Вопросы лексикографии. 2025. № 37. С. 96–118
Russian Journal of Lexicography. 2025, 37, pp. 96–118

Научная статья
УДК 811.112.2 (Немецкий язык)
doi: 10.17223/22274200/37/5

Структура словарных статей электронного двуязычного словаря и его интеграция с корпусом

Анна Юрьевна Егорова[1]

[1] *Федеральный исследовательский центр «Информатика и управление» РАН, Москва, Россия, anna.yu.egorova@yandex.ru*

Аннотация. Рассмотрена задача интеграции электронного двуязычного словаря с корпусом параллельных текстов на примере разрабатываемой в ФИЦ ИУ РАН лексикографической информационной системы (ЛГИС). Сопоставлены возможности интеграции в ЛГИС и подход к интеграции в других корпусных лексикографических ресурсах. Дано описание полей и способов связи, обеспечивающих интеграцию словаря с корпусом через интерфейсную базу данных. В первой версии ЛГИС связи словаря с корпусом реализованы для трех зон словарной статьи (заглавного слова, значения и идиоматики).

Ключевые слова: лексикографическая информационная система, электронный словарь, немецкий язык, русский язык, корпус, база данных, интеграция

Благодарности. Автор выражает благодарности А.А. Гончарову и Д.А. Бахматову за помощь в подборе и анализе примеров, а также анонимным рецензентам за ценные замечания, работа над которыми помогла улучшить статью. Исследование выполнено в ФИЦ ИУ РАН за счет гранта Российского научного фонда № 24-18-00155, с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

Для цитирования: *Егорова А.Ю.* Структура словарных статей электронного двуязычного словаря и его интеграция с корпусом // Вопросы лексикографии. 2025. № 37. С. 96–118. doi: 10.17223/22274200/37/5

Original article

Structure of dictionary entries of a digital bilingual dictionary and its integration with the corpus

Anna Yu. Egorova[1]

[1] *Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russian Federation, anna.yu.egorova@yandex.ru*

Abstract. The article considers the problem of integrating a digital bilingual dictionary with a corpus of parallel texts on the example of the lexicographic information system (LGIS) being developed at the FRC CSC RAS. The relevance of this issue is due to the increased attention to the creation and updating of digital dictionaries based on corpus data both in Russia and abroad. In particular, on the 25th December 2024 the Russian Government issued Resolution No. 1892 "On the National Dictionary Fund", which set the task of automating the lexicographic activity in the Russian Federation. It involves converting existing paper dictionaries into digital format and a regular subsequent updating of these dictionaries. The aim of this article is to describe fields and methods that ensure the integration of a digital German-Russian dictionary with a text corpus in the framework of LGIS. To achieve this aim, a review of corpus-oriented lexicographic resources existing in Russia and abroad was conducted. The proposed approach to integration in LGIS was compared with integration solutions in similar resources. Six types of fields and three methods of integration were described, ensuring the connection of a dictionary entry with a corpus of parallel texts using the interface database. In the first version of LGIS, the links of a dictionary entry with a corpus are implemented in three zones of the dictionary entry, namely, in the zones of the headword, of description of its meanings, and of idioms. The material of the study comprises a German-Russian dictionary (about 40 000 dictionary entries) and parallel texts of the German-Russian subcorpus of the Russian National Corpus (about 18 million word usages). The methods of analysis, synthesis, and structuring of inherited lexicographic resources were used as a methodological base. The main result of the study is the description of six types of integration fields that provide a transit from the dictionary to the text corpus, in particular: (1) from the headword of the dictionary entry, (2) from the meaning of the focal language unit (in the first version of LGIS from the meanings of German modal

verbs), (3) from the phraseme, (4) from the meanings of the phraseme, (5) from the idiom, and (6) from the meanings of the idiom. These links are implemented in three ways of integrating the digital dictionary with the corpus of texts: (1) through a search in the corpus by the lemma of the entry word, (2) through a table of annotated translation correspondences, and (3) using a search query for phraseme or idiom forms. The scientific novelty of the work lies in the categorization of fields and methods that ensure integration of the dictionary with the corpus.

Keywords: lexicographic information system, digital dictionary, German, Russian, corpus, database, integration

Acknowledgments: The author is grateful to A.A. Goncharov and D.A. Bakhmatov for their assistance in selecting and analyzing the examples, and to the anonymous reviewers for their valuable comments, which helped to improve the article. This study is funded by the Russian Science Foundation, Project No. 24-18-00155. It was carried out using the infrastructure of the Center for Collective Use "High-Performance Computing and Big Data" of the FRC CSC RAS (Moscow).

