

**Современная лексикография, дополненная лингвистическая реальность  
и проблемы терминографии**

**Н.К. Рябцева (Институт языкознания РАН)**

**Аннотация**

В статье показано, что под воздействием стремительно развивающихся цифровых технологий современная лексикография переживает кардинальные преобразования лексикографических принципов представления лингвистической информации в различных информационных и интеллектуальных системах и приложениях. Выявлены новые активные тенденции в эволюции как самой лексикографии, так и в формировании и функционировании различных инновационных лексикографических ресурсов. Особое внимание в статье уделяется самой актуальной проблеме в области лингвистической поддержки межъязыковой научной коммуникации — лексикографическому представлению современной терминологии, которая под воздействием цифровых технологий приобретает принципиально новые свойства: динамичность, многокомпонентность, контекстуальность, междисциплинарность и др.

**Ключевые слова**

Цифровые технологии, развитие лексикографических ресурсов, автоматическое лексикографирование, терминологические кластеры, межъязыковая асимметрия.

**1. Вводные замечания**

С начала двухтысячных годов и по настоящее время лексикография как главный способ представления лингвистических знаний (наравне с грамматикой), претерпевает кардинальные изменения. Они вызваны стремительно развивающимися цифровыми технологиями, которые все более активно используются в лексикографии и порождают в ней принципиально новое направление — электронная лексикография, *electronic*

*lexicography, e-lexicography* [Abel, Bracco 2006; Granger, Paquot 2010; Tarp 2021] (здесь и далее *курсивом* даны составные лингвистические термины — Н. Р.). Последняя, в свою очередь, занимается созданием не только обычных и электронных словарей (*e-dictionaries*), но и играет все более заметную роль в создании и совершенствовании различных систем класса «искусственный интеллект» (ИИ), в их лингвистическом обеспечении, расширении их применения и взаимодействия со смежными ресурсами и мн. др. [Kilgarriff 2009; Abel 2010; Fuertes-Olivera et al. 2018; Tarp 2022].

Важнейшим направлением стремительной эволюции самой лексикографии выступает создание лексикографических ресурсов принципиально нового «активного» типа: электронных словарей, лексикографических баз данных, предметных лексиконов, онтологий и мн. др., отличительной чертой которых выступает многофункциональность, многоканальность доступа (*multiple access routes*), интеграция общих, специальных и фразеологических словарей (*efficient integration of phraseology and specialised dictionaries*) и мн. др. Основой вызов здесь связан с необходимостью разработки принципиально нового формата представления лексикографических данных, обусловленного инновационными возможностями цифровых технологий (*innovative changes to the dictionary structure afforded by the electronic medium*), которые призваны расширять принципы обращения к электронным лингвистическим ресурсам, учитывать различные интересы, цели и задачи различных пользователей, автоматизировать доступ к ресурсам и настройку на конкретного пользователя, кастомизацию (*automating customisation of dictionaries in function of users' needs*) и т.д. [Tarp 2019].

Так, традиционный лексикографический принцип представления лингвистической информации — словарная статья, по целому ряду причин не может быть использована в неизменном виде при формировании электронных лексикографических ресурсов и лингвистического обеспечения в системах класса ИИ [Huang, Tarp 2021]. Центральным направлением развития электронной лексикографии выступает ее интеграция в инновационные лингвистические ресурсы: автоматизированные системы обучения

чтению, письму, родному и иностранному языку, переводу, коммуникации и пр. (*language learning and teaching tools*) [Granger, Paquot 2010; Wachter et al. 2014; Fuertes-Olivera, Tarp 2020].

В настоящее время лексикография с уверенностью вошла в эпоху интернета, значительно обогатив лексикографические описания за счет использования интернет-приложений и получив возможность интеграции с ними в новые эффективные ресурсы класса *Augmented Writing* (AW) и др., обеспечивающие помощь в чтении, переводе и написании текстов разного типа, вида и содержания [Abel 2010, 3; Abel, Bracco 2006; Simonsen 2020; Tarp 2020; Huang, Tarp 2021]. Происходящие сдвиги в новой лексикографической практике позволяют говорить о новом содержании понятия «словарь», который становится «электронным ассистентом», (*electronic*) «e—Assistant», («e—Ассистент»). («Lexicography has definitively entered the internet age. Applying *web database applications* in a context of *enriched lexicographic descriptions* has opened new perspectives for the integration of resources and the development of new and efficient *reading, translation and writing assistants* and has in fact *reshaped the very definition of 'dictionary'*») [Verlinde 2010, 341].

## **2. Лексикография и e-Ассистенты «Дополненное письмо»: *Augmented Writing Assistants***

В процессе развития цифровых технологий и систем искусственного интеллекта наравне с ресурсами *дополненная реальность* развиваются процессоры класса *дополненная лингвистическая реальность*. Они воплощаются в настоящее время в целом ряде инновационных программ и приложений, которые объединяются уже устоявшимся в английском языке понятием *Augmented Writing* (AW), образованным по образу и подобию хорошо известного термина *Augmented Reality* (AR), которому в русском языке соответствует название *дополненная реальность*. «Дополненная» в том смысле, что обычные (исходно — перцептивные) возможности пользователя с их помощью дополняются возможностями цифровых технологий [Riabtseva 2016; 2017]. *Дополненная*

лингвистическая реальность основана на приложениях / ресурсах типа *Writing Assistants* (WA, *Текстовый ассистент*) и способна обеспечивать значительное расширение лингвистических возможностей пользователя в процессе порождения (написания, перевода, редактирования и т. д.) текста на естественном языке, родном или иностранном. Главной задачей развития систем класса *дополненная лингвистическая реальность* — *Writing Assistants* (WA) и *Augmented Writing* (AW) выступает дальнейшее совершенствование их лингвистического обеспечения, которое, несмотря на его постоянную эволюцию, все еще оставляет желать лучшего в связи с тем, что автоматическая обработка естественного языка, ввиду чрезвычайной сложности его устройства, пока еще значительно ограничена всего несколькими возможностями, которые, правда, постепенно и постоянно расширяются, в первую очередь, благодаря совершенствованию технологий *Text Production Technologies* (TPT) («Lexicography has an important role to play in the development of new advanced Text Production Technologies and the lexicographic augmentation of AW could be an important step in the right direction») [Simonsen 2020, 513].

В настоящее время приложения класса WA и AW развиваются в следующих направлениях: 1) Приложения по проверке правописания и грамматики: *Spellcheckers* и *Grammar checkers*, например, *Grammarly* и *WhiteSmoke*. 2) Текстопорождающие приложения типа *TalktoTransformer* или *Articoolo*, позволяющие пользователю порождать текст на основе ключевых слов. 3) Приложения типа «L2 writing assistants», оказывающие помощь в порождении текста на иностранном языке, например, *Text Assistant* [Tarp et al. 2017]. 4) Приложения по проверке тональности типа *Persado* и *MessagePath*, интегрированные в рабочее место специалиста по редактированию рекламных, маркетинговых и т.п. текстов. 5) Приложения по анализу специальных текстов типа *Textio* [Marconi], помогающие анализировать тексты конкурсных заявлений о приеме на работу, составлять тексты объявлений о вакансиях и др. [Simonsen 2020, 509—514].

Первостепенную роль в развитии технологий WA и AW сыграли образовательные системы

класса *Computer-Assisted Language Learning* (CALL). Их центральным компонентом являются технологии искусственного интеллекта (ИИ), которые заимствуются в AW-технологии в виде алгоритмов порождения текста и автоматического исправления грамматических ошибок. В настоящее время *CALL-applications* развиваются в нескольких направлениях: *CELL* (Computer-Enhanced Language Learning), *TELL* (Technology-Enhanced Language Learning), *ICALL* (Intelligent Computer-Assisted Instruction), *NCALL* (Network based CALL) [Abel 2010: 3], *Web Assisted Language Learning* [WALL] и др. Активно разрабатываются также разнообразные экспертные приложения типа *Expert writing support* и соответствующие ресурсы общего, специального и узкого назначения. По способу обращения и по удобству использования они значительно превосходят традиционные бумажные словари и справочники. К ним относятся такие приложения как *Word Combinatory assistant*, *Automatic Lexical Error Detection*, *Lexical Error Correction assistant*, *Translation Memory tools (TMT)*, *Context-aware dictionaries*, *Web-based dictionaries*, *Process-oriented Writing support*, *Writing Aid*, *Text Assistant*, *Write Assistant*, [Write Assistant], *Wordfast* [Wordfast], *Computer-based Writing Instructions for text producers and learners*, *Augmented Writing Platform*, *Collaborative Writing*, *Common Business Writing Mistakes* и мн. др. [Banks 2019; Simonsen 2021]. Особые приложения класса *Augmented Writing* и *Writing Assistants* (AW и WA) разрабатываются также для различных специалистов-текстовиков (*text producers*): репортеров, обозревателей, журналистов, переводчиков и т. п., особенно работающих в коллективе или в команде (*Computer-supported Collaborative Writing*). Важное место они занимают в приложениях по обучению, развитию и практике научной речи на (иностранном) английском языке — *L2 Academic Writing Assistants*. Так, ресурсы типа *ColloCaid* оказывают поддержку в написании научных текстов на иностранном английском языке: «*ColloCaid is Writing Assistant designed to help English for Academic Purposes (EAP) writers with collocations*» [Frankenberg-Garcia et al. 2021].

