

Вестник Ивановского государственного университета.

Серия: Гуманитарные науки. 2024. Специальный выпуск. С. 131—137.

Ivanovo State University Bulletin. Series: Humanities. 2024. Special issue. P. 131—137.

Научная статья

УДК 81'373/374

DOI: 10.46726/H.2024.4.15

СОЗДАНИЕ ГЛОССАРИЕВ ФОРМИРУЮЩИХСЯ ТЕРМИНОЛОГИЙ: ОПЫТ ЛИНГВИСТИЧЕСКОЙ ЛАБОРАТОРИИ ФЛМК МГИМО

Валентина Александровна Иконникова,

Анастасия Дмитриевна Мезенцова

Одинцовский филиал МГИМО МИД России, г. Одинцово, Россия,

v.ikonnikova@odin.mgimo.ru, a.mezentsova@odin.mgimo.ru

Аннотация. На материале новых отраслевых русскоязычных и англоязычных терминологий на основе теории языкового субстрата и понятийного суперстрата терминологического знака в статье анализируются особенности употребления актуальной терминологии, не закрепленной в словарях, — «терминов текущей кризисной ситуации». Обозначаются признаки сформированных терминосистем и формирующихся терминологий. Авторы описывают процесс составления онлайн глоссариев, разработанных коллективом Лингвистической лаборатории ФЛМК МГИМО в 2022—2024 гг.: глоссариев юридических терминов санкционного регулирования и контрактного права, а также терминов нейро- и психолингвистики. Представлен опыт создания терминологической базы данных (ТБД), разработки параметров словарной статьи, методики отбора терминов и формулирования собственных дефиниций и вариантов перевода.

Ключевые слова: формирующиеся терминологии, теория терминологического знака, актуальные термины «текущей кризисной ситуации», цифровые онлайн глоссарии, терминологическая база данных

Для цитирования: Иконникова В.А., Мезенцова А.Д. Создание глоссариев формирующихся терминологий: опыт Лингвистической лаборатории ФЛМК МГИМО // Вестник Ивановского государственного университета. Серия: Гуманитарные науки. 2024. Специальный выпуск. С. 131—137.

Лингвистическая лаборатория факультета лингвистики и межкультурной коммуникации (ФЛМК) функционирует на базе кампуса МГИМО-Одинцово в рамках проекта «Цифра и инновации в образовании», который, в свою очередь, является частью федеральной программы академического лидерства «Приоритет 2030», реализуемой, в том числе, в МГИМО МИД России. В исследованиях Лингвистической лаборатории принимают активное участие преподаватели и студенты старших курсов бакалавриата и магистратуры ФЛМК, приглашенные эксперты в области терминографии, специального перевода, ИТ, обработки данных естественных языков (natural language processing) и т. д.

В 2022—2024 гг. Лингвистическая лаборатория ФЛМК активно работает над проектом «Корпусные исследования языка и речи». Целью данного проекта является повышение качества цифровых лингвистических инструментов. Результатом должна стать разработка уникального цифрового лингвистического инструмента для перевода и обработки терминов и профессиональных текстов, а также

© Иконникова В.А., Мезенцова А.Д., 2024

внедрение высокотехнологичных методик в процесс языковой подготовки и подготовки лингвистов, переводчиков, юристов-международников. В своих исследованиях терминоведы — сотрудники Лаборатории опираются в том числе на теорию В.М. Лейчика [Лейчик], оперирующую понятиями языкового субстрата и логического суперстрата терминологического знака.

Несмотря на то, что на современном лексикографическом рынке известно довольно много словарей, регистрирующих общественно-политическую лексику, речь идет в основном об одноязычных справочниках [Маник], переводных англо-русских и русско-английских словарей такой лексики, выпущенных после 2020 г., практически нет. Также довольно сложно рекомендовать, например, студентам, обучающимся по направлениям «Лингвистика» или «Юриспруденция» какой-либо один словарь актуальной специальной терминологии, меняющейся буквально на глазах.

Новые понятия обозначаются новыми лексическими единицами, в т. ч. терминологическими, образованными средствами своего языка или являющимися заимствованиями. Иногда новые понятия обозначаются уже существующими единицами, получающими новые значения (семантический способ терминообразования). Формирующиеся новые терминологии отличаются рядом признаков от устоявшихся терминосистем. Так, если для сложившихся областей знания и их терминосистем характерны а) наличие теории, изложенной в докторских и кандидатских диссертациях, монографиях и иных научных публикациях; б) наличие предмета и специфических методов исследования, то для формирующихся терминологий и подсистем свойственны а) наличие активно развивающихся, часто противоречивых и взаимоисключающих концепций; б) стихийно использующаяся терминология, а также значительное количество терминов, привлеченных из смежных областей знания. Тем не менее эти признаки делают новые тематические области знания и их терминологии все более интересными и важными для лингвистического и терминографического изучения.