For citation: Egorova, A.Yu. (2025) Structure of dictionary entries of a digital bilingual dictionary and its integration with the corpus. *Voprosy leksikografii – Russian Journal of Lexicography*. 37, pp. 96–118. (In Russian). doi: 10.17223/22274200/37/5

Введение

Цель статьи – описать поля и три способа связи, обеспечивающие интеграцию электронного немецко-русского словаря с корпусом параллельных текстов в рамках создания лексикографической информационной системы (ЛГИС), а также сопоставить предлагаемый подход к интеграции в ЛГИС с решениями по интеграции в аналогичных лексикографических ресурсах.

Для достижения поставленной цели были сформулированы следующие *задачи*: 1) дать краткий обзор российских и зарубежных корпусно-ориентированных лексикографических ресурсов; 2) описать структуру словарной статьи электронного немецко-русского словаря ЛГИС, созданного на основе наследуемых лексикографических ресурсов; 3) описать и проиллюстрировать виды полей связи словаря

с корпусом параллельных текстов, а также предлагаемые способы, реализующие эти связи. Для разработки электронного словаря ЛГИС был использован печатный «Немецко-русский словарь актуальной лексики» [1], который, в свою очередь, опирается на наследуемые лексикографические ресурсы, включающие двуязычные [2; 3] и одноязычные [4–6] словари.

В рамках ЛГИС интеграция словаря с корпусом реализуется как по словам и устойчивым словосочетаниям, так и по их значениям для трех категорий языковых единиц: немецких модальных глаголов и двух категорий устойчивых словосочетаний (фразем[1] и идиом[2]). ЛГИС включает три компонента: электронный словарь, корпус параллельных текстов объемом 18 млн словоупотреблений (см. работы [7; 8]), взятых из немецко-русского подкорпуса Национального корпуса русского языка [9], и связующую их интерфейсную базу данных (БД), спроектированную в рамках концепции надкорпусных БД [10].

Актуальность проблемы интеграции словарей с корпусами текстов обусловлена возросшим вниманием к созданию и актуализации электронных словарей как в России, так и за рубежом. 25 декабря 2024 г. Правительство РФ утвердило постановление № 1892 «О Национальном словарном фонде», в сообщении о котором сказано: «Лексикографическая деятельность, ориентированная на печатное издание словарей, не даёт возможности получить полноценное представление об актуальном состоянии языка, отстаёт от культурно-языковых потребностей общества. Создание Национального словарного фонда позволит устранить этот пробел» [11]. Основная идея этого постановления состоит в том, чтобы перевести в электронный формат

[1] Фраземы – устойчивые словосочетания со слабой степенью идиоматичности, то есть такие, в которых опорное слово сохраняет одно из своих словарных значений (дефиниция, используемая в рамках проекта РНФ № 24-18-00155).

[2] Идиомы – устойчивые словосочетания, для которых невозможно или крайне затруднительно определить, в каком именно значении то или иное слово употреблено в составе идиомы, поэтому идиомы никогда не располагаются внутри описания одного из значений многозначного слова, а находятся в конце словарной статьи за ромбом (дефиниция, используемая в рамках вышеуказанного проекта РНФ).

или словарные БД несколько десятков словарей, которые существуют в бумажной форме, и обеспечить к ним бесплатный доступ через Интернет. Кроме того, планируется не реже, чем раз в пять лет, проводить их актуализацию. Таким образом, предполагается автоматизировать лексикографическую деятельность.

Актуальность интеграции словаря с корпусом как одной из первостепенных задач цифровой лексикографии подробно обоснована в работе [12]. *Новизна* настоящего исследования заключается в категоризации полей связи и способов, обеспечивающих интеграцию электронного двуязычного словаря с корпусом параллельных текстов.

Корпусно-ориентированные лексикографические ресурсы

Прежде чем перейти к описанию разрабатываемой ЛГИС и ее сопоставлению с другими ресурсами, представляется важным привести краткую классификацию словарей (по основанию степени корпусной поддержки) [13], которые создавались с использованием корпусных данных, но без обеспечения непосредственной интеграции словаря с текстами корпуса.

В обозначенной классификации выделяются три типа словарей, которые будут проиллюстрированы примерами словарей русского языка.

1) Первый тип – словари, основанные на корпусе, т. е. такие словари, концепция которых была сформирована вне корпусной парадигмы, но при создании которых активно применялись корпусы (прежде всего, для создания эмпирической базы). К этому типу словарей относятся, например, «Активный словарь русского языка» под руководством Ю.Д. Апресяна [14] и «Толковый словарь русской разговорной речи» под ред. Л.П. Крысина [15].

