva — processing of illustrative material, description of models of terminological phrases, Elena Yu. Kartseva — collection of illustrative material on the subject area, writing the text.

Введение

Благодаря достижениям современной корпусной лингвистики стало возможным при проведении лингвистических исследований осуществлять сбор иллюстративного материала при помощи специализированных электронных корпусов. Сферы применения корпусов текстов включают автоматизированное извлечение информации, обучение на основе данных, текстовые поиски в крупномасштабных коллекциях текстов с использованием методов обработки естественного языка, автоматическую классификацию текстов, преподавание языков для специальных целей, проведение лингвистических исследований [1].

Распространенными являются корпусы, содержащие тексты художественной литературы, однако наличия корпусов, основанных исключительно на литературно-художественных произведениях, недостаточно для эффективного применения корпусной лингвистики в области лингвистических исследований [2]. Для осуществления исследований в определенной предметной области, т.е. изучения языков для специальных целей, необходимо наличие корпусов научно-технических текстов, которые в настоящее время находятся в закрытом доступе или вовсе отсутствуют. Описание подязыка определенной предметной области требует создания специального корпуса узкоспециальных текстов, так как общие корпуса не всегда подходят для изучения определенных предметных областей в силу их большого объема и разнообразного материала.

Тексты корпусов обычно размечаются для удобства пользования, т.е. текстам и содержащимся в них языковым единицам приписываются специальные метки. Размеченные корпуса обеспечивают специализированными

поисковыми системами, реализующими грамматические и лексические виды поиска [3]. В зависимости от целей создания корпуса в него включают дополнительные виды разметки [4]. Так, для корпуса научно-технических текстов наибольшую значимость приобретает терминологическая разметка, так как именно термины выступают основным средством передачи информации.

Работа с корпусами научно-технических текстов требует особый инструментарий для выявления устойчивых терминологических словосочетаний [5]. Стоит отметить, что в настоящее время лексико-грамматические шаблоны многокомпонентных терминов исследователи создают для каждого специализированного корпуса отдельно, например, для корпусной лингвистики [6], информационной безопасности [7], нанотехнологии [8] и др. В такой ситуации возникает необходимость в разработке специализированной технологии, позволяющей последовательно обрабатывать коллекции текстов разных предметных областей и фиксировать каждую модель многокомпонентного терминологического словосочетания в отдельной базе данных. М.Г. Кружков в работе [3] использует термин для таких специализированных баз данных — «надкорпусные базы данных». Создание надкорпусной базы структурных моделей английских терминологических сочетаний позволит, с одной стороны, использовать уже созданную базу структурных моделей терминологических словосочетаний, а с другой стороны — дополнять и актуализировать базу данных новыми структурными моделями терминов с минимальными временными и человеческими затратами.

Материал и методы исследования

Автоматизация терминологической разметки, которая является обязательным элементом корпуса научно-технических текстов, представляет собой сложную теоретическую и практическую задачу, решение которой требует взаимопроникновения частных методов лингвистики и программирования [9]. Накоплены значительные фонды научно-технических текстов, их обработка требует наличия средств автоматической обработки текстов, учитывающих их стилевые особенности. Так, наибольшую сложность при автоматической разметке представляют собой многокомпонентные термины, выражающие одно понятие, но представленные несколькими лексическими единицами в своем составе. Отмеченные обстоятельства указывают на необходимость создания «надкорпусной» базы структурных моделей терминологических словосочетаний, что позволит использовать разработанную базу для дальнейшей терминологической разметки корпусов научно-технических текстов разных предметных областей.

Материалом исследования послужили терминологические словосочетания предметной области «Welding types», собранные методом сплошной выборки из иллюстративных источников. При обработке иллюстративного материала использован метод структурного анализа для разработки моделей двух-, трех- и четырех компонентных терминологических словосочетаний

как наиболее вероятных в научно-технических текстах. Терминологические словосочетания с пятью и более компонентами не являются часто используемыми моделями, в связи с чем они не были включены в базу.

Таким образом, целью статьи является выявление структурных моделей английских терминологических сочетаний предметной области «Welding types» как основы надкорпусной базы данных о структуре английских терминов для повышения качества автоматической разметки специализированных корпусов научно-технических текстов.

Терминосистема предметной области «Welding types»

На всем протяжении развития общества, науки и техники человек сталкивался с необходимостью не только изобретать способы получения нужной ему продукции, но и потребностью называть придуманные технологии и инструментарий. То есть еще в древности зародились зачатки современных терминологий. Являясь вербализованными результатами речемыслительной деятельности человека, термины проявляются как многофункциональные образования. Во-первых, с исторической точки зрения представляют собой элементы когниции, т.е. отражают знания, зафиксированные в определенный момент развития отрасли. Во-вторых, термины могут способствовать развитию данной предметной области, иногда опережая и вызывая теоретические и практические изыскания. В-третьих, будучи элементами не только языка для специальных целей, но и просто языка, термины являются инструментом для общения в данной предметной области, представляют собой понятийные единицы для передачи накопленных знаний и опыта.