Важно подчеркнуть, что общественно-политическая лексика, а также терминология юриспруденции, цифровых технологий, иных новых областей знания постоянно развиваются, а новые лексические единицы часто обозначают ту или иную «текущую кризисную ситуацию» [Иконникова, Цверкун]. При этом онлайн и бумажные словари обновляются иногда недостаточно часто, чтобы обеспечить переводчика информацией о подобной лексике — она еще не вошла, а возможно, никогда не войдет в академические словари, но тем не менее на данный момент является актуальной, частотной и нуждается в корректном переводе. По этой причине не только студенты, но и переводчики, журналисты, дипломаты обращаются порой к проектам Multitran или Reverso context, надежность которых иногда вызывает сомнения, поскольку переводы выполнены людьми разной квалификации и уровня знаний, а предлагаемые варианты переводных эквивалентов всегда контекстно обусловлены.

Кроме того, имеющимися печатными общими переводными словарями общественно-политической и юридической лексики неудобно или практически невозможно быстро воспользоваться в условиях переговоров, устного (последовательного и синхронного) перевода, в учебных целях. Нужны цифровые версии небольших по объему, мгновенно доступных в телефоне или на планшете тематических глоссариев с англоязычными контекстами употребления. Не столько важен перевод контекстов на русский язык (предполагается, что адресат-переводчик их поймет), сколько удобство и скорость, актуальность, встречаемость в СМИ в последние месяцы, в текущей ситуации, в условиях актуальной на текущий

момент международной повестки дня (предполагается, что терминологическая база данных доступна с телефона, планшета, ПК). В таком случае нет необходимости искать по нескольким (тем более бумажным) словарям специфичные, встречающиеся в контекстах, но по разным причинам не включенные в словари единицы; нет необходимости тратить время и собирать «разбросанные» по разным словарям / корпусам самые востребованные для переводчика данные об этих единицах.

Некоторые терминологические единицы не включаются в крупные словари, т. к., возможно, выйдут из употребления вместе с потерявшей актуальность тематикой (например, “yellow jackets” / «жёлтые жилеты»), но специалисты используют такие единицы, а следовательно, переводчику необходимо их переводить. Так, например, единица *substate*, включенная в глоссарии ФЛМК, не зафиксирована в переводных англо-русских словарях ни как существительное, ни как прилагательное; тем не менее эксперты-международники употребляют единицу «субгосударственный» [Артеев]. Следовательно, необходимость в онлайн глоссариях с максимально актуализированным словником, дефинициями, вариантами перевода, пометами и справкой возрастает постоянно, поскольку постоянно растет скорость информационного потока, и все больше ценится время, сэкономленное на обработке больших объемов текстовых данных.

Разработкой такого инновационного цифрового продукта занимаются лингвисты и программисты Лингвистической лаборатории ФЛМК. С 2022 г. ведется работа над созданием собственной онлайн-платформы, включающей мгновенно доступные для пользователя тематические глоссарии актуальной терминологии, прошедшие экспертную валидацию, а также инструменты автоматической разметки загружаемых текстовых файлов для поиска коллокаций, контекстов, дефиниций, антонимов, синонимов (где возможно) и другой важной для переводчика и юриста-международника информации.

Терминологическая база данных Лингвистической лаборатории ФЛМК МГИМО может стать улучшенным аналогом следующих существующих инструментов:

- 1) ресурсов по созданию собственных корпусов и программ памяти перевода, которые сейчас не функционируют на территории России и имеют несколько другие функции;
- 2) онлайн агрегаторов переводов, имеющих в сети и предоставляющих переводы разного качества, в том числе ошибочные, с неподходящими контекстами;
- 3) онлайн или бумажных словарей, которые не дают нужных значений (новых или специальных);
- 4) GPT-моделей, не функционирующих на территории России и допускающих ошибки в силу обучения на любых запросах (в т. ч. неверных).

Работа над созданием глоссариев проводилась в два этапа. На первом этапе (2022 г.) каждый участник команды исследователей (автор глоссария) выбрал себе микро-тему в рамках юридической терминологии (терминология организации и проведения референдумов, терминология антидопингового регулирования, терминология контрактного права, терминология международных договоров о государственных границах и др.). Затем автором был произведен сбор англоязычных текстов международных и национальных нормативно-правовых актов и комментариев к ним, а также практик международных организаций по выбранной теме, из которых был составлен корпус для исследования.