2) В качестве второго типа выделяют словари, «вдохновленные» корпусом. Такие словари концептуально основываются на корпусе текстов, т. к. встроенные в корпус инструменты (такие как, например, разметка, наличие подкорпусов, автоматический подсчет количественных данных) позволяют формировать, а затем описывать языковые массивы [13]. Примером словаря, «вдохновленного» корпусом,

может послужить «Частотный словарь современного русского языка» [16].

3) Третий тип – словари, «управляемые» корпусом. Такие словари нельзя назвать словарями в привычном понимании, они представляют собой скорее лингвистические инструменты, работающие на базе корпусов. Это автоматизированные системы, которые с опорой на текстовые массивы и без участия лексикографа выдают пользователю запрашиваемую словарную информацию. В качестве примера можно упомянуть русскоязычный ресурс CoCoCo (Collocations, Colligations, and Corpora) [17].

Все перечисленные словари были созданы в результате развития корпусной лексикографии [18], но важно еще раз отметить, что в отличие от описываемых ниже ресурсов (DWDS и разрабатываемой ЛГИС) эти словари не обладают онлайновой интеграцией с корпусом текстов, которые были в той или иной мере задействованы при их создании.

Что касается задачи интеграции словаря с корпусом текстов, то в России и за рубежом [19–21] выполняются исследовательские проекты, цель которых заключается в разработке подходов к формированию и онлайновой интеграции электронных словарей с корпусами, что дает возможность оперативно обновлять электронные словари и предоставлять пользователям доступ к большому спектру примеров контекстного использования слов и устойчивых словосочетаний [22]. В частности, в проекте Берлинско-Бранденбургской академии наук под названием DWDS (Digitales Wörterbuch der deutschen Sprache) уже реализована онлайновая интеграция словарей с корпусами, но это сделано только для одноязычных словарей [23].

ЛГИС, создаваемая в ФИЦ ИУ РАН, имеет ряд принципиальных отличий от отечественных и зарубежных корпусно-ориентированных лексикографических проектов, включая DWDS (см. подробнее об этом проекте в работах [24; 25]). Основные отличия ЛГИС заключаются в следующем.

Во-первых, ЛГИС предполагает возможность перехода не только к примерам корпуса параллельных текстов по заглавному слову словарной статьи, но и к примерам с конкретным значением заглавного

слова (для фокусных языковых единиц, ФЯЕ). Под фокусными языковыми единицами понимаются те категории языковых единиц, для которых в ЛГИС реализована связь словаря и корпуса текстов не только по словам и устойчивым словосочетаниям, но и по их значениям [26]. Что касается DWDS, то данный ресурс обеспечивает переход к корпусам текстов от заглавных слов и устойчивых словосочетаний без учета их значения в текстовых фрагментах корпуса [23].

Во-вторых, первая версия ЛГИС ориентирована на пару языков, в отличие от одноязычных словарей проекта DWDS, а принципы ее создания могут быть масштабированы на несколько языков.

Среди электронных одноязычных словарей других германских языков можно отметить словарь английского языка Oxford English Dictionary [27], в словарных статьях которого используются заранее отобранные лексикографами из корпуса примеры предложений с заглавным словом статьи, но у пользователя отсутствует возможность перехода к другим примерам корпуса с этим заглавным словом, т. е. онлайновая интеграция не реализована. Подробнее об электронном словаре Oxford English Dictionary см. в работе [28]. Еще одним примером основанного на корпусе словаря английского языка может послужить Merriam-Webster Dictionary [29].

Перечисленные функциональные возможности ЛГИС по интеграции, отсутствующие в рассмотренных лексикографических ресурсах, обеспечиваются во многом благодаря многоуровневой структуре словарной статьи разработанного электронного словаря ЛГИС. Его интерфейс включает поля, которые используются для интеграции электронного словаря с корпусом параллельных текстов (далее – поля связи), в том числе и по значениям, что отсутствует и в отечественных, и в зарубежных аналогах. Описанию полей и способов связи посвящены следующие разделы статьи.