Создаваемая таким образом *Дополненная лингвистическая реальность* класса *Augmented Writing* и *Writing Assistants* (AW и WA) позволяет пользователям более активно

и свободно порождать тексты, используя автоматические подсказки самого разного типа — от правописания и грамматики до сочетаемости и терминологии. Так что современные приложения типа (*L2*) *Writing Assistants* (WA), *Context-aware Dictionaries* и др. становятся все более популярными среди специалистов [Tarp et al. 2017; Tarp et al. 2019], особенно благодаря своей способности не просто предоставлять имеющуюся информацию, но еще и подсказывать, предсказывать или корректировать варианты развертывания текста. Их дальнейшее развитие, в свою очередь, влечет кардинальные преобразования в области лексикографии вообще и двуязычной лексикографии в частности — в виде их «переворачивания с ног на голову», «turning lexicography upside down» [Simonsen 2020, 512] и «turning *bilingual lexicography* upside down» [Tarp 2022, 67], благодаря созданию систем, способных выполнять *автоматическое лексикографирование*: разделять выполнение лексикографических задач между человеком и машиной, создавать лексикографические продукты нового типа — лексикографические базы данных, а также использовать технологии ИИ по оперированию знаниями [Schryver 2012, 130], тем самым активно продвигаясь в направлении DIKIW: *Data, Information, Knowledge, Intelligence, Wisdom* [Liew 2013].

Для этого необходимы сведения принципиально нового типа: знания о мире (*world knowledge*), реляционные знания (*relational knowledge*) и контекстуально связанная лексикографическая информация (*curated lexicographic data*) [Simonsen 2020, 512]. Поскольку современные технологии AW и WA пока недостаточно учитывают контекст, то именно автоматический сбор соответствующих контекстуальных данных может предоставить необходимую для этого лексикографическую информацию в виде знаний о мире и реляционных знаний. Они размещаются в специальном лексикографическом корпусе, обращение к которому осуществляется при автоматическом поиске текстовых ресурсов, содержащих нужную информацию и соответствующий ей контекст. Для получения более качественного результата по использованию контекстуальным и лексикографическим образом представленных знаний о мире, в системы AW и WA

вводится специальное вспомогательное окно для пост-редактирования выходного текста («*an auxiliary post-editing window providing help in post-editing the output text*») [Simonsen 2020, 513].

Получаемая таким образом лексикографическая аугментация технологий AW и WA достигается путем разделения труда между машиной и редактором, разграничения понятий «данные — информация — знания — интеллект» («*data, information, knowledge and intelligence*»), а также использования *Открытой III-модели понимания естественного языка (OpenAI's Natural Language Understanding (NLU) model)* [OpenAi 2019], которая была разработана на большом массиве текстов и хорошо зарекомендовала себя в алгоритмах предсказания развертывания текста [Simonsen 2020]. На данный момент важнейшим средством повышения качества лексикографических баз данных является их пополнение за счет обработки интернет-текстов WebText большого объема (*a dataset of millions of webpages*). Отмечается также, что лексикография может способствовать повышению качества приложений AW и WA именно за счет концентрированного лексикографического контекстуального представления знаний о мире. При этом предлагается часть работы переложить на автоматическое получение и представление информации [Colson 2019]. Построение таким образом алгоритмов лексикографического продвижения технологий класса AW и WA и создание систем автоматического (искусственного) лексикографирования (*e-Lexicographer, Artificial Lexicographer*) может значительно повысить их качество [Simonsen 2020, 513; Tarp 2022, 66].

Таким образом, развивающиеся технологии AW и WA стимулируют дальнейшее совершенствование лексикографических принципов представления лингвистической и экстралингвистической информации и знаний, вызывая симбиоз цифровых технологий и лингвистики и способствуя созданию лексикографических ресурсов нового типа — лексикографических баз данных, а также интегральных и инкорпорированных словарей [Simonsen 2021; Tarp 2022].

### **3. Отличие словаря от лексикографической базы данных. Интегральные и инкорпорированные словари**

Основным средством представления лексикографической информации в традиционном бумажном словаре является словарная статья, которая имеет свою организацию, последовательность и логику представления информации, особо структурированное содержание, дополнительный аппарат: ссылки, сокращения, обозначения и т.п. Основным способом обращения к словарной статье выступает знакомство с ее содержанием и (сплошной) поиск в ней нужной (запрашиваемой) лингвистической информации: значение слова, производные, сочетаемость, устойчивые словосочетания, перевод на другой язык, способы написания и т.п. Основным средством предоставления лексикографической информации в цифровых информационных ресурсах является всплывающее окно, которое появляется, когда пользователь выделил элемент текста, который, например, ему нужно перевести на другой (родной или иностранный) язык.

Всплывающее окно не может содержать всю лексикографическую информацию, относящуюся к выделенному слову. Есть несколько вариантов ее сокращения: давать только самую главную, основную и наиболее часто требующуюся, т. е. первостепенную по значимости информацию; дать сокращенный вариант каждого пункта словарной статьи; т.е. представить реферат словарной статьи и т. п. Так, известный лексикограф С. Тарп в свое время подчеркивал, что современный высоко операциональный лексикографический онлайн-справочник должен представлять собой поисковый инструмент, обеспечивающий онлайн-поиск по лексической базе данных и интернету — *leximat*. Его задачей является предоставление лексикографической информации в зависимости от запроса пользователя («*high-performance lexicographic reference tool should be a kind of search engine enabling queries on a lexical database or on the web. This tool, called leximat, should allow any user with a particular communicative or cognitive need to access lexicographic data*») [Тарп 2008, 123].



Современные лексикографы создают не только и не столько словари, сколько лексикографические базы данных, которые контекстно, коммуникативно (стилистически) и когнитивно связаны, и могут быть использованы для создания и совершенствования как традиционных и электронных словарей, так и высокотехнологичных лингвистических приложений. В результате влияние современных цифровых технологий на лексикографическую практику носит кардинальный характер и преобразует ее всю до основания: бумажные словари преобразуются в цифровые интерактивные многофункциональные лексикографические базы данных [Pajzs 2009], обращение к словарной информации проходит без выхода из основного рода коммуникативной деятельности, а предъявление лексикографической информации все более ориентируется на контекстуально-связанное и когнитивно-ориентированное представление. В результате происходящие в лексикографии фундаментальные преобразования ставят перед разработчиками лексикографического обеспечения различных лингвистических процессоров принципиально новые научные, лингвистические, лексикографические, коммуникативные и когнитивные задачи. Главные из них — обеспечить наличие в системе как можно более полного, можно сказать, максимального объема лингвистической информации — с одной стороны, и обеспечить ее организацию таким образом, чтобы на любой запрос пользователя выдавался максимально сокращенный, минимальный вариант, причем максимально релевантной для данного случая информации.