Далее, с помощью цифровых корпусных инструментов, автор формировал список лексических единиц (терминов и терминологических сочетаний,

специальной и иной тематической лексики), вошедших в пилотную версию глоссария. Ключевыми параметрами при составлении списка терминов и специальной лексики были:

- а) принцип относительной частотности в корпусе;
- б) важная роль в тексте конкретной тематики (ключевые понятия на основании экспертного мнения исследователя, принимавшего решение о важности лексической единицы после изучения правовых актов и медиатекстов по узкой теме глоссария);
- в) актуальность для целей перевода и обучения (обусловленная критериями а и б).

Предоставленный корпусным менеджером список частотных лексических единиц не переходил в глоссарий в «первозданном» виде, а дополнительно редактировался и валидировался исследователем. Так, например, в большинстве случаев самыми частотными единицами в сгенерированных корпусах программа выделяла артикли, предлоги или наименования государств (например, “India” в глоссарии «Договоры о государственных границах»). Однако очевидно, что с точки зрения подготовки переводчиков к работе с определенной темой, подобные, пусть и самые частотные, лексические единицы нерелевантны, и они в глоссарии включены не были.

Перевод осуществлялся при помощи специализированных словарей юридической и иной терминологии, словарей для общих целей, а также одноязычных толковых и переводных словарей с опорой на контекстное значение в рамках нормативно-правовых актов, загруженных в корпус. Были случаи, когда перевод выполнялся автором глоссариев самостоятельно, поскольку в словарях новое значение лексической единицы не было зафиксировано. Перевод выполнялся на основе имевшихся вариантов перевода в словарях, а также исходя из контекстов, представленных в корпусах, собранных из актуальных текстов, опубликованных за последние несколько месяцев.

Пилотные версии глоссариев, разработанные в 2022 г., были размещены на отечественной платформе Myefe¹ и представлены на Международном научно-практическом семинаре «Лингвокультурологические исследования дискурса медиадипломатии»² 6—7 октября 2022 г. и на международном научно-практическом семинаре «Язык. Культура. Перевод: лаборатория актуальных смыслов»³ 16—17 февраля 2023 г., проходивших в МГИМО-Одинцово.

В 2023 г. проводился второй этап работы. Были созданы корпуса медиатекстов, отобранных из англоязычных СМИ за 2022—2023 гг. В ходе обработки корпуса моделировался список терминов, иной специальной и тематической лексики. Далее проходило сопоставление полученного списка лексических единиц с аналогичным тематическим списком, разработанным в ходе работы проекта в 2022 г., проводившегося на основе нормативно-правовых актов. Из обоих списков отбирались совпадающие и наиболее частотные и актуальные лексические единицы для формирования глоссария.

Лексические единицы тематического корпуса фиксировались в сводной таблице со следующими колонками: термин, помета о сфере употребления,

¹ Myefe. URL: <https://myefe.ru/> (дата обращения: 11.06.2024).

² Семинар «Лингвокультурологические исследования дискурса медиадипломатии». URL: <https://clck.ru/3EPvnq> (дата обращения: 13.06.2024).

³ Международный семинар «Язык. Культура. Перевод: лаборатория актуальных смыслов». URL: <https://clck.ru/3EPvqE> (дата обращения: 13.06.2024).

региональность (амер. / брит.), дефиниция на английском языке, переводной эквивалент, синоним / антоним (если есть), коллокация, производные слова, пример употребления, ссылка на контекст в СМИ. В таблице также фиксировалась полисемия лексических единиц (при наличии) и разные коллокационные возможности их отдельных значений.

Всего было сформировано 10 глоссариев: «Международное санкционное регулирование / International sanctions» (Е.В. Глушко), «Практика организации и проведения референдумов / Referendum Glossary» (С.А. Маник), Договоры о государственных границах / Boundary treaties (А.Д. Мезенцова), Международное и национальное контрактное право / Contract law (А.С. Галюченко) и др.

Экспертизу глоссариев осуществили сотрудники Института русского языка им. В.В. Виноградова РАН. На основе комментариев, вопросов, рекомендаций и замечаний экспертов были уточнены и расширены параметры словарной статьи глоссариев, созданных Лингвистической лабораторией в 2022 г., доработаны методология исследования и функционал будущей цифровой платформы.