Структура словарной статьи

Структура словарной статьи электронного словаря на первом уровне включает 10 зон, представленных в табл. 1. Некоторые зоны словарной статьи делятся на поля, которые образуют многоуровневую

Таблица 1

Зоны словарной статьи электронного словаря ЛГИС

1.	Зона заглавного слова: связь статьи с корпусом в интерфейсе словарной статьи через поиск по лемме
2.	Зона лексических омонимов
3.	Зона альтернативных вариантов входа в словарную статью
4.	Зона грамматических омонимов
5.	Зона этимологии
6.	Зона фонетической транскрипции
7.	Зона грамматической информации о заглавном слове в целом
8.	Зона помет, относящихся к заглавному слову в целом
9.	Зона значения
10.	Зона идиоматики

структуру, при этом число уровней структуризации различных зон может быть разным. В настоящей статье будет описана структура только тех зон словарной статьи, поля связи в интерфейсе которых используются для интеграции электронного словаря с корпусом параллельных текстов. В данном разделе представлена структура трех обозначенных зон и их полей, которые дополняются полями связи в интерфейсе словарной статьи.

Три зоны словарной статьи, а именно зоны заглавного слова, значения заглавного слова и идиоматики, дополняются в интерфейсе словаря полями связи с параллельными текстами корпуса. В табл. 2 и 3 отображены поля связи (их названия начинаются со слова «связь») с указанием используемого способа интеграции. Первое поле связи относится к зоне 1 «Заглавное слово». Более подробное описание полей связи и способов интеграции дано в следующем разделе.

Одним из ключевых понятий для описания полей связи электронного словаря с корпусом является понятие аннотированного переводного соответствия (АПС) [30]. АПС включает в себя фрагмент текста оригинала, некую ФЯЕ и признаки ее употребления, перевод фрагмента, а также использованный вариант перевода ФЯЕ и признаки его

Таблица 2

Структура зоны значения заглавного слова

<table>
<tr><th>Зона</th><th colspan="3">Поля зоны</th><th>Вид и способ связи словарной статьи с корпусом в ее интерфейсе</th></tr>
<tr><td rowspan="12">Значение 1</td><td colspan="3">Номер значения</td><td rowspan="6">Связь для первого значения модального глагола в качестве ФЯЕ с корпусом через таблицу АПС, сформированных для всех ФЯЕ</td></tr>
<tr><td colspan="3">Пометы</td></tr>
<tr><td colspan="3">Комментарий</td></tr>
<tr><td colspan="3">Варианты перевода</td></tr>
<tr><td colspan="3">Примеры употребления с переводом</td></tr>
<tr><td rowspan="4">Подзначение 1_1</td><td colspan="2">Пометы</td></tr>
<tr><td colspan="2">Комментарий</td></tr>
<tr><td colspan="2">Варианты перевода</td></tr>
<tr><td colspan="2">Примеры употребления с переводом</td></tr>
<tr><td colspan="3">Подзначение 1s</td><td rowspan="5">Связь фраземы с корпусом через поисковый запрос, доступный в интерфейсе словарной статьи</td></tr>
<tr><td rowspan="5">Фразема 1 с использованием заглавного слова в значении 1</td><td colspan="2">Фразема</td></tr>
<tr><td rowspan="3">Значение 1</td><td>Пометы</td><td rowspan="3">Связь фраземы в первом значении с корпусом через таблицу АПС, сформированных для всех ФЯЕ</td></tr>
<tr><td>Варианты перевода</td></tr>
<tr><td>Примеры употребления с переводом</td></tr>
<tr><td>Значение m</td><td></td><td>Связь для m-го значения фраземы через таблицу АПС</td></tr>
<tr><td>Значение n</td><td></td><td></td><td></td><td>Связь для n-го значения ФЯЕ через таблицу АПС</td></tr>
</table>

Таблица 3

Структура зоны идиоматики

Зона	**Поля зоны**		**Вид и способ связи словарной статьи с корпусом в ее интерфейсе**
Зона идиоматики	Идиома		Связь идиомы с корпусом через поисковый запрос, доступный в интерфейсе словарной статьи
	Значение 1	Пометы	Связь идиомы в первом значении с корпусом через таблицу АПС, сформированных для всех ФЯЕ
		Варианты перевода	
		Примеры употребления с переводом	
	Значение *n*		Связь идиомы в *n*-м значении через таблицу АПС

употребления. АПС формируются в интерфейсной БД лингвистами-экспертами вручную.