В условиях научно-технического прогресса отраслевые терминологии и терминосистемы представляют собой постоянно изменяющиеся информационные базы по разным видам деятельности. Каждая вновь появляющаяся в настоящее время предметная область предполагает необходимость номинации. В связи с этим возникает существенный вопрос: являются ли принципы зарождения и развития терминов, формирования их структуры, а также их группировки в терминологии хаотичными, т.е. специфичными для каждой терминологии, или эти процессы подчиняются не просто отдельным общим принципам, а основополагающим законам [10]. Ответ на этот вопрос является основополагающим для планирования работ по унификации и стандартизации терминов с целью избежать множественных синонимичных названий, омонимии и полисемии в терминологии.

Для выявления именно устойчивых терминологических сочетаний и рассмотрения уже устоявшихся структурных моделей с последующей возможностью их разметки в корпусах научно-технических текстов была выбрана терминосистема предметной области «Welding types». Тексты предметной области «Сварка» во многом терминологичны, и поэтому требуется разработка принципов и методов автоматизированного выделения терминов и терминологических сочетаний [6].

Участие языкового субстрата в структуре термина, существование термина в виде лексической единицы определенного естественного языка проявляется в том, что термин обладает определенной формальной структурой. Подавляющее большинство терминов имеет формальную структуру лексических единиц того естественного языка, на основе которого они образуются [11]. Термин может быть однокомпонентным и состоять из ключевого слова или представлять собой терминологическую группу, в состав которой входит ключевое слово или ядро группы, одно или несколько левых определений и одно или несколько правых или предложных определений, которые уточняют или модифицируют смысл терминологической единицы [12].

Наиболее сложным явлением в процессе автоматической разметки терминов в корпусе научно-технических текстов представляют собой многокомпонентные термины — терминологические словосочетания, образованные лексическим и синтаксическим способами, то есть по определенным моделям. Способ образования терминов в виде цепочки слов часто используется на практике [13].

В основе анализа терминологических словосочетаний лежит вычленение исходного терминологического словосочетания и определение последовательности присоединения к нему остальных элементов. Исходным терминологическим словосочетанием, как правило, является двухкомпонентное субстантивное терминологическое словосочетание, которое в рамках трех-, четырехкомпонентного терминологического словосочетания характеризуется более тесными структурно-семантическими отношениями [14].

Таким образом, при создании системы автоматической разметки корпуса научно-технических текстов необходимо установить все возможные структурные модели русских многокомпонентных терминов на примере предметной области «Welding types».

Структурные модели терминологических единиц предметной области «Welding types»

При рассмотрении терминологии «Welding types» следует отметить присутствие небольшого количества **однокомпонентных** терминов. Их малое количество в основном массиве говорит о невысокой продуктивности данной модели в исследуемой области (*arcogen, rewelding, tacking, Union-Melt, welding*). Скорее всего это связано с узостью или, наоборот, слишком большим объемом понятия и недостаточной дифференциацией технологий, ими обозначаемых (как в случае с термином *welding*) [15].

Анализ английского массива терминов исследуемой понятийной области показал, что наиболее продуктивным способом номинации является синтаксический способ — образование терминов-словосочетаний, состоящих из двух, трех и более компонентов. При этом можно наблюдать, как растет семантическая дифференциация терминов по мере развития каждого направления предметной области. При рассмотрении выявленных многокомпонентных

терминов в 90% случаев зафиксирована структура, которая подразумевает расположение ядерного элемента, т.е. элемента, ответственного за номинативный характер лексической единицы, в конце словосочетания. Элемент, отвечающий за расширение понятийного наполнения термина (определение), находится в препозиции (левое определение). Преобладание такой модели обусловлено самой понятийной областью, в которой видовой, атрибутивный, элемент описывает способ защиты, материал или процесс, с помощью которого и происходит соединение материалов, т.е. сварка. Во всех представленных далее схемах ядерным компонентом является термин *welding* (сварка) как родовой термин данной терминологии.

Наиболее распространенными (порядка 60%) терминами-словосочетаниями являются **двухкомпонентные** образования, что, с одной стороны, свидетельствует о достаточной информативности выявленных терминов, с другой же стороны, говорит об удобстве их использования в устной и письменной профессиональной коммуникации [16—17]. Дальнейший анализ структуры представленных терминов показал наличие моделей, образованных с помощью разнообразных принципов, характерных для английского языка как языкового субстрата.

Наибольшее количество примеров продемонстрировало структуру «существительное + существительное (ядерный элемент)» (*arc welding, blacksmith welding, braze welding, diffusion welding, explosion welding, acetylene welding*), где первое существительное выступает в роли препозитивного определения. В половине случаев определение выражено простым существительным (*deck welding, die welding, fire welding, flame welding, flash welding, gravity welding, thermit welding, wash welding*, и др.) Но в массиве выделяется существенная часть сложных существительных-определений, демонстрирующих структурное разнообразие, характерное для англоязычного словообразования и, как следствие, образования терминов.