Параллельно с подготовкой глоссариев проходила разработка собственной онлайн-платформы — терминологической базы данных (ТБД), куда могут быть загружены подготовленные тематические корпуса и глоссарии к ним.

На данный момент завершен первый этап разработки платформы. Технические специалисты Лаборатории создали платформу для загрузки разработанных терминологами Лаборатории глоссариев, а также программу морфологической и синтаксической разметки загружаемых контекстов употребления терминов. Программа формирует синтаксическое древо предложения и распознает части речи слов. В загруженных файлах можно выполнять конкордантный поиск. Пилотная версия ТБД успешно прошла специализированную экспертизу.

Ценным опытом для сотрудников Лингвистической лаборатории ФЛМК стал заказ Центра нейролингвистических исследований когнитивного потенциала МГИМО на разработку пилотной версии (2023) и версии 1.0 (2024) Русскоязычного глоссария нейролингвистической терминологии. Для работы над данным проектом был проанализирован широкий текстовый и видео материал, посвященный проблематике нейро- и психолингвистики, нейропсихологии, биологии и анатомии, физиологии, медицины и т. д. Были разработаны две методики отбора терминологических единиц (В.А. Иконниковой и Е.В. Глушко), принципы дальнейшего отбора лексики для словника глоссария и особенности репрезентации продукта на платформе Muefe и в ТБД Лингвистической лаборатории (А.С. Галюченко). В ходе работы над русскоязычной терминологией нейролингвистики авторы пришли к следующим выводам: 1) русскоязычная терминология нейролингвистики находится в состоянии формирования; 2) заимствования терминологических единиц путем транслитерации и калькирования — один из способов пополнения терминологии нейролингвистики; вопрос об обоснованности и ассимиляции заимствований является открытым; 3) корпусные инструменты обработки текстов и терминов, а также методика Лингвистической лаборатории ФЛМК МГИМО позволяют сделать наиболее достоверную выборку терминов для глоссария; 4) концепция В.М. Лейчика о структуре терминологического знака и идеи о создании ТБД — одна из базовых теорий для современных терминологических исследований и практики.

Результаты работы Лингвистической лаборатории в 2023 г. были представлены на презентации пилотной версии цифровой терминологической базы данных Лингвистической лаборатории ФЛМК МГИМО в рамках XV Конвента

Российской ассоциации международных исследований (РАМИ)⁴, а также на Круглом столе⁵. На мероприятии присутствовали представители ДЛЮ МИД России, руководители бюро перевода крупных международных юридических компаний, юристы-международники, переводчики, преподаватели перевода, студенты старших курсов бакалавриата ФЛМК, а также магистранты программы «Перевод и современные технологии в лингвистике» ФЛМК. Участники и гости презентации проявили интерес к представленному цифровому лингвистическому продукту. Перспективы работы над терминологической базой данных были изложены на пленарном и секционных заседаниях, а также на дискуссионных площадках Международного научно-практического форума «Язык. Культура. Перевод: цифровые технологии и вербальная коммуникация»⁶ 16—17 февраля 2024 г. Завершение работы над ТБД планируется в 2024—2025 г. (версия 1.0, адаптированная под заказчика).

Проведенная работа позволяет констатировать следующие особенности исследования новых терминологий и параметры терминографической фиксации их единиц: 1) необходимость фиксации терминопорождения в онлайн режиме, обусловленная возрастающей скоростью возникновения и исчезновения новой специальной лексики; 2) важность качества работы по определению критериев и источников отбора новых терминов, определения степени их терминологичности (критерии частотности, релевантного контекста, корректности применения данных корпусных инструментов обработки текстов и критический анализ результатов); 3) необходимость уточнения новых требований к тематическим онлайн-гlossариям актуальной специальной лексики: особенности параметров словарной статьи, особенности онлайн платформ и терминологических баз данных как систем управления (в т. ч. новыми) терминологиями.