Рассмотрим данное понятие на примере АПС (см. табл. 4), сформированного для модального глагола *sollen*. В первой версии ЛГИС в качестве ФЯЕ выбраны модальные глаголы, а также некоторые многозначные фраземы и идиомы немецкого языка. Предполагается, что во второй версии ЛГИС перечень ФЯЕ может быть расширен. В первом столбце табл. 4 представлен фрагмент текста оригинала на немецком языке с аналитической формой глагола *sollen* (*hätte sollen*). Во втором столбце указан инфинитив модального глагола *sollen*, также описаны признаки его употребления, а именно: наличие грамматического субъекта *ich* в 1 л. ед. ч. (1sg); грамматическая форма глагола (*hätte sollen*, Plusquamperfectum conjunctivi); наличие у модального глагола подчиненного инфинитива (*annehmen*, +Inf), а также значение модального глагола по электронному словарю (sollen-02). Словарное описание второго значения глагола *sollen* дано в следующем разделе. Третий столбец содержит перевод фрагмента оригинала на русский язык. В четвертом столбце указан вариант перевода модального глагола *sollen* в данном фрагменте (модальный предикатив *надо*), а также

Таблица 4

Пример АПС с глаголом *sollen*

Фрагмент оригинала	**Модальный глагол в оригинале и признаки его употребления**	**Фрагмент перевода**	**Вариант перевода и признаки его употребления**
Ich **hätte** sein Angebot, mir Kaffee zu kochen, *annehmen* und ihn noch etwas festhalten **sollen**.	**sollen** <1sg> <PqpConj> <+Inf> <sollen-02>	**Надо было** *принять* его предложение – пусть бы сварил мне кофе и еще немного посидел.	**надо** <Past> <+Inf>

даны признаки употребления варианта перевода: грамматическая форма прошедшего времени (Past) и наличие подчиненного инфинитива *принять* (+Inf).

В табл. 2 представлена структура зоны значения заглавного слова (№ 9 в табл. 1) и указаны поля связи для зоны значения. В табл. 3 даны поля связи для зоны идиоматики (№ 10 в табл. 1). Отметим, что табл. 2 и 3 описывают частный случай структуры словарной статьи, в которой есть только одна фразема и/или одна идиома. Для проектирования ЛГИС используются таблицы в общем виде, которые концептуально построены на тех же принципах, что и табл. 2, 3. В настоящей статье не представляется возможным продемонстрировать таблицы в общем виде из-за их большого размера.

Поля связи словарной статьи с корпусом

Интеграция электронного словаря с корпусом текстов реализуется за счет полей связи словарной статьи с корпусом, перечисленных в табл. 1–3. Опишем их функциональное назначение и проиллюстрируем с помощью примеров. Ниже поля связи перечислены в порядке упоминания шести их видов в табл. 1–3, также указаны способы, реализующие связи словаря с корпусом.

1. **Поле связи для перехода от заглавного слова** (см. табл. 1) к тем фрагментам текстов корпуса, которые содержат любую словоформу заглавного слова: связь реализуется с помощью поиска по лемме только для немецких заглавных слов.

Например, использовав поиск по лемме для глагола *sollen,* можно перейти к фрагментам корпуса, которые содержат любую из его словоформ. В примере (1) содержится аналитическая форма *hätte sollen*, в примере (2) – словоформа *soll*.

(1) Ich hätte sein Angebot, mir Kaffee zu kochen, annehmen und ihn noch etwas festhalten sollen. [Heinrich Böll. Ansichten eines Clowns (1963)]

Надо было принять его предложение – пусть бы сварил мне кофе и еще немного посидел. [Пер. Л.Б. Черная. Глазами клоуна (1964)]

(2) Man soll dort auch passabel essen. [Erich Maria Remarque. Der schwarze Obelisk (1956)]

Говорят, там кормят довольно прилично. [Пер. В. Станевич. Чёрный обелиск (1961)]

2. **Поле связи для перехода от словарного описания значения ФЯЕ** (см. табл. 2) к тем фрагментам текстов корпуса, которые содержат любую словоформу ФЯЕ в этом значении: связь реализуется через таблицу АПС, сформированных для всех ФЯЕ.

Приведем примеры АПС, сформированных для модального глагола *sollen*. Ниже в сокращенной форме (без примеров и комментариев) представлены словарные описания трех из 14 значений глагола *sollen* (описания других значений не могут быть представлены ввиду ограничений статьи по объему):

– словарное значение 1: осуществление какого-л. действия по чьему-л. указанию, по закону, по правилам, нормам и т.п., а также по собственному желанию, ‘до́лжен’, ‘сле́дует’; ‘сто́ит’, ‘хоте́лось бы’;

– словарное значение 2: указание на несостоявшееся действие ‘сле́довало бы’, ‘сто́ило бы’, ‘ну́жно бы́ло бы’, ‘до́лжен был бы (по мнению говорящего)’;

– словарное значение 3: указание на получение информации от других лиц; при переводе на русский язык предложение начинают словами ‘говоря́т’, ‘полага́ют’ и т.п.