Атрибутивный элемент может быть выражен существительными, представляющими собой сложные лексемы с нулевым интерфиксом и дефисным написанием (*air-acetylene welding, argon-arc welding, carbon-electrode welding, energy-storage welding, friction-stir welding, electron-beam welding, radio-frequency welding, solid-phase welding, solid-state welding* и др.).

Еще одной частотной моделью атрибутивного элемента может быть сочетание основ прилагательного (*bare-electrode welding, high-frequency welding, magnetic-pulse welding*), причастия (*heated-tool welding, controlled-transfer welding*), числительного (*three-phase welding, two-electrode welding*) и существительного, которое, являясь главным компонентом модели, отвечает за принадлежность полученного определения к разряду существительных.

Следующей, менее продуктивной, группой моделей, выделяемой в общем массиве двухкомпонентных терминов, является «прилагательное + существительное (ядерный элемент)». Как и в случае, когда атрибутивный элемент выражен существительным, мы имеем дело с простыми (*chemical welding, cold*

welding, dielectric welding, electric welding, plastic welding,) и сложными (*gaselectrical welding, electro-chemical welding*) прилагательными. Сложные прилагательные представляют собой образования, полученные сложением основ с использованием интерфикса -o-. В половине случаев используется дефисное написание (chemical (усечение суффиксов прилагательного) + o + mechanical — *chemico-mechanical welding*; electric (усечение суффикса прилагательного) + o + mechanical — *electro-mechanical welding*; electric (усечение суффикса прилагательного) + o + slag — *electro-slag welding*). Другая часть сложных прилагательных представлена терминами также с интерфиксом -o-, но со слитным написанием (*electropercussive welding, electrostatic welding, electromagnetic welding*).

Другая малопродуктивная модель с препозитивным определением строится по схеме «причастие + существительное (ядерный элемент)» (*powdered welding, flux-cored welding, pressure-controlled welding*), где причастие может быть простым по своей структуре, как в случае с *powdered welding*, и может быть сложным, как в терминах *flux-cored welding* и *pressure-controlled welding*.

Модель, по своей структуре противоположная представленным ранее, т.к. ядерный элемент в ней находится перед атрибутивным элементом, показала свою малую продуктивность. Образованная по схеме «существительное (ядерный элемент) + существительное с предлогом (*in, with*)» (*welding in air, welding in space, welding-on with pressure*), она встречается только приблизительно в 1% случаев номинации. Кроме того, эволюция терминологии показывает, что некоторые подобные термины преобразуются в синонимичные в соответствии с моделью «существительное + существительное (ядерный элемент)». Так, например, вместо *welding in space* в настоящее время чаще встречается *space welding*, а *welding with pressure* в ряде случаев функционирует параллельно с термином *pressure welding*, хотя в некоторых случаях их значения и расходятся.

Трехкомпонентные терминологические словосочетания занимают второе место по встречаемости в предметной области «**Welding types**». Наиболее продуктивной является структурная модель, образованная присоединением препозитивного определения слева к ядерному элементу, как и в двухкомпонентных образованиях. При проведении структурного анализа атрибутивного элемента выявляется наличие как простых, так и сложных моделей.

В качестве атрибутивного элемента в подобной конструкции в большинстве выступают, как и в двухкомпонентных образованиях, существительные, прилагательные и причастия. Сложность для понимания подобных конструкций может представлять выявление смыслового присоединения атрибутивного элемента к ядерному.

Так, в модели «прилагательное + существительное + существительное (ядерный элемент)» (*atomic hydrogen welding, cold pressure welding, electromagnetic arc welding, magnetic discharge welding, straight flash welding, continuous feed welding, independent arc welding*, и др.), прилагательное может

образовывать понятийное единство с первым существительным, образуя самостоятельно функционирующие отдельно от рассматриваемого терминополья терминологические двухкомпонентные словосочетания (*atomic hydrogen, continuous feed, independent arc, magnetic discharge*, и др.), в дальнейшем присоединяемые к ядерному элементу. Но, в то же время, возможны параллельные связи внутри трехкомпонентного словосочетания, например *cold pressure welding* — это одновременно и *cold welding*, и *pressure welding*; *manual arc welding* является сочетанием *manual welding* с *arc welding*; а *remote laser welding* представляет собой комбинацию *remote welding* и *laser welding*.

Подобным образом строится модель, в которой вместо прилагательного первым компонентом является причастие настоящего или прошедшего времени «причастие + существительное + существительное (ядерный элемент)» (*closed joint welding, hammered resistance welding, hidden arc welding, submerged arc welding, rotating arc welding*). В подобной модели проблемы с определением понятийных взаимоотношений в пределах модели отсутствуют, т.к. причастие определяет ближайшее к нему существительное.