Список литературы / References

- Артеев С.П. Субгосударственные акторы в системе международных отношений // Вестник МГИМО-Университета. 2016. № 5 (50). С. 167—171.
(Arteev S.P. The third level of international relations: characteristics of sub-state actors, *MGIMO Review of International Relations*, 2016, no. 5 (50), pp. 167—171. — In Russ.)
- Иконникова В.А., Цверкун Ю.Б. Лингвокультурологические исследования англоязычных отраслевых терминологий и современные технологии в лингвистике. Монография. М.: Русайнс, 2021. 186 с.
(Ikonnikova V.A., Tsverkun Yu.B. Linguistic and cultural research on Anglo-American branch terminologies and modern technologies in linguistics, Moscow, 2021, 186 p. — In Russ.)
- Лейчик В.М. Терминоведение: Предмет, методы, структура. М.: КомКнига, 2006. 256 с.
(Lejchik V.M. Terminology Science: Object, methods, structure, Moscow, 2006, 256 p. — In Russ.)
- Маник С.А. Англоязычная политическая лексикография: формирование, развитие, современное состояние: дис. ... д-ра филол. наук. Иваново, 2019. 572 с.
(Manik S.A. English-language political lexicography: genesis, development, current state: dis. ... Doctor of Sciences (Philology), Ivanovo, 2019, 572 p. — In Russ.)

⁴ РАМИ: презентация пилотной версии цифровой терминологической базы данных МГИМО. URL: <https://mgimo.ru/about/news/departments/xv-risa-presentation-linguistic-laboratory/> (дата обращения: 13.06.2024).

⁵ Нейролингвистика, ИОТ и тенденции языковой политики. URL: <https://mgimo.ru/about/news/main/priorityfest-2023-lang/> (дата обращения: 13.06.2024).

⁶ Форум «Язык. Культура. Перевод: цифровые технологии и вербальная коммуникация». URL: <https://clck.ru/3EPvtZ> (дата обращения: 13.06.2024).

CREATING GLOSSARIES OF EMERGING TERMINOLOGY: THE EXPERIENCE OF THE MGIMO L&CCC LINGUISTIC LABORATORY

Valentina A. Ikonnikova, Anastasia D. Mezentsova

Moscow State Institute of International Relations (University) of the Ministry of Foreign Affairs of the Russian Federation, Odintsovo branch, Odintsovo, Russian Federation, v.ikonnikova@odin.mgimo.ru, a.mezentsova@odin.mgimo.ru

Abstract. The report studies characteristic features and contexts in which relevant “current crisis terminology” not fixed in dictionaries is used. The research is based on the new thematic Russian and English terminologies as well as on the theory of terminological sign. Peculiarities of well-formed terminological systems and developing terminologies are stated. The authors describe the process of compiling online glossaries created by the members of the Linguistic Laboratory that functions within the MGIMO University School of Linguistics and Cross-Cultural Communication — glossaries of law terms of sanctions regulation acts and contract law, terms of neuro- and psycholinguistics. The article shares the experience of creating a terminological data base (TDB), as well as working out lexicographical entry parameters, methods of terms selection and developing new definitions and variants of translation.

Keywords: developing terminologies, terminological sign theory, relevant term of the “current crisis”, digital online glossaries and terminological data base

For citation: Ikonnikova V.A., Mezentsova A.D. Creating glossaries of emerging terminology: the experience of the MGIMO L&CCC Linguistic Laboratory, *Ivanovo State University Bulletin, Series: Humanities*, 2024, Special issue, pp. 131—137.

Статья поступила в редакцию 14.06.2024; одобрена после рецензирования 26.06.2024; принята к публикации 29.06.2024.

The article was submitted 14.06.2024; approved after reviewing 26.06.2024; accepted for publication 29.06.2024.

Информация об авторах / Information about the authors

Иконникова Валентина Александровна — доктор филологических наук, декан факультета лингвистики и межкультурной коммуникации, Одинцовский филиал МГИМО МИД России, г. Одинцово, Россия, v.ikonnikova@odin.mgimo.ru

Ikonnikova Valentina Alexandrovna — Doctor of Sciences (Philology), Dean of the School of Linguistics and Cross-Cultural Communication (L&CCC), Moscow State Institute of International Relations (University) of the Ministry of Foreign Affairs of the Russian Federation, Odintsovo branch, Odintsovo, Russian Federation, v.ikonnikova@odin.mgimo.ru

Мезенцова Анастасия Дмитриевна — специалист по учебно-методической работе деканата факультета лингвистики и межкультурной коммуникации, преподаватель кафедры лингвистики и переводоведения, Одинцовский филиал МГИМО МИД России, г. Одинцово, Россия, a.mezentsova@odin.mgimo.ru

Mezentsova Anastasia Dmitrievna — specialist in educational and methodological work of the School of Linguistics & Cross-Cultural Communication Dean’s Office, teacher at the Department of Linguistics and Interpretation & Translation Studies, Moscow State Institute of International Relations (University) of the Ministry of Foreign Affairs of the Russian Federation, Odintsovo branch, Odintsovo, Russian Federation, a.mezentsova@odin.mgimo.ru