В настоящее время электронные словари и лексикографические базы данных разбиваются на ориентированные на человека или машину (*human-oriented vs. machine-oriented*), ориентированные на тип использующего его приложения: предназначенные преимущественно или опосредованно для приложений класса CALL, AW и WA, а также «лексикографически ориентированные» (*primarily vs. secondarily CALL-, AW-, WA-oriented or lexicography-oriented applications*). Кроме того, все типы словарей могут совмещать в себе содержание и функции друг друга и тем самым становятся многофункциональными («This can be interpreted as *a new facet of the multifunctionality of a dictionary*») (Abel 2010, 6;

Kilgarriff 2009; Simonsen 2021) и интегральными [Huang, Tarp 2021, 69]. Современные цифровые технологии развиваются стремительно и предъявляют все новые требования к лингвистическому обеспечению, в первую очередь, лексикографического типа. В настоящий момент соответствующая тенденция может быть охарактеризована как тройственное преобразование автономных словарей, печатных и онлайн, в интегральные, стандартизованных — в более персонифицированные, а собственно словари преобразуются в лексикографические базы данных для различных целей (the overall transition can be further specified as *a triple transformation from the traditional stand-alone dictionary, either printed or online, to the integrated dictionary; from the standardized dictionary to a more personalized dictionary; and from the dictionary as such to lexicographic data for different uses*) [Huang, Tarp 2021, 69]. Так, интеграция словарей в цифровые продукты *E-readers, Writing Assistants, Translation Tools, Learning Apps* описана в [Tarp et al. 2017; Loewen et al. 2019].

В зависимости от режима обращения словари могут быть встроенными (*embedded*), обращение к которым требует выхода из программы, и инкорпорированными / интегрированными (*integrated*), не требующими выхода (that are *activated by clicking on a word in the course text*). Последние более перспективны, поскольку способны реагировать на текущие индивидуальные (*personalized*) потребности пользователя. Если встроенные словари не «понимают», какую именно информацию пользователь запрашивает, то инкорпорированные, напротив, обладают свойством контекстуальности (*context-aware*), т. е. «понимают», какое значение слова из запрашиваемых является релевантным в данном контексте [Huang, Tarp 2021, 71—74]. Таким образом, преимущества инкорпорированных словарей заключаются в их персонифицированности, эффективности и оперативности, а их качество зависит от объема содержащейся в них лексикографической базы данных, а также от дизайна пользовательского интерфейса.

Лексикографическая база данных и интерфейс тесно между собой связаны. Если один из них работает плохо, то пользователь в любом случае останется недоволен работой

приложения. Так, наибольшие проблемы вызывает подача информации по следующим вопросам: выбор словарной единицы (*lemma selection*): слово, словосочетание, сложное слово (*single words, multiword units, compound words*), подача многозначных слов (*treatment of polysemous words*), описание грамматических категорий; переводные эквиваленты (*translation equivalents*); наличие избыточной информации (*data overload*) или отсутствие необходимой (*data underload*); позиция всплывающего окна (*position of the pop-up window*).

Самым сложным вопросом для современных приложений является подача словосочетаний, для которых пока еще не выработан последовательный и дружественный алгоритм предъявления [Huang, Tarp 2021, 77]. При этом часто неудовлетворительная работа лексикографической базы данных объясняется низким качеством лексикографического описания именно словосочетаний разных типов и видов и способов обращения к ним (*a low database quality originating from a poor compilation or storage of the lexicographic data that are presented in the pop-up window on demand*) [Huang, Tarp 2021, 78].

Лексикографическая база данных может содержать излишнюю информацию, но может и не содержать необходимую (*data overload and data underload*) [Huang, Tarp 2021, 88], давать устаревшие или неточные дефиниции (*inadequate definitions and old-fashioned features*), неточные или неправильные переводы и др. В результате компьютерные словари могут не только способствовать усвоению материала, но и препятствовать этому разными «способами» (*All these problems show that the dictionaries not only assist the learners with lexicographic data. They also create all sorts of obstacles*) [Huang, Tarp 2021, 86]. Так, инкорпорированные словари часто оказываются или недостаточно полными, или содержащими излишне подробную информацию. Не все лексические единицы, которые встречаются в текстах обучающего приложения, в нем содержатся, так же как и не все необходимые словосочетания. Значения многозначного слова могут быть представлены не в полном объеме, или доступ к ним затруднен из-за перегрузки системы избыточной информацией. Соответствующие недостатки должны быть устранены в рабочем порядке,

поэтому основным требованием к инкорпорированным словарям является их динамичный характер — постоянное обновление и улучшение.

Популярность цифровых приложений по обучению иностранному языку активно растет во всем мире. Обычно они содержат инкорпорированные одноязычные и двуязычные словари. Их основным недостатком пока является недостаточно полное описание лексики вообще и многозначных слов и устойчивых словосочетаний в частности. Их совершенствование, как считают специалисты, может быть основано на «контекстуальном подходе» к лексикографическому представлению соответствующих явлений (*context-adapted approach*), точнее, содержания всплывающих в процессе чтения или написания текста окон. Как уже отмечалось, в настоящее время они или содержат избыточную информацию, или недостаточную. И в том, и в другом случае пользователь ставится в затруднительное положение. Выход из него заключается в том, что всплывающее в процессе работы пользователя окно с подсказкой должно содержать перечень вариантов для дальнейшего поиска, а не всю соответствующую запросу, часто избыточную, информацию. [Fuertes-Olivera, Tarp 2014, 64; 2020]. (The ideal online dictionary should contain “as much data as possible in terms of all possible consultations”, whereas its individual articles “should include as little data as possible” *to avoid information overload*. From this perspective, a personalized service requires that the dictionary only presents the exact “types and amount of data”, which the individual user needs in each consultation, “neither more nor less”) [Fuertes-Olivera, Tarp 2014, 64]. Эти и смежные задачи специалисты пытаются решить с помощью принципиально новых лингвистических технологий обработки естественного языка — создания лингвистических процессоров класса e-Lexicographer (Artificial Lexicographer).

#### 4. Электронный лексикограф *Artificial Lexicographer*: задачи и перспективы

Технологическая революция в индустрии производства и использования знаний, вызванная стремительным развитием цифровых технологий и их применением в системах класса *искусственный интеллект*, особенно связанных с обработкой естественного языка,

вызывает, в свою очередь, кардинальные изменения почти во всех областях знаний и обработки информации, а такая область, как лексикография преобразуется в производящую информацию благодаря развитию интеллектуальных систем под названием *Artificial Lexicographer* (e-Lexicographer).

Современные цифровые лингвистические технологии класса *искусственный интеллект*, типа машинного перевода (МП), *Machine-aided Translation*, *Write Assistant* (WA), *Augmented Writing* (AW) и т. п., развиваются ускоренными темпами и воздействуют на принципы функционирования лингвистического обеспечения: требуют его постоянного обновления и развития, причем сразу в нескольких направлениях [Тарп 2019]. При этом наиболее активные трансформации в принципах и способах представления лингвистической информации происходят в лексикографии. Так, традиционная, бумажная лексикография представляет свои результаты в виде традиционных словарей, в которых, как правило, словарная статья одного слова, последовательно описывающая все его, скажем, 10 или 15 значений, может занимать четыре печатных страницы и более. Компьютерная лексикография, напротив, вынуждена подавать свою информацию не последовательно и полно, а так, чтобы она отвечала на стоящий в данный момент перед пользователем вопрос. Например, при последовательном переводе на английский язык (с любого другого языка) фразы «Он попросил меня закрыть (дверь)» автоматическая система может предложить в качестве эквивалента «закрыть» следующие варианты (в специальном окне — *in the suggestion window*): *close, shut, lock, seal* и др. Если пользователь не знает последнего слова, то он обращается за консультацией — кликает на это слово и тем самым активирует консультативное окно (*activates the consultation window*). Принципы организации информации в данном окне представляют собой наибольшую сложность при моделировании потребностей пользователя, который не будет читать словарную статью в несколько десятков строк, как в бумажном словаре. Поэтому здесь должен действовать принцип «первой выдается наиболее релевантная информация в данной ситуации» (*The*

*most relevant data are provided first*). Для ее определения необходимы специальные исследования потребностей пользователей конкретными лингвистическими ресурсами.