Рассматривая структуру, где второе существительное заменено прилагательным, «прилагательное + прилагательное + существительное (ядерный элемент)», выявляется другое, параллельное, подчинение опорному элементу. Так, *electromagnetic percussive welding* представляет собой сочетание *electromagnetic welding* и *percussive welding*; а *electrostatic percussive welding* сочетает в себе предметные области *electrostatic welding* и *percussive welding*.

Еще одной продуктивной моделью трехкомпонентных терминологических словосочетаний является «существительное + существительное + существительное (ядерный элемент)» (*carbon arc welding, laser hybrid welding, oxyacetylene pressure welding, stick electrode welding, pressure thermit welding*). Данные модели являются еще одним примером, демонстрирующим разницу в понятийном соподчинении в рамках представленной структуры. С одной стороны, можно говорить о параллельном соподчинении, как в случае с *fusion thermit welding, contact arc welding, pulse spot welding, pressure contact welding*, где каждое из зависимых слов может образовывать самостоятельно функционирующее терминологическое словосочетание с ядерным элементом — *fusion welding* и *thermit welding, contact welding* и *arc welding, spot welding* и *pulse welding, pulse welding* и *spot welding*. Другой, так же часто встречаемой комбинацией понятий в модели, является присоединение еще одного атрибутивного элемента слева к уже имеющемуся двухкомпонентному терминологическому словосочетанию, что также является достаточно логичным явлением, отражающим процесс дифференциации понятия, как, например, *vacuum (diffusion welding)* — расширение понятия, выраженное термином *vacuum*, добавлено к словосочетанию *diffusion welding*.

Но данная модель может представлять и другое комбинирование понятий. Так в сочетании (*laser (hybrid) welding*) исходным является *laser welding*, а термин *hybrid* относится ко всему словосочетанию, являясь своего рода

термином-вставкой. В сочетании *dip transfer welding* разбивка на отдельные словосочетания невозможна, т.к. обозначаемое понятие может быть отражено триединым нечленимым образованием.

Еще одной особенностью структурирования **трехкомпонентных** словосочетаний оказалось, что наиболее продуктивная модель строится по принципу сочетания сложных терминослов.

Сложным по своему составу может быть первый компонент, как в случае с *bare-metal arc welding*, *bare-wire arc welding*, *constant-pressure pressure welding*, *laser-plasma hybrid welding*, и др., составленный в большинстве случаев из основ прилагательного и существительного (*constant-temperature pressure welding*, *magnetic-flux arc welding*) или, в меньшей степени, существительного и существительного (*oxy-city gas welding*), давая в итоге сложное существительное. В таком случае модель сводится к уже рассмотренной ранее — «существительное + существительное + существительное (ядерный элемент)». Но, данной модели может соответствовать и образование, в котором сложным по своему составу является второй компонент словосочетания. Так, например, в терминах *argon metal-arc welding*, *argon tungsten-arc welding*, *gas carbon-arc welding*, *gas metal-arc welding*, *twin carbon-arc welding* второй компонент является комбинацией двух основ существительных.

Существенная часть массива трехкомпонентных словосочетаний представлена словосочетаниями, одним из компонентов которых является причастие настоящего и прошедшего времени, и один из компонентов является сложным словом. При этом могут наблюдаться следующие модели «прилагательное + сложное существительное + существительное (ядерный элемент)» (*automatic covered-electrode welding*); «причастие + сложное существительное + существительное (ядерный элемент)» (*controlled tungsten-arc welding*); «сложное существительное + причастие + существительное (ядерный элемент)» (*inert-gas shielded welding*); «сложное причастие + существительное + существительное (ядерный элемент)» (*gas-shielded arc welding*, *vapour-shielded arc welding*, *CO2-shielded arc welding*, *self-adjusting arc welding*, *semi-submerged arc welding*).

Сложными по своему составу могут быть и оба атрибутивных компонента, причем наиболее часто встречается модель «сложное существительное + сложное существительное + существительное (ядерный элемент)» (*metal-arc inert-gas welding*, *inert-gas carbon-arc welding*, *inert-gas metal-arc welding*, *impregnated-tape metal-arc welding*), и намного реже используется модель «сложное существительное + сложное причастие + существительное (ядерный элемент)» (*magnetic-flux gas-shielded welding*). Сложные компоненты в данном случае могут быть образованы при помощи соединения основ прилагательных (*magnetic*, *inert*), существительных (*metal*, *flux*, *carbon*) и причастий (*shielded*, *impregnated*). Еще одним вариантом получения трехкомпонентного словосочетания при помощи соединения сложных терминослов является использование аббревиатур. Хотя в исследуемой тематической

области обнаружен только один пример подобного образования (*a-c single-phase welding*), но, возможно, потенциал подобного подхода к созданию сложных терминов недооценен.