Таблица 5

Примеры АПС с глаголом *sollen* в словарном значении № 3

Фрагмент оригинала	**Модальный глагол в оригинале и признаки его употребления**	**Фрагмент перевода**	**Вариант перевода и признаки его употребления**
Man **soll** dort auch passabel *essen*.	**sollen** <man> <Praes> <+Inf> <sollen-03>	**Говорят**, там *кормят* довольно прилично.	**говорят**
In einem Glas Wassers **sollen** neuerdings ganz kleine Tierchen *schwimmen*, die man früher nicht gesehen hat;	**sollen** <3pl> <Praes> <+Inf> <sollen-03>	В стакане воды, **дескать**, *плавают* малюсенькие зверушки, которых раньше никто не видел;	**дескать**

В табл. 4 и 5 представлены три АПС с разными словоформами глагола *sollen* во втором и третьем значениях. Так, в табл. 4 дано АПС с глаголом *sollen* (аналитическая форма *hätte sollen*) во втором его значении с описанием элементов АПС.

В табл. 5 даны два АПС с глаголом *sollen* в третьем значении. В первом примере словоформа *soll* (3 лицо единственного числа глагола *sollen* с неопределенным местоимением <man>, <Praes> – настоящее время) переведена на русский язык с помощью вводного слова «говорят». Во втором примере словоформа *sollen* (<3pl> – третье лицо множественного числа глагола *sollen*; <Praes> – настоящее время) переведена с помощью частицы «дескать». В данных примерах для вариантов перевода не было выявлено признаков употребления, значимых для анализа значений глагола *sollen.*

Рисунок 1 иллюстрирует связь № 2 для перехода от словарного описания значения ФЯЕ к текстам корпуса. На нем дано графическое представление связи между словарным описанием третьего значения

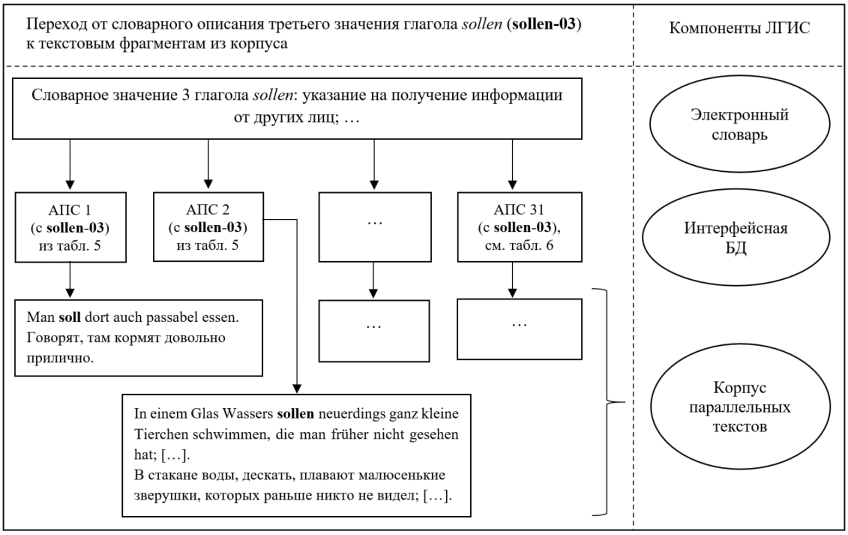

Рис. 1. Схема связи для перехода от словарного описания третьего значения глагола *sollen* к корпусу

глагола *sollen* и текстовыми фрагментами из корпуса с глаголом *sollen* в данном значении. На уровне компонентов ЛГИС эта связь проходит от электронного словаря (описание третьего значения глагола *sollen*) через интерфейсную БД (сформированные АПС 1 и 2 из табл. 5) к корпусу параллельных текстов (соответствующие пары фрагментов текстов оригинала и перевода). Таким образом, используется двойной переход от словарного значения к АПС, а затем от АПС к текстовым фрагментам из корпуса.

3. **Поле связи для перехода от фраземы** (см. табл. 2) к тем текстовым фрагментам корпуса, которые содержат фразему в любой форме без учета ее значения (с помощью синтаксического поискового запроса): связь будет реализована по аналогии с полем связи № 5.

4. **Поле связи для перехода от словарного описания значения фраземы** (см. табл. 2) к тем фрагментам текстов корпуса, которые содержат какую-либо форму фраземы в конкретном значении (через таблицу АПС, сформированных для всех ФЯЕ): связь будет реализована по аналогии с полями связи № 2 и 6.

5. **Поле связи для перехода от идиомы** (см. табл. 3) к тем фрагментам текстов корпуса, которые содержат идиому в любой форме без учета ее значения (с помощью поискового запроса).