В результате происходящих в компьютерной лексикографии кардинальных преобразований в формулировке задач и проблем в ней развивается капитальная революция лексикографических основ (*a total revolution of the lexicographic discipline in all its major dimensions*) [Tarp 2019, 229]. Она проявляется, в том числе, и в экономическом отношении: если спрос на традиционные бумажные словари значительно снизился, вызвав нерентабельность их выпуска и обновления, то спрос на (бесплатные) цифровые лексикографические ресурсы резко увеличился, вызвав потребность в их коренной перестройке, особенно для стремительно развивающихся приложений класса *искусственный интеллект*. На данном этапе возникла, как уже указывалось, самая революционная цифровая идея в области лингвистического лексикографического обеспечения подобных интеллектуальных ресурсов — создание процессора под названием *Artificial Lexicographer*, по образу и подобию систем *искусственный интеллект* (Artificial Intelligence). В отличие от человека-лексикографа (*human lexicographer*), искусственный лексикограф («*artificial lexicographer*») не предлагает прочитать всю словарную статью к данному слову, а выбирает наиболее вероятный и нужный в данный момент вариант лексикографической информации и помещает его в окно *Подсказка*.

##### **5. Инновации в лексикографии: цифровизация лексикографических подсказок и разработка процессоров Lexicographic Assistants**

Использование современных технологий в лексикографической практике позволило значительно повысить качество, объем, содержание, возможности и даже цели современных электронных словарей. В настоящее время происходит их переориентация из лексикографических продуктов в лексикографические службы, платформы, инструменты и т. д. (*developing lexicographic services instead of lexicographic products* [Fuertes-Olivera 2019, 25], *as well as platforms, tools, etc.* [Tarp 2019, 231]. Их дальнейшее качественное развитие требует взаимодействия специалистов разных направлений, как гуманитарных,

так и технических. Основными направлениями развития онлайн-лексикографии становятся качественная с профессиональной точки зрения интеграция различных лексикографических источников в единое целое и как можно более точная ориентация на пользователя. «Время доморощенных баз данных закончилось» (*The days with homemade databases are long over*) [Tarp 2019, 232], нужны конкурентоспособные услуги и службы, а также их индустриализация [Fuertes-Olivera 2019].

Со стороны собственно лексикографии осуществляется переход от статичности традиционных словарей к динамике словарей электронных, от неопределенного пользователя к ориентации на пользователя и адаптации к его задачам, от неизменного вида лексикографической информации к изменяемому. Плюс еще введение гиперссылок, различных режимов просмотра содержания, его визуализации и т. п. (*The former static lexicographic structure has increasingly been replaced by a dynamic structure that adapts to different user needs in different types of consultation. This implies that the amount and organisation of the displayed lexicographic data are fluctuating. And to this should be added that many dictionary articles have been broken up and require clicks, scrolling down and other techniques to be visualized in their totality*) [Tarp 2019, 233]. Создаваемые таким образом лексикографические ассистенты (*Lexicographic Assistants*) становятся доступными на любых электронных устройствах и в различных лингвистических приложениях. Главное, что принципы организации лексикографической информации в них значительно отличаются от традиционных. Так, лексикографическое обеспечение систем типа *Write Assistant* (WA) и *Augmented Writing* (AW) значительно отличается от обычного, словарного оформления, в котором, с одной стороны, много лишнего (на данном шаге его использования), а с другой — нет принципиально важного, того, что нужно онлайн-пользователю в данный момент времени. И несмотря на то, что электронные словари чаще всего составляются с использованием лингвистической информации, импортированной из нескольких разных словарей — обычно от двух до шести — их все равно оказывается недостаточно для выполнения текущей задачи [Fuertes-Olivera et al. 2018]. Поэтому,

несмотря на возможность интеграции в одном онлайн-словаре информации из различных источников, необходимы сведения о том, какую именно информацию хочет получить пользователь на конкретном шаге использования «умной» системы.

Важнейшей задачей в создании интеллектуальных систем класса *дополненное письмо*, типа WA и AW, выступает выбор способа подачи материала: как выбрать из четырех страниц текста словарной статьи, содержащей описание одной лексикографической единицы, именно ту информацию, которая нужна пользователю сейчас и которую следует поместить первой в окно *Подсказка* на данном этапе? Как определить ее релевантность в данном контексте? Первым шагом на данном пути выступает максимальное редуцирование лексикографической информации, выдаваемой на первом шаге поддержки, до подсказки в одну строку (*short one-line definitions*), и ее расширение, если необходима дальнейшая помощь, на втором шаге и тем самым разделение обязанностей между окном-подсказкой (*suggestion window*) и окном-консультантом (*consultation window*) [Bothma, Tarp 2014; Tarp 2019, 242]. Следующим этапом является контекстуализация и персонализация лексикографической информации [Tarp, Gouws 2019]. Она производится посредством установления отношения пользователей к принципам представления информации в окне-подсказке и окне-консультанте и (*a test-driven development of the lexicographic product*).

Новые задачи, которые ставят перед лексикографией новые технологии, в частности, в виде систем дополненного письма, не могут быть решены традиционными лексикографическими методами анализа потребностей пользователя. Основным способом их решения выступают сами цифровые технологии, которые способны фиксировать действия пользователя и его информационные потребности. В результате лексикографы, всегда занимавшиеся составлением словарей, в настоящее время поняли, что теперь, в связи с развитием цифровых технологий, их практическая деятельность принципиально изменяется [Tarp 2019, 231]. Если на заре своего рождения лексикография создала два принципиально новых аппарата описания языка — словарную статью и словарный формат



— алфавитную организацию словарных статей, то в настоящее время возникли принципиально новые цифровые формы презентации лексикографических продуктов (*new digital forms of presentation of the lexicographic products*). Словарь перестал быть последовательностью словарных статей, а словарные статьи предъявляются пользователю не в обычном последовательном, статичном и полном виде, а динамично и направленно, в зависимости от запроса пользователя. Происходит естественная и необходимая адаптация лексикографических данных к новому цифровому пространству (*a natural and necessary lexicographic adaptation to the new digital environment*) [Tarp 2019, 231].

Главные особенности этого пространства заключаются в том, что пользователи, активно используя различные гаджеты для личной, профессиональной и другой коммуникативной деятельности (переписки, составления документов и других текстов, их перевода и т. д.), нуждаются в орфографической, межъязыковой и др. поддержке, желательно в виде онлайн-подсказок. Они обычно размещаются в специальном окне и автоматически генерируются встроенной в ресурс моделью языка и соответствующим лингвистическим обеспечением (*suggestions, generated automatically by the language model*). Кроме того, соответствующий ресурс, типа WA и AW, может подсказывать иноязычные эквиваленты, которые извлекаются из лексикографической базы данных (*lexicographic database*) или веб—словарей (*web-based dictionaries*). Возможны еще и другие виды лингвистических подсказок, рекомендаций и контекстуальных предложений. Заполнение соответствующих окон, подразумевающих различные виды лингвистической/лексикографической информации, частично может быть генерировано автоматически/алгоритмически, особенно относительно лексических и грамматических контекстов. Так, задача по упорядочению приоритетов в очередности подачи рекомендованных вариантов продолжения текста или его перевода в зависимости от контекста (*words listed in a prioritized, context-aware order*) вполне может оказаться под силу лингвистическому процессору класса Artificial Lexicographer и Lexicographic Assistant [Tarp 2019, 236].

Исходный вопрос при этом заключается в том, чтобы точно определить, какого рода лингвистическую информацию Artificial Lexicographer и Lexicographic Assistant может дать по умолчанию и по запросу, и каким образом ее лучше представить количественно и качественно в соответствующем пользовательском окне (*without data overload and in no user-unfriendly way*) и в каком порядке [Tarp 2019, 238]. Важным операциональным принципом при этом выступает возможность предоставления последующей информации (*additional lexicographic data*) посредством метатекстовых комментариев/гиперссылок типа «синонимы», «словосочетания» и т.п. (*with metatexts for further access: synonyms, collocations, etc.*). Такая контекстуализация лексикографической информации происходит в процессе апробирования различных итеративных форматов представления данных пользователю (*the use of iterative processes with test-driven development (TDD)*). Автоматическое отслеживание обращений пользователя к различным видам лексикографической информации (*lexicographic user research*) позволяет определить пользовательские приоритеты и оперативно вносить коррективы в порядок предоставления различных подсказок в соответствующий процессор Artificial Lexicographer и Lexicographic Assistant [Tarp 2019, 240—244].