Следует отметить в единичных случаях встречающуюся в исследуемой понятийной области модель, в которой ядерный компонент находится не в конечной позиции, а в центральной «существительное + существительное (ядерный элемент) + существительное с предлогом in» (*diffusion welding in vacuum*) и даже в начальной «существительное (ядерный элемент) + прилагательное/причастие + существительное с предлогом (under, with)» (*welding under controlled atmosphere, welding with independent arc*). Редкая встречаемость подобных моделей свидетельствует об их невысокой продуктивности в английском терминообразовании.

По материалам лексикографических источников **четырёхкомпонентные** терминологические словосочетания встречаются гораздо реже **двух- и трёхкомпонентных**. Хотя точное количество терминов, которые можно отнести к той или иной группе, выяснить затруднительно, так как во многих случаях терминоэлемент «*arc*» может присутствовать или отсутствовать в терминологическом словосочетании без потери смысла, что отражается в словарях выделением скобками элиминируемого элемента (*gas-shielded (arc) welding*). Практика показывает, что в научных текстах по данной предметной области данный элемент зачастую элиминируется, что превращает **четырёхкомпонентные** словосочетания в **трёхкомпонентные** эллиптические варианты терминов, в которых пропущен один из компонентов, а **трёхкомпонентные** термины, соответственно, в **двухкомпонентные**.

Большинство выявленных моделей демонстрирует структуру, проявляющуюся на этапе анализа **двух- и трёхкомпонентных** словосочетаний. Ее можно свести к упрощенной **двухкомпонентной** модели «[препозитивное определение] + существительное (ядерный элемент)», которая уже показала свою продуктивность, т.к. чаще всего встречается во всех рассмотренных терминологических словосочетаниях. В рассматриваемой упрощенной модели [препозитивное определение] представляет собой комбинацию простых и сложных терминослов, присоединяемых к ядерному элементу или напрямую, или образуя самостоятельно функционирующие словосочетания, соединяющиеся в сложный термин самым непредсказуемым образом. Так, например, уже имеющийся **трёхкомпонентный** термин *electron beam welding* получил уточняющий элемент *non-vacuum*, в ходе чего образовалось **четырёхкомпонентное** образование *non-vacuum electron beam welding*. Или можно увидеть объединение 2-х двухкомпонентных словосочетаний (*controlled rate + arc welding*) в одно **четырёхкомпонентное** *controlled rate arc welding*. По такой же схеме образуются терминологические словосочетания *stored energy resistance welding* и *vibrating electrode arc welding*. Еще одна модель образуется по принципу «матрешки», когда к ядерному элементу последовательно присоединяются дифференцирующие термины: к имеющейся **двухкомпонентной** модели

laser welding сначала добавлен 3-й компонент *remote*, а, потом, в одном случае — компонент *robot*, а в другом — *scanner-based*; в итоге получаем два термина — *scanner-based remote laser welding* и *robot remote laser welding*. Другие варианты образования **четырёхкомпонентных** словосочетаний на базе уже имеющихся **двухкомпонентных** представлены:

- параллельным присоединением к двухкомпонентному термину *arc welding* не связанных между собой однословных терминов, обозначающих специфику процесса сварки (*gas + tungsten + (arc welding)*);

- параллельным присоединением однослового термина *metal* и **двухкомпонентного** словосочетания *active gas* к ядерному элементу (*metal (active gas) welding*).

В общем массиве **четырёхкомпонентных** словосочетаний выявлены единичные случаи применения постпозитивных определений, представленных набором существительных с предлогом (*welding in the atmosphere of illuminating gas*).

В выявленном массиве терминов предметной области «Welding types» наблюдается также наличие немногочисленных многокомпонентных словосочетаний с пятью и шестью составляющими. Наличие подобных терминологических словосочетаний свидетельствует, с одной стороны, о стремлении все более детализировать семантику терминологических словосочетаний, отражая эволюцию существующих процессов и образование на их основе новых, с другими характеристиками (*consumable electrode inert arc welding, gas shielded metal arc welding, shielded inert gas metal arc welding, welding with consumable electrodes in an atmosphere of CO2 gas*). С другой стороны, наличие многокомпонентных словосочетаний с большим количеством составляющих является сложным для восприятия и коммуникации, что ведет к их последующему упрощению. Скорее всего процесс упрощения словосочетаний до 3-х и 4-х компонентных мы наблюдаем в сочетаниях, использующих в своих моделях сложные слова (*laser-plasma hybrid welding, gas carbon-arc welding, a-c single-phase welding, impregnated-tape metal-arc welding* и др.).