Например, словарная статья заглавного слова *Mann* содержит описание двух значений идиомы *an den Mann brí́ngen (etw. A)*: 1) ‘найти́ покупа́теля (на что-л.)’, ‘прода́ть (что-л.)’, ‘сбыть с рук (что-л.)’; 2) ‘рассказа́ть, донести́ до слу́шателя (что-л. – анекдот и т.п.)’. Синтаксический поиск форм идиомы *an den Mann bringen* позволит перейти к тем фрагментам корпуса (примеры 3 и 4), которые содержат форму идиомы *(brachte an den Mann; an den Mann bringen)* без учета ее значения.

3) Er warb tatsächlich um Mitleid mit den Fundsachen, und er brachte Thermoskannen und Prinz-Heinrich-Mützen an den Mann […]. [Siegfried Lenz. Fundbüro (2003)]

Он действительно добивался сочувствия к потерянным вещам и удачно сбыл кому-то чайники-термосы и фуражки а-ля принц Генрих, в каких обычно ходят рыболовы […]. [Пер. Г. Косарик. Бюро находок (2004)]

4) Er stand brüsk auf. „Ich hoffe, Sie werden Gelegenheit finden, Ihren feinen Witz an den Mann zu bringen“, sagte er. [Dieter Noll. Die Abenteuer des Werner Holt. Roman einer Heimkehr (1960–1963)]

– Надеюсь, – Хольт порывисто встал, – вам представится возможность рассказать ваш пикантный анекдот кому следует. [Пер. Е. Закс, Н. Ман. Приключения Вернера Хольта (1964)]

6. **Поле связи для перехода от описания значения идиомы** (см. табл. 3) к тем фрагментам текстов корпуса, которые содержат какую-либо форму идиомы в конкретном значении (через таблицу АПС, сформированных для всех ФЯЕ).

Например, с помощью таблицы АПС будет реализован переход от первого значения идиомы *an den Mann bringen* (‘найти́ покупа́теля (на что-л.)’, ‘прода́ть (что-л.)’, ‘сбыть с рук (что-л.)’) к примеру (3), в котором анализируемая идиома переведена с помощью глагола «сбыть». От второго значения идиомы (‘рассказа́ть’, ‘донести́ до слу́шателя (что-л. – анекдот и т.п.)’) будет обеспечен переход к примеру (4), в котором идиома переведена с помощью глагола «расска-

зать». Данный тип связи будет реализован с применением тех же программных принципов, которые лежат в основе связи словарного описания значений модальных глаголов с фрагментами из корпуса (поле связи № 2).

Шесть видов перечисленных полей связи для зон и полей словарной статьи соотнесены в табл. 1–3 с тремя способами интеграции электронного словаря с корпусом параллельных текстов: 1) через поиск по лемме заглавного слова, 2) через таблицу АПС и 3) с помощью поискового запроса форм фраземы или идиомы без учета их значений.

Таблица 6

Распределение АПС по словарным значениям глагола *sollen*

№ значения глагола *sollen* в статье электронного словаря	**Число АПС с глаголом *sollen* в указанном значении**
1	420
2	81
3	31
4	38
5	16
6	0[1]
7	28
8	34
9	63
10	23
11	34
12	4
13	1
14	0
Всего АПС:	773

[1] На момент написания статьи не было сформировано АПС для значений 6 и 14 глагола *sollen*, что может быть связано с низкой частотностью его употребления в данных значениях. Дальнейшая работа с ЛГИС предполагает увеличение числа АПС, что может привести к обнаружению соответствующих контекстов.

Первый способ интеграции электронного словаря с корпусом обеспечивается для всех заглавных слов на немецком языке, представленных в разработанном электронном немецко-русском словаре (всего в словаре содержатся около 40 000 словарных статей), и реализуется через поиск по лемме заглавного слова. Примеры (1) и (2) иллюстрируют результат поиска таких пар, где какой-либо из фрагментов содержит словоформу искомого слова *sollen*.

Второй способ интеграции реализуется с помощью таблиц АПС с описанием конкретных значений ФЯЕ. Примеры АПС для значений модального глагола *sollen* представлены в табл. 4 и 5. Всего в ЦКП «Информатика» ФИЦ ИУ РАН было сформировано 773 АПС (см. табл. 6) для глагола *sollen* в разных его значениях (описания первых трех из 14 словарных значений *sollen* представлены выше).

Третий способ интеграции электронного словаря с корпусом реализуется с помощью синтаксического поискового запроса форм фразем и идиом. Примеры (3) и (4) иллюстрируют результат поискового запроса идиомы *an den Mann bringen* в разных формах.