Опыт функционирования приложений WA, AW и т. п. в целом показывает, что трансформация лексикографических принципов представления информации за счет разработки процедур искусственного (автоматического) лексикографирования Artificial Lexicographer и Lexicographic Assistant требует от лексикографов решения принципиально новых лексикографических задач: создание новых подходов к выявлению релевантной лингвистической информации в зависимости от текущих задач использующего данное приложение пользователя. Это, в свою очередь, предполагает совершенствование (и сжатие) традиционных словарных толкований у значений многозначных слов, их эксплицитную контекстуализацию, а также ориентацию на кластерное порождение текста на родном и иностранном языке, использование при этом принципов машинного перевода

и многое другое [Tarp et al. 2017, 494—521; Fuertes-Olivera et al. 2018, 152—176; Tarp et al. 2019, 250—268; Tarp 2019, 224—249; Huang, Tarp 2021, 68—92; Tarp 2022, 66—87].

## **6. Кластеризация дискурса и проблемы идиоматики**

Проблеме кластеризации дискурса и выделению устойчивых оборотов и словосочетаний, их переводу и т.п., посвящена огромная специальная лингвистическая литература (включая словари, онлайн ресурсы и учебники): в связи с преподаванием английского как иностранного, в связи с ярко выраженной идиоматичностью иноязычных оборотов и словосочетаний, а также по причине невозможности их пословного перевода с одного языка на другой и т.д. (подробнее см. [Рябцева 2019; 2021а; 2021б; Riabtseva 2020]).

В статье [Chen 2022], посвященной обучению поиска словосочетаний в словаре английского языка и их переводу на родной язык (и наоборот), подчеркивается, что обучение использованию устойчивых словосочетаний на иностранном, английском, языке — это сложный и долгий процесс, что использование устойчивых словосочетаний на иностранном языке вызывает большие трудности у учащихся, а изучение эффективности различных способов обучения идиоматике: словарный, текстовый и другие, показывает необходимость их сочетания, при том что словари словосочетаний должны постоянно обновляться, размещаться онлайн и совершенствовать аппарат обращения к ним и расширяться за счет корпусных материалов и примеров [Chen 2022, 1—30]. В самой статье [Chen 2022] используется множество специальных устойчивых словосочетаний: *dictionary lookup behaviour, dictionary use instruction, trained dictionary users, task-based dictionary training, a product-based approach, a multiple case study approach, a five-step dictionary training session, collocation dictionary skills development, Japanese learners' Collocation Dictionary retrieval performance, user-friendly search routes, a special-purpose Dictionary of Collocations, an increasing number of dictionary-based collocation studies, overshadowing the non-reportable perceptual and memory processes* и т. п. При этом также отмечается, что онлайн словари имеют свои недостатки: высвечивают слишком много (дополнительной) информации (*Long entries pose problems for many dictionary users*), которая отвлекает

пользователя и затрудняет поиск нужных слов и выражений (*The design of search routes should be more user-friendly*) [Chen 2022, 26]. Тем не менее, онлайн словари оказываются самыми удобными в использовании, особенно если учащиеся получают навыки обращения к ним (*With e-dictionaries gaining popularity, the dictionary use instruction should be based on the consultation habits and reference needs of e-dictionary users* [Chen 2022, 2—27]).

В межъязыковом отношении наиболее актуальной задачей повышения качества научной коммуникации за счет передовых цифровых технологий выступает создание лексикографических ресурсов по переводу с родного языка на английский и с английского языка на родной язык терминологических словосочетаний, отражающих современное стремительное технологическое развитие почти всех научных дисциплин, включая саму лингвистику, лексикографию и научную коммуникацию в целом.

#### **7. Словари терминологических словосочетаний: терминологическая кластеризация научного дискурса в межъязыковом аспекте**

Современная лингвистическая терминология развивается стремительно под воздействием, в первую очередь, цифровых технологий, позволяющих активно и в бесконечно большом объеме обрабатывать значительное количество разнообразной лингвистической и экстралингвистической информации и тем самым не только развиваться самой, но и развивать сами эти технологии. Результаты этих процессов ярко проявляются в терминологии, которая становится многокомпонентной, развернутой и в то же время компактной, сжатой (*compressed*), динамичной, активной, контекстуальной, междисциплинарной и инновационной по форме и содержанию. При этом ведущие специалисты подчеркивают, что в настоящее время двуязычная лексикография, особенно в области терминологии, а также в области инноваций в терминологии и даже в форме онлайн, значительно отстает от постоянно растущих потребностей в качественных терминологических ресурсах, как со стороны переводчика-человека, так и искусственных приложений класса CALL, WA и AW и др. (*Contemporary terminological resources... have considerable importance for translators and should meet their requirements as far as possible.*

Unfortunately, these *resources are frequently of poor quality and do not adequately satisfy their needs*) [Durán-Muñoz 2010, 55—66; Fuertes-Olivera, Esandi-Baztan 2020].

В настоящее время важнейшей задачей в области терминографии выступает также интеграция терминологических ресурсов в лексикографические порталы (ср. Integrating Terminological Resources in Dictionary Portals) [Fuertes-Olivera, Esandi-Baztan 2020]. В этом отношении необходима унификация представления терминографической информации, ее систематизация и интеграция, а также конвергенция с лексикографией (*convergence of terminology and lexicography*). Необходима также типология лексикографического описания словарных единиц, которая позволит сделать процесс формирования лексикографического портала более легким, простым, унифицированным, системным и автоматизированным (*to make the process of compilation easier, more uniform, and more readily systematized*) [Fuertes-Olivera, Esandi-Baztan 2020, 93].

Одной из важнейших лингвистических задач в области цифровой терминографии в настоящее время выступает разносторонняя терминологическая поддержка межъязыковой научной коммуникации: разработка двуязычных терминологических динамических ресурсов, порталов и т. д., а также унификация лексикографического представления терминографической информации в них в связи с тем, что современная научная терминология носит многокомпонентный характер и включает в свой состав преимущественно развернутые терминологические словосочетания-кластеры. Их перевод с родного языка на английский и с английского языка на родной составляет большую коммуникативную и лексикографическую проблему по целому ряду причин. Главная из них это активное появление и использование терминов-неологизмов — новых специальных понятий уточняющего и потому многокомпонентного характера, которых пока нет даже в онлайн словарях и перевод которых с английского языка на родной и обратно представляет большие трудности [Gouws 2016; Riabtseva 2022]. Так, Яндекс Переводчик перевел с английского языка на русский довольно простое словосочетание — название статьи «Collocation, Colligation and Encoding Dictionaries» [Siepmann 2006]

следующим образом: \*«Словари *коллокации*, *Сопоставления* и кодирования». В данном переводе курсивом выделены ошибки: сущ. *коллокации* должно стоять во мн. ч., сущ. *Сопоставления* переведено некорректно и стоит в грамматически и орфографически неправильной форме. Подобные квази-переводы с разного рода стилистическими, грамматическими, орфографическими и др. погрешностями и ошибками порождают квази-научный текст, засоренный языковыми аномалиями. Большинство автоматических онлайн-переводчиков делают подобные малые и большие ошибки в переводе именно самых важных в научной коммуникации компонентов речи — развернутых терминологических словосочетаний, поскольку действуют обычно на основе пословного перевода, без учета контекста, тематики, грамматической нормы, лингвоспецифических синтаксических правил и т.д.

#### **8. Активные модели и структуры терминологических словосочетаний-кластеров в английском языке и проблемы перевода**

Наиболее активными, продуктивными, частотными и аутентичными моделями построения многокомпонентных терминологических словосочетаний-кластеров (*multi-word terminological expressions / collocations / clusters*) в современных лингвистических исследованиях на английском языке являются следующие, каждая из которых допускает дальнейшее расширение дополнительными атрибутивными (Adj), номинативными (N), глагольными (V), адвербиальными (Adv), нумеративными (Num) и другими компонентами (в частности, сокращениями):

##### **1) N-Ved (N):**

*N-based* (N): *usage-based* (technology), *games-based* (grammar teaching), *IT-based* (grammar learning), a *product-based* approach, *web-based* dictionaries, *web-based* error interface, *web-based* management, (customizable) *web-based* (lexicographic) tool, *web-based* (dictionary management) system for bilingual dictionaries; a (free) *web-based* learning environment for French vocabulary, an intelligent *server-based* tutoring system (for mobile

devices), (to provide) a concordance-*based* viewing perspective, (an increasing number of) dictionary-*based* collocation studies;

N/Adv-*driven* (N): technology-*driven* (approach), innovation-*driven* growth, in a (basically) technology-*driven* way, corpus-*driven* lexicography, AI-*driven* decision making, a pedagogically-*driven* search tool (for querying corpora), a test-*driven* development of lexicographic products, (the use of iterative processes with) test-*driven* development, a pedagogy-*driven* design for online language teaching and learning, commercially *driven* integrated dictionary portals, commercially-*driven* language tools constructing;

N-*assisted* (N): computer *assisted* (pronunciation training); a lexicography-*assisted* writing;

N-*oriented* (N): goal-*oriented* (language courseware design), meaning-*oriented* classification of collocations; process-*oriented* writing support with the digital Writing Aid;

N-*supported* (N): web-*supported* (language learning), computer-*supported* collaborative writing;

N-*adapted* (N): context-*adapted* (approach);

N-*challenged* (N): auditory *challenged* (learners);

N-*delivered* (N): (a suite of) internet-*delivered* (diagnostic) tests;

N-*derived* (N): corpus-*derived* (evidence);

N-*directed* (N): self-*directed* (L2 vocabulary) learning;

N-*enhanced* (N): technology-*enhanced* (language learning);

N-*generated* (N): user-*generated* (activity);

N-*informed* (N): corpus-*informed* research;

N/Adv-*interested* (N): (cooperation between) technology-*interested* linguists and pedagogically-*interested* linguists;

N-*mediated* (N): computer-*mediated* (communication);

N-*tagged* (N): error-*tagged* (learner corpora);

N-*ed* N: *networked* (intercultural) language study;

**2) N-Adj (N):**

N-*specific* (N): subject-*specific* (terminology), discipline-*specific* corpora;

N-*independent* (N): platform-*independent* (tools), technology-*independent* content structuring, a language-*independent* system (for parsing unrestricted text);

N-*efficient* (N): time-*efficient* (method), a space-*efficient* data structure, cost-*efficient* production (of content for different medias and target audiences);

N-*adaptive* (N): computer-*adaptive* (language testing), computer-*adaptive* tests (of reading);

N-*aware* (N): context-*aware* (dictionaries), context-*aware* order, (words listed in) a prioritized, context-*aware* order;

**3) Adj-N (N):**

cross-N (N): cross-language (approach); a cross-platform (ubiquitous language learning) service (via mobile phone and interactive television);

high-N (N): high-performance (lexicographic reference) tool;

**4) N-Adv (N):**

user-*friendly* (interface), user-*friendly* search routes, in (no) user-*unfriendly* way;

**5) Num-N N:** (short) *one-line* definitions;

**6) Adj-Adj (N):** well-*defined* (learner populations);

**7) Adv-N N:** up-*market* reference tools, up-*market* online resources;

**8) Adv-to-V N:** easy-*to-consult* interfaces.

Подобные новые, особенно развернутые, терминологические словосочетания-коллокации указанных типов и видов специального (как правило, междисциплинарного) и потому терминологического характера: 1) часто отсутствуют в общих и специальных словарях; 2) различными автоматическими переводчиками (Google Translate, Reverso, Яндекс Переводчик и другими) переводятся по-разному, преимущественно пословно; 3) их автоматический перевод или неточен, или нетерминологичен; 4) их способность втягивать в свой состав различные уточняющие компоненты контекстологического



характера значительно затрудняет их перевод с английского языка на родной и делает их порождение на иностранном английском языке почти невозможным; 5) соответствующие двуязычные автоматические ресурсы пока слабо ориентированы на распознавание и перевод терминологических словосочетаний, особенно контекстуального характера, когда в составной термин включается два и более контекстуально и в смысловом отношении связанных с ним понятия, ср. *a commercially driven language tool, dynamic access structures with different access possibilities, broadening the customer base of online dictionaries, additional lexicographic data with metatexts for further access*, etc. Так, естественное для специального текста на английском языке выражение *a commercially driven language tool* автоматическими переводчиками переводится на русский как ?«коммерческий языковой инструмент», т. е. выражением, которое не используется в специальных текстах на русском языке.

Между тем способность к активной контекстуализации современной терминологии в английском языке является ее принципиально важным свойством. Она позволяет включать в состав термина большое количество дополнительных, уточняющих, конкретизирующих и т. п. компонентов, например: (to develop) *a pedagogically-driven search tool for querying corpora (in such a way that allows) language teachers to access corpora from their teaching perspectives and experiences* [Granger 2007, с. 69], а также объединятся с аналогичными терминологическими конструкциями в единое целое, ср.: *new digital forms of presentation of the lexicographic products; a natural and necessary lexicographic adaptation to the new digital environment*; см. выделенные курсивом ИГ: *Long entries pose problems for many dictionary users; With e-dictionaries gaining popularity, the dictionary use instruction should be based on the consultation habits and reference needs of e-dictionary users; The design of search routes should be more user-friendly* [Chen 2022, 26—27]; *High-performance lexicographic reference tool should be a kind of search engine enabling queries on a lexical database or on the web* [Tarp 2008, 123].

Важнейшим свойством в организации терминологии выступает также ее стремление к симметрии: *primarily vs secondarily CALL-oriented or lexicography-oriented applications*; «*human-oriented vs. machine-oriented tools/ applications («lexicons»)* [Abel 2010, 6]; *Involvement Load Hypothesis vs. Cognitive Load Theory* [Chen 2022]. Еще одним весьма характерным и показательным свойством современной терминологии на английском языке в виде развернутых терминологических словосочетаний-кластеров является их активное использование в названиях научных публикаций, например: *A Logic Based Implementation of Head-Driven Phrase Structure Grammar* [Popowich, Vogel 1991]; *Evolving approaches to Web-supported language learning: From platforms to platform-independent tools* [Wible 2007]; *L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography* [Tarp et al. 2017].

Главной проблемой в переводе развернутых терминологических словосочетаний (терминологических кластеров) с английского языка на русский выступает их фразеологизированный и идиоматичный характер, связанный с их межъязыковой асимметрией: отсутствием в русском языке аналогичных лексико-синтаксических конструкций. Это в свою очередь предопределяет невозможность их пословного перевода, который составляет основу всех видов автоматического (машинного) перевода (МП). Современные системы МП, правда, стремительно эволюционируют и способны решать переводческие задачи не алгоритмически, а лексикографически, включая в свои лексикографические базы данных все возможные и разнообразные обороты, выражения, конструкции и «кластеры» целиком, равно как и их межъязыковые эквиваленты. Однако сама современная терминология развивается так быстро и динамично, что словари не успевают ее фиксировать. Кроме того, способность терминологических словосочетаний контекстуализироваться — включать в свой состав дополнительные контекстно связанные уточняющие специальные понятия исключает их полное перечисление в словаре, например: *Some projects mainly re-use already existing lexicographic data in a basically technology-driven way without explicitly applying any specific lexicographic approach* [Granger

2007, 61]; *This is a corpus-based approach to writing definitions taking into account thematic variation in the selection of information* [IJL—2022].

В результате видно, что нужны более явные и адекватные правила перевода развернутых терминологических словосочетаний на иностранный английский и на родной русский язык. Исходным моментом тут выступает исчисление их структурных типов в английском языке, их разбиение на автономные ядерные коллокации, задание порождающих моделей их развертывания и сворачивания и др. Так, по своей структуре многокомпонентные именные терминологические словосочетания английского языка можно разделить на следующие типы: 1) составные (двусоставные) именные группы (ИГ) вида N+N и ИГ, состоящие из трех и более именных компонентов вида (ADJ+)N+N(+N): *language learning, target audiences, speech recognition, learner corpora, user research; online resources, suggestion window, consultation window, dictionary compilation, dictionary lookup behaviour, definition writing guidelines, web database applications, multiple access routes, (digital bilingual) lexicographic products design, an online dictionary writing system*; 2) сложные (двухсоставные ИГ с пассивным причастием) вида N+V-ed: *mobile-assisted, user-generated, user-oriented, NLP based*; 3) сложно-составные (трех- и четырех компонентные, один из которых пассивное причастие) вида N+V-ed+N(+N): *a logic—based implementation, user-generated activity, mobile assisted language learning*, 4) усложненные (четырёх—пяти компонентные) вида N+V-ed+N+N(+N): *head-driven phrase structure grammar, computer-mediated communication for language learning, avenues for future learner-corpus-informed research*; 5) комбинированные (два и более составных компонента в виде ИГ): *cost-efficient production of content for different medias and target audiences, user-oriented NLP-based teaching platform; computer assisted pronunciation training applications using automatic speech recognition; a cross-platform ubiquitous language learning service via mobile phone and interactive television*.