Заключение

Для упрощения подхода к разметке корпусов терминологических текстов появилась необходимость в выборе достаточно устоявшегося массива терминов, который мог бы продемонстрировать характерные структурные образования, присущие развитым терминологиям и, особенно терминосистемам. В качестве исследуемого массива была выбрана терминосистема предметной области «Welding types», которая, с одной стороны, имеет давнюю историю, а, с другой стороны, основной этап ее наполнения пришелся на XX век, что упрощает подбор и анализ терминов. В связи с тем, что терминосистема предметной области достаточно изучена и стандартизирована, предполагается создание на ее основе надкорпусной базы данных, описывающей структурные модели терминологических словосочетаний предметной области «Welding

types». Рассматривая формальную структуру элементов терминосистемы «Welding types», следует отметить, что наиболее продуктивной моделью является сочетание ядерного элемента с именем существительным или прилагательным в функции препозитивного определения. Наиболее явно данная модель прослеживается у двухкомпонентных словосочетаний, но анализ более сложных образований показывает, что модель «левое определений, присоединенное к ядру термина» присутствует и в них, демонстрируя родовые признаки. Дальнейшее усложнение терминологического словосочетания чаще всего происходит с усложнением постпозитивного определения, несущего в себе видовые особенности. Такой подход позволит не только исследовать частные характеристики моделей терминологических словосочетаний, но и формирование всех возможных продуктивных моделей терминологических словосочетаний, которые будут совпадать со словосочетаниями из других предметных областей.

Библиографический список

1. *Нагель О.В.* Корпусная лингвистика и ее использование в компьютеризированном языковом обучении // *Язык и культура*. 2008. № 4. С. 53—59.
2. *Захаров В.П.* Корпуса русского языка // *Труды института русского языка имени В.В. Виноградова*. 2015. № 6. С. 20—65.
3. *Кружков М.Г.* Информационные ресурсы контрастивных лингвистических исследований: электронные корпуса текстов // *Системы и средства информатики*. 2015. Т. 25. № 2. С. 140—159.
4. *Лесников С.В.* Виды разметок текстовых корпусов русского языка // *Научно-техническая информация. Серия 2. Информационные процессы и системы*. 2019. № 9. С. 27—30.
5. *Захаров В.П., Хохлова М.В.* Автоматическое выявление терминологических словосочетаний // *Структурная и прикладная лингвистика*. 2014. № 10. С. 182—200.
6. *Захаров В.П., Хохлова М.В.* Автоматическое извлечение терминов из специальных текстов с использованием дистрибутивно-статистического метода как инструмент создания тезаурусов // *Структурная и прикладная лингвистика*. 2012. № 9. С. 222—233.
7. *Loukachevitch N., Dobrov B.* Ontological Resources for Representing Security Domain in Information-Analytical System // *Open Semantic Technology for Intelligent Systems*. 2018. № 8. С. 185—191.
8. *Морев Н.А.* К проблеме лингвистического анализа терминологии в области нанотехнологий (о необходимости разработки исследовательского корпуса терминологических единиц) // *Вестник МГЛУ*. 2012. № 13 (646). С. 115—124.
9. *Sidnyaev N.I., Butenko I.I., Bolotova E.E.* Formal Grammar Theory in Recognition Methods of Unknown Objects Automatic Documentation and Mathematical Linguistics. 2020. Vol. 54. Issue 4. P. 215—225.
10. *Николаева Н.С.* Особенности происхождения основных терминов терминосистемы «Сварка» (на материале английской и русской терминологий) // *Вестник Московского государственного областного университета. Серия: Лингвистика*. 2011. № 1. С. 132—138.
11. *Лейчик В.М.* О языковом субстрате термина // *Вопросы языкознания*. 1986. № 5. С. 87—97.
12. *Лейчик В.М.* Исходные понятия, основные положения, определения современного терминоведения и терминографии // *Вестник Харьковского политехнического университета*. 1994. № 1. С. 147—180.
13. *Лотте Д.С.* Основы построения научно-технической терминологии. М.: Изд-во АН СССР, 1961.

14. *Циткина Ф.А.* Терминология и перевод. Львов: Высшая школа, 1988.
15. *Николаева Н.С.* Особенности начального этапа в формировании терминологии (донаучный и протонаучный периоды) [Электронный ресурс] // Гуманитарный вестник. 2013. № 7. Режим доступа: <http://hmbul.bmstu.ru/catalog/pedagog/engped/87.html> (дата обращения: 25.10.2020).
16. *Лейчик В.М.* Оптимальная длина и структура термина // Вопросы языкознания. 1981. № 2. С. 63—73.
17. *Гринева-Гринева С.В., Сорокина Э.А.* Опыт описания формальной структуры термина (на материале английской терминологии лексикологии // Вестник Московского государственного областного университета. Серия: Лингвистика. 2020. № 5. С. 74—85.