Заключение

В статье описаны шесть видов полей и три способа связи электронного словаря с корпусом текстов в рамках создаваемой лексикографической информационной системы.

Поля связи используются для перехода к текстам корпуса: 1) от заглавного слова, 2) от значения фокусной языковой единицы, 3) от фраземы, 4) от отдельных значений фраземы, 5) от идиомы и 6) от отдельных значений идиомы. Эти связи шести видов реализуются тремя способами интеграции электронного словаря с корпусом параллельных текстов: 1) через поиск по лемме заглавного слова, 2) через таблицу аннотированных переводных соответствий и 3) с помощью поискового запроса форм фраземы или идиомы без учета их значений.

Разрабатываемая лексикографическая информационная система концептуально предусматривает расширение перечня фокусных языковых единиц, а также наращивание числа сформированных аннотированных переводных соответствий, которые, с одной стороны, иллюстрируют зафиксированные в словаре значения фокусных языковых единиц, а с другой стороны, предоставляют эмпирический материал для пополнения словарных статей новыми значениями и примерами.

Список источников

1. *Немецко-русский* словарь актуальной лексики: около 50 000 лексических единиц / под общ. рук. Д.О. Добровольского. М. : Азбуковник, 2025 (в печати).
2. *Большой* немецко-русский словарь = Das Grosse Deutsch-Russische Woerterbuch: в 2 т. / сост. Е.И. Лепинг и др., под рук. О.И. Москальской. М. : Русский язык, 1980.
3. *Новый* большой немецко-русский словарь / под общ. рук. Д.О. Добровольского. В 3 т.: около 500 000 лексических единиц. М. : АСТ, Астрель, 2008–2010.
4. *Duden.* Das große Wörterbuch der deutschen Sprache in zehn Bänden. 3., völlig neu bearb. und erw. Aufl. Mannheim etc. : Dudenverl., 1999.
5. *Duden* online. URL: http://www.duden.de/ (дата обращения: 18.06.2025).
6. *DWDS-Wörterbuch*. URL: https://www.dwds.de/d/woerterbuecher (дата обращения: 18.06.2025).
7. *Добровольский Д.О., Кретов А.А., Шаров С.А.* Корпус параллельных текстов // Научно-техническая информация. Сер. 2: Информационные процессы и системы. 2005. № 6. С. 16–27.
8. *Добровольский Д.О.* Корпусный подход к исследованию фразеологии: новые результаты по данным параллельных корпусов // Вестник Санкт-Петербургского университета. Язык и литература. 2020. Т. 17, вып. 3. С. 398–411.
9. *Национальный* корпус русского языка. URL: https://ruscorpora.ru/ (дата обращения: 06.04.2025).
10. *Кружков М.Г.* Концепция построения надкорпусных баз данных // Системы и средства информатики. 2021. Т. 31, № 3. С. 101–112.

11. *Постановление* Правительства РФ от 25 декабря 2024 г. № 1892 «О Национальном словарном фонде». URL: http://government.ru/docs/53894/ (дата обращения: 06.04.2025).
12. *Плунгян В.А., Рахилина Е.В.* О цифровой лексикографии // Труды Института русского языка им. В.В. Виноградова. 2025. № 1 (43). С. 360–366.
13. *Ольховская А.И.* Корпуса на службе у лексикографии: применение корпусных технологий при составлении словарей // Русский язык за рубежом. 2023. № 1. С. 77–83.
14. *Активный* словарь русского языка / отв. ред. Ю.Д. Апресян. М. : Языки славянской культуры, 2014. Т. 1. 408 с.
15. *Толковый* словарь русской разговорной речи. Вып. 1–4 / отв. ред. Л.П. Крысин. М. : Языки славянской культуры, 2014–2021.
16. *Ляшевская О.Н., Шаров С.А.* Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М. : Азбуковник, 2009. 1087 с.
17. *Collocations*, Colligations, and Corpora (CoCoCo). URL: https://cosyco.ru/cococo/ (дата обращения: 06.04.2025).
18. *Ooi V.B.Y.* Computer corpus lexicography. Edinburgh : Edinburgh University Press, 1998. 224 p.
19. *Frankenberg-Garcia A., Rees G., Lew R.* Slipping through the Cracks in e-Lexicography // International Journal of Lexicography. 2021. № 34 (2). P. 206–234.
20. *Rees G.* Using corpora to write dictionaries // The Routledge handbook of corpus linguistics / ed. by A. O'Keeffe, M. McCarthy. London; New York : Routledge, 2022