Основным принципом порождения соответствующих переводов при этом выступает учет перевода предшествующего (более простого типа) словосочетания при

переводе словосочетания более сложного типа, в противном случае системность и последовательность в организации терминологии не будет соблюдена. Однако существенно гораздо более трудная проблема при переводе терминологических развернутых словосочетаний с английского языка на русский заключается в том, что в подавляющем большинстве случаев их дословный, близкий к тексту перевод оказывается по-русски семантически недостаточным и требует уточнений и конкретизации. Так, автоматический (дословный) перевод развернутого терминологического словосочетания *the use of iterative processes with test-driven development* как «использование итерационных процессов с разработкой через тестирование» (Google Translate) семантически недостаточен и стилистически неудачен: разработка чего? В тексте ему соответствует следующий смысл (контекстуальный перевод): ‘апробирование различных итеративных форматов представления данных пользователю’, ср.: Based on the results from observation, D. Norman advocates the use of iterative processes with test-driven development [Tarp, Gouws 2020, 475]. Ср. также: *dictionary-based collocation studies* — «изучение словосочетаний на основе ?словаря», *an increasing number of dictionary-based collocation studies* — «растущее число исследований словосочетаний на основе ?словарей» (Google Translate).

Как видно из проведенного исследования (и ряда предшествующих публикаций), чаще всего перевод с английского языка на русский язык сложных составных (многокомпонентных) терминов требует экспликации разного рода смыслов, имплицитно представленных в компактных и свернутых терминологических выражениях на английском языке, т. е. в явном представлении дополнительных пояснительных компонентов в переводе. Так, предельно лапидарное англ. *lexicographic user research* не может быть передано (при первом упоминании) на русский язык дословно и буквально, как это делают все автоматические переводчики, а именно: ‘лексикографическое исследование пользователей’. Напротив, многословное и эксплицитное представление соответствующего сложного смысла в виде естественной для русского языка развернутой фразы типа *автоматическое отслеживание обращений пользователя к различным видам*

*лексикографической информации*, которая по своему содержанию приближается к дефиниции соответствующего сложно-составного понятия, позволяет полностью пояснить сущность соответствующего технического процесса, который «позволяет определить пользовательские приоритеты и оперативно вносить коррективы в порядок предоставления различных подсказок и др. лингвистической информации в соответствующий лексикографический процессор» (см. [Тагп 2019, 244]).

Следует особо отметить, что самой сложной задачей в современной межъязыковой научной коммуникации является не столько перевод терминологических словосочетаний с английского языка на русский, сколько противоположный процесс: перевод с родного, русского языка на английский. Он оказывается еще в большей степени фразеологизированным и идиоматичным, чем перевод с английского языка на русский, а также более наукоемким: необходимо (у)знать точно, как именно заданный смысл идиоматично выражается на иностранном английском языке. Самому сконструировать варианты перевода типа «*corpus-derived evidence*», «*self-directed L2 vocabulary learning*» и т. п. почти невозможно по той простой причине, что у них нет прямых, дословных аналогов-терминов в русском языке. Особенно показательным является тот факт, что подобные развернутые идиоматичные терминологические словосочетания не используются (по указанным выше причинам) при переводе названий научных публикаций с русского языка на английский.

## **9. Заключение**

В заключение следует подчеркнуть, что современные алгоритмы автоматического перевода научных текстов с русского языка на английский и с английского языка на русский требуют дальнейшего совершенствования, особенно с учетом отличительных особенностей современной научной терминологии и ее свернутого, идиоматического и в то же время активного, динамичного и многокомпонентного характера — с одной стороны, и с учетом новейших тенденций в цифровой лексикографии — с другой. Последние показывают, что благодаря новейшим цифровым технологиям и их стремительному

развитию, а также в ответ на предъявляемые к ней принципиально новые требования, сама лексикография и создаваемые ею современные лексикографические ресурсы активно приобретают новые конструктивные свойства: динамичность, операциональность, интерактивность, многофункциональность, многоканальность (доступа), интегративность, инкорпорированность, персонифицированность (*personalization*, ориентация на (конкретного) пользователя), адаптивность (учет пользовательских задач), (автоматизированная) пользовательская настройка (*automated customization*), контекстуализация, а также индустриализация: разработка комплексных цифровых лексикографических служб и их симбиоз с интеллектуальными системами автоматизированной обработки естественного языка. Показательно, что новые задачи, которые ставят перед лексикографией новые технологии, не могут быть решены традиционными лексикографическими методами анализа потребностей пользователя. Основным способом их решения становятся цифровые технологии, которые способны фиксировать и прогнозировать действия пользователя и его информационные потребности.

Особое место и значение в современной межъязыковой научной коммуникации занимает двуязычная терминография. Соответствующие двуязычные ресурсы пока слабо ориентированы на распознавание и перевод современной, высокотехничной, активной, динамичной, предельно содержательной терминологии, органично вписывающейся в текущий контекст, особенно в виде развернутых многокомпонентных терминологических словосочетаний-кластеров, от принципов включения которых в двуязычные терминологические ресурсы в значительной степени зависит уровень развития межъязыковой научной коммуникации.

### Список сокращений

Adj — Adjective

Adv — Adverb

AW — Augmented Writing

CALL — Computer-assisted Language Learning

N — Noun

Num — Numeral

V — Verb

WA — Write Assistant

ИГ — Именная группа

ИИ — Искусственный интеллект

МП — Машинный перевод

### **Источники**

IJL—2022 — International Journal of Lexicography. 2022. //

<https://academic.oup.com/ijl>.

GPT—2 — Natural Language Understanding model «Generative Pre-trained Transformer 2 (GPT-2)»: an open-source artificial intelligence created by OpenAI . 2019.

WALL — Web Assisted Language Learning // <http://wall.eila.univ-paris-diderot.fr>.

Write Assistant — <https://www.writeassistant.com/da>.

Wordfast — <https://www.wordfast.com/>.

### **Литература**

Рябцева Н. К. «Конструкционная грамматика» и научная коммуникация: Межъязыковой аспект // Научный диалог. 2019. № 6. С. 50—71. DOI: <https://doi.org/10.24224/2227-1295-2019-6-50-71>.

Рябцева Н. К. Термины как когнитивные доминанты в научной коммуникации и их межъязыковые соответствия // Когнитивные исследования языка. 2021а. № 1 (44). С. 356—364.

Рябцева Н. К. Переводческие парадоксы и их когнитивная интерпретация / Н. К. Рябцева // Когнитивные исследования языка. 2021б. № 4 (47). С. 383—390.

Abel A. Towards a systematic classification framework for dictionaries and CALL // S. Granger, M. Paquot (eds). eLexicography in the 21st century: New challenges, new applications.

Proceedings of eLex-2009. Louvain-la-Neuve, 22-24 October, 2009. Cahiers du Cental, 2010. № 7. Pp. 3—11.

Abel A., Bracco St. From an online dictionary to an online dictionary writing system // G.-M. de Schryver (ed.). DWS 2006: Proceedings of the Fourth International Workshop on Dictionary Writing Systems. Pretoria: (SF) Press, 2006. Pp. 25—34.

Banks C. What is an Augmented Writing Platform? 2019. // <https://medium.com/swlh/what-is-an-augmented-writing-platform-b28fa588a1c5>.

Boelhouwer B., Sijens H., Dykstra A. Dictionary Portals. Pedro A. Fuertes-Olivera (Ed.). The Routledge Handbook of Lexicography. 2018. Pp. 754—766.

Bothma T. J. D., Tarp S. Why Relevance Theory Is Relevant for Lexicography // Lexicographica, 2014. № 30. Pp. 350—378.

Colson E. What AI-Driven Decision Making Looks Like. 2019. // <https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like> .

Chen Yu. Effect of Learning Conditions on Collocation Gains: A Case Study of Task-based Dictionary Use Instruction // Lexikos, 2022, № 32 (AFRILEX-reeks/series 32: 2022). Pp. 1—30. (1679-Article Text-3062-1-10-20220227.pdf).

Durán-Muñoz I. Specialised lexicographical resources: a survey of translators' needs // S. Granger, M. Paquot (eds). eLexicography in the 21st century: New challenges, new applications. Proceedings of eLex-2009, Louvain-la-Neuve, 22-24 October, 2009. Cahiers du Cental, 2010, № 7. Pp. 55—66.

Frankenberg-Garcia A., Lew R., Rees G. Roberts J. C., Sharma N., Butcher P. *ColloCaid*: Around 30 thousand academic English collocations and examples of collocations in context curated from expert writing corpora. 2021 // <http://www.collocaid.uk/>

Fuertes-Olivera P.A., Tarp S. Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography. Berlin/Boston: De Gruyter, 2014.



Fuertes-Olivera P.A. Designing and Making Commercially Driven Integrated Dictionary Portals: the Diccionarios Valladolid-Uva // *Lexicography*, 2019, № 6(1). Pp. 21—41. doi.org/10.1007/s40607-019-00056-8.

Fuertes-Olivera P.A., Esandi-Baztan M.A. Integrating Terminological Resources in Dictionary Portals: The Case of the Diccionarios Valladolid-UVa // *Lexikos*, 2020, № 30 (AFRILEX-reeks/series 30). Pp. 90—110.

Fuertes-Olivera P.A., Tarp S. A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant // *Lexicographica*, 2020, № 36. Pp. 257—286.

Fuertes-Olivera P.A., Tarp S., Sepstrup P. New Insights in the Design and Compilation of Digital Bilingual Lexicographical Products: The Case of the Diccionarios Valladolid-UVa. *Lexikos*, 2018, № 28. Pp. 152—176.

Gouws R.H. Increasing the Scope of the Treatment of Specialised Language Terms in General Dictionaries // Schierholz, S., R.H. Gouws, Z. Hollós and W. Wolski (Eds.). *Wörterbuchforschung und Lexikographie*. Berlin: De Gruyter, 2016. Pp. 101—119 // <https://doi.org/10.1515/9783110474251-007>.

Granger S. (ed.) *Optimizing the role of language in Technology-Enhanced Learning*. Proceedings of the expert workshop organized by the Integrated Digital Language Learning (IDILL) seed grant project. Louvain-la-Neuve (Belgium), 4—5 October 2007.

Granger S., Paquot M. Preface // *Cahiers du Cental*, 2010, № 7. «eLexicography in the 21st century New challenges, new applications». Proceedings of eLex 2009. Louvain-la-Neuve 22-24 October 2009. Sylviane Granger and Magali Paquot (eds). UKL Presses, 2010. Pp 1—5.

Huang F., Tarp S. Dictionaries Integrated into English Learning Apps: Critical Comments and Suggestions for Improvement // *Lexikos*. 2021, № 31. P. 68—92. ([suphttp://lexikos.journals.ac.za](http://lexikos.journals.ac.za)) // <https://doi.org/10.5788/31-1-1626>.

Liew A. DIKIW: Data, Information, Knowledge, Intelligence, Wisdom and their Interrelationships // *Business Management Dynamics*. 2013, Vol. 2, Issue 10. Pp. 49—62.

Loewen S., Crowther D., Isbell D. R., Kim K. M., Maloney J., Miller Z. F., Rawal H. Mobile-Assisted Language Learning: A Duolingo Case Study // *ReCALL*, 2019, № 31(3). Pp. 293—311.

Kilgarriff A. How to monetise a web presence (and Hoover a moose). A report on the e-lexicography conference at Louvain-la-Neuve, Belgium, 22—24 October 2009 // <http://www.uclouvain.be/en-307398.htm>

Marconi F. The Year of Augmented Writing. 2022. *Textio* // <http://www.techxplore.com>

OpenAi2019 // <https://openai.com/>.

Pajzs J. On the possibility of creating multifunctional lexicographical databases // H. Bergenholtz, S. Nielsen, S. Tarp (eds). *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern: Peter Lang, 2009. Pp. 327—354.

Popowich F., Vogel C. A Logic Based Implementation of Head-Driven Phrase Structure Grammar // C. Brown and G. Koch (eds). *Natural Language Understanding and Logic Programming, III*. Elsevier, North-Holland. 1991. Pp. 227—246.

Riabtseva N. K. «From augmented reality to augmented human»: Contemporary information technologies and their impact on human intelligence // *Cognitive Modeling: Proceedings of the Fourth International Forum on Cognitive Modeling (11—18 September, 2016, Lloret de Mar, Spain)*. In 2 parts. Part 1. *Cognitive Modeling in Linguistics: Proceedings of the XVII International Conference «Cognitive Modeling in Linguistics. CML-2016»*. Rostov-on-Don: Science and Studies Foundation, 2016. Pp. 412—421.

Riabtseva N. K. Contemporary digital technologies and innovations in human intelligence // *Russian Linguistic Bulletin*. 2017, № 3 (11). Pp. 7—16. DOI: 10.18454/RULB.11.15

Riabtseva N. K. Cross-Cultural Communication: Its Asymmetry and Authenticity // *Nauchnyi dialog*. 2020, № 4. Pp. 130—150. DOI: <https://doi.org/10.24224/2227-1295-2020-4-130-150>. (In Russ.).

Riabtseva N. K. Contemporary terminological problems in a cross-linguistic perspective // Nauchny Dialog. 2022. № 8. (In Russ.). (In press)

de Schryver Gilles-Maurice. Lexicography in the crystal ball: Facts, trends and outlook. // Fjeld R. V., Torjusen J. M. (eds). Proceedings of the 15th EURALEX International Congress, 711. August, 2012. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, 2012. Pp. 93—163.

Siepmann D. Collocation, Colligation and Encoding Dictionaries. Part II: Lexicographic Aspects // International Journal of Lexicography. 2006, № 19(1). Pp. 1—39.

Simonsen H. K. Augmented Writing and Lexicography: A Symbiotic Relationship? // EURALEX. Proceedings. 2020. Pp. 509—514.

Tarp S. Lexicography in the borderland between knowledge and non-knowledge. Tübingen: Niemeyer, 2008.

Tarp S. Connecting the Dots: Tradition and Disruption in Lexicography // Lexikos, 2019, № 29. Pp. 224—249.

Tarp S. Turning Bilingual Lexicography Upside Down: Improving Quality and Productivity with New Methods and Technology // Lexikos. 2022, № 32 (AFRILEX-reeks/series 32). Pp. 66—87.

Tarp S., Fisker K., Sepstrup P. L2-Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography // Lexikos. 2017, № 27. Pp. 494—521.

Tarp S., Gouws R. H. Lexicographic Contextualization and Personalization: A New Perspective // Lexikos. 2019, № 29. P. 250—268.

Verlinde S. The Base *lexicale du français*: a multi-purpose lexicographic tool // Cahiers du Cental, 2010, n° 7. Pp. 335—342.

De Wachter L., Verlinde S., D'Hertefelt M., Peeters G., Tounsi L. How to deal with students' writing problems? Process-oriented writing support with the digital Writing Aid Dutch // Rak Rafal (Ed.) The 25th International Conference on Computational Linguistics, 2014. Dublin. Proceedings of the Conference. System Demonstrations. 2014, Pp. 20—25.

Wible D. Evolving approaches to Web-supported language learning: From platforms to platform-independent tools // Granger S. (ed.) *Optimizing the role of language in Technology-Enhanced Learning*. Proceedings of the expert workshop organized by the Integrated Digital Language Learning (IDILL) seed grant project, Louvain-la-Neuve (Belgium), 4-5 October 2007. Pp. 65—69.

**Contemporary lexicography, «augmented linguistic reality»  
and terminographic challenges**

**N. K. Riabtseva (Institute of Linguistics, Russian Academy of Sciences)**

The paper shows that rapidly developing digital technologies radically transform contemporary lexicography and its principles in representing linguistic information, particularly for various intellectual systems and applications. New active trends in the evolution of lexicography and in the compiling and functioning of innovative lexicographic resources are revealed. Particular attention is paid to a most urgent problem within contemporary cross-linguistic scientific communication — the lexicographic representation of modern terminology, which is stimulated by digital technologies, acquires fundamentally new properties and becomes more dynamic, multiword, compositional, context-aware, cross-disciplinary, etc.

Key words: Digital technologies, lexicographic resources augmentation, Artificial Lexicographer, terminological clusters, cross-linguistic asymmetry.