Список источников иллюстративного материала

1. *Золотых В.Т.* Англо-русский словарь по сварочному производству, под ред. А.А. Ерохина. М.: Сов. Энциклопедия, 1967.
2. *Николаева Н.С.* Обучение чтению литературы на английском языке по специальности «Сварочное производство»: лексический справочник. М.: Изд. МГТУ им. Н.Э. Баумана, 2015.
3. *Cordovi Marcel A.* Glossary of Foreign Welding Terms, M.A. Cordovi (comp). New York: American Welding Society, 1944.
4. Extensive Experience with Both Consumable Guide and High Speed Conventional Electroslag Welding [Electronic resource]. Режим доступа: <http://www.netwelding.com/serv04.htm#BackgroundDetails> (дата обращения: 17.06.2020).
5. Friction Stir Welding // Wikipedia: the free encyclopedia [Electronic resource]. Режим доступа: http://en.wikipedia.org/wiki/Friction_stir_welding (дата обращения: 11.05.2020).
6. Fusion Splicing // Wikipedia: the free encyclopedia [Electronic resource]. Режим доступа: http://en.wikipedia.org/wiki/Fusion_splicing (дата обращения: 11.05.2020).
7. History of MIG (GMAW) Welding [Electronic resource]. Режим доступа: http://www.netwelding.com/History_MIG%201.htm#AtomicHydrogen (дата обращения: 14.05.2020).
8. History of Welding — in the Beginning [Electronic resource]. Режим доступа: <http://www.netwelding.com/History%20of%20Welding.htm> (дата обращения: 11.05.2020).
9. Laser-Hybrid Welding // Wikipedia: the free encyclopedia. [Electronic resource]. Режим доступа: http://en.wikipedia.org/wiki/Laser-hybrid_welding (дата обращения: 12.05.2020).
10. *Mackenzie L.B.* The Welding Encyclopedia, T.B. Jefferson (comp.). New York: McGray-Hill, 1951.
11. Terms and Definitions. AWS A 3.0-69. // Prep. By AWS Com. On definitions and symbols. Under the dir. of AWS Technical activities com. New York: American Welding Society, 1969.
12. Welding handbook // Fundamentals of welding. Vol. 1. Miami: American Welding Society, 1976.

References

1. Nagel', O.V. (2008). Corpus linguistics and its use in computer-based language teaching. *Language and culture*, 4, 53—59. (In Russ.).
2. Zakharov, V.P. (2015). Corpora of the Russian Language. *Proceedings of the V.V. Vinogradov Russian Language Institute*, 6, 20—65. (In Russ.).
3. Kruzhev, M.G. (2015). Information resources for contrastive studies: electronic text corpora. *Systems and means of informatics*, 25(2), 140—159. (In Russ.).
4. Lesnikov, S.V. (2019). The types of marking of text corpora of the Russian language. *Scientific and Technical Information. Series 2. Information Processes and Systems*, 9, 27—30. (In Russ.).

5. Zakharov, V.P. & Khokhlova, M.V. (2014). Automatic extracting of terminological phrases. *Structural and Applied Linguistics*, 10, 182—200. (In Russ.).
6. Zakharov, V.P. & Khokhlova, M.V. (2012). Automatic term extraction and statistical analysis in a special text corpus as a tool for thesaurus construction. *Structural and Applied Linguistics*, 9, 222—233. (In Russ.).
7. Loukachevitch, N. & Dobrov, B. (2018). Ontological Resources for Representing Security Domain in Information-Analytical System. *Open Semantic Technology for Intelligent Systems*, 8, 185—191.
8. Morev, N.A. (2012). To the problem of linguistic analysis of terminology in the field of nanotechnology (on the need to develop a research corpus of terminological units). *Vestnik of Moscow State Linguistic University. Humanities*, 13(646), 115—124. (In Russ.).
9. Sidnyaev, N.I., Butenko, I.I. & Bolotova, E.E. (2020). Formal Grammar Theory in Recognition Methods of Unknown Objects. *Automatic Documentation and Mathematical Linguistics*, 54(4), 215—225.
10. Nikolaeva, N.S. (2011). Peculiarities of origin of the basic terms of the “Welding” system (on the material of English and Russian terminology). *Bulletin of the Moscow Region State University. Series: Linguistics*, 1, 132—138. (In Russ.).
11. Lejchik, V.M. (1986). About the language substrate of the term. *Topics in the study of language*, 5, 87—97. (In Russ.).
12. Lejchik, V.M. (1994). Original concepts, basic provisions, definitions of modern terminology and terminology. *Bulletin of Kharkiv Polytechnic Institute*, 1, 147—180. (In Russ.).
13. Lotte, D.S. (1961). *Basics of building scientific and technical terminology*. Moscow: Izdatel'stvo AN SSSR. (In Russ.).
14. Tsitkina, F.A. (1988). *Terminology and translation*. L'vov: Vysshaya shkola. (In Russ.).
15. Nikolaeva, N.S. (2013). Peculiarities of the initial stage in the formation of the terminology (the prescientific and protoscientific periods). *Humanities Bulletin of BMSTU*. URL: <http://hmbul.bmstu.ru/catalog/pedagog/engped/87.html>. (accessed: 25.10.2020). (In Russ.).
16. Lejchik, V.M. (1981). Optimal length and structure of the term *Voprosy Jazykoznanija* (*Topics in the study of language*), 2, 63—73. (In Russ.).
17. Grinev-Grinevich, S.V. (2020). On terminology aspects of scientific and technical translation. *Bulletin of the Moscow Region State University. Series: Linguistics*, 6(2), 74—78. (In Russ.).

List of sources of illustrative material

1. Zolotykh, V.T. (1967). *English-Russian dictionary of welding production*, A.A. Erokhin (ed.). Moscow: Sovetskaya Encyclopedia. (In Russ.).
2. Nikolaeva, N.S. (2015). *Teaching reading literature in English in the specialty “Welding”: a lexical guide*. Moscow: MSTU named after N.E. Bauman publ. (In Russ.).
3. Cordovi, Marcel A. (1944). *Glossary of Foreign Welding Terms*, M.A. Cordovi (comp.). New York: American Welding Society.
4. Extensive Experience with Both Consumable Guide and High Speed Conventional Electroslag Welding [Electronic resource]. URL: [http://www.netwelding.com/serv04.htm#Background Details](http://www.netwelding.com/serv04.htm#Background%20Details) (accessed: 06.17.2020).
5. Friction Stir Welding. In: *Wikipedia: the free encyclopedia* [Electronic resource]. URL: http://en.wikipedia.org/wiki/Friction_stir_welding (accessed: 05.11.2020).
6. Fusion Splicing. In: *Wikipedia: the free encyclopedia* [Electronic resource]. URL: http://en.wikipedia.org/wiki/Fusion_splicing (accessed: 05.11.2020).
7. History of MIG (GMAW) Welding [Electronic resource]. URL: [http://www.netwelding.com/History_MIG%201.htm#Atomic Hydrogen](http://www.netwelding.com/History_MIG%201.htm#Atomic%20Hydrogen) (accessed: 05.14.2020).
8. History of Welding — in the Beginning [Electronic resource]. URL: <http://www.netwelding.com/History%20of%20Welding.htm> (accessed: 05.11.2020).
9. Laser-Hybrid Welding. In: *Wikipedia: the free encyclopedia*. [Electronic resource]. URL: http://en.wikipedia.org/wiki/Laser-hybrid_welding (accessed: 05.12.2020).

10. Mackenzie, L.B. (1951). *The Welding Encyclopedia*, T.B. Jefferson (ed.). New York: McGraw-Hill.
11. Terms and Definitions (1969). AWS A 3.0-69, prep. By AWS Com. On definitions and symbols. Under the dir. of AWS Technical activities com. New York: American Welding Society.
12. Welding handbook (1976). In: *Fundamentals of welding*. Vol. 1. Miami: American Welding Society.

Информация об авторах:

Бутенко Юлия Ивановна, кандидат технических наук, доцент кафедры «Романо-германские языки» факультета «Лингвистика», Московский государственный технический университет им. Н.Э. Баумана; ассистент кафедры иностранных языков факультета гуманитарных и социальных наук, Российский университет дружбы народов; *сфера научных интересов*: компьютерная лингвистика, терминоведение, интеллектуальные информационные технологии; *e-mail*: iubutenko@bmstu.ru Scopus Author ID: 57212144571; ResearcherID: AAH-1049-2020

Николаева Наталья Сергеевна, кандидат филологических наук, доцент кафедры «Английский язык для машиностроительных специальностей» факультета «Лингвистика», Московский государственный технический университет им. Н.Э. Баумана; *сфера научных интересов*: сравнительно-историческое языкознание, терминоведение, структурная лингвистика; *e-mail*: natalynic@yandex.ru ResearcherID AAD-4868-2022

Карцева Елена Юрьевна, кандидат исторических наук, доцент кафедры иностранных языков факультета гуманитарных и социальных наук, Российский университет дружбы народов; *сфера научных интересов*: английский язык, переводоведение, терминоведение; *e-mail*: kartseva_eyu@rudn.ru Scopus Author ID: 57205196522; ResearcherID A-6432-2017

Information about the authors:

Iuliia I. Butenko, PhD, Candidate of Technology, Associate Professor of Roman-Germanic Languages Department, Bauman Moscow State Technical University; assistant professor of Foreign Languages Department, Peoples' Friendship University of Russia (RUDN University); *Research interests* computer linguistics, terminology science, intelligent information technologies; *e-mail*: iubutenko@bmstu.ru Scopus Author ID: 57212144571; ResearcherID: AAH-1049-2020

Nataliya S. Nikolaeva, PhD, Candidate of Philology, Associate Professor of English for Mechanical Engineering Department, Bauman Moscow State Technical University; *Research interests*: comparative linguistics, terminology science, structural linguistics; *e-mail*: natalynic@yandex.ru ResearcherID AAD-4868-2022

Elena U. Kartseva, PhD, Candidate of History, Associate Professor of Foreign Languages Department, Peoples' Friendship University of Russia (RUDN University); *research interests*: English language, translation studies, terminology science; *e-mail*: kartseva_eyu@rudn.ru Scopus Author ID: 57205196522; ResearcherID A-6432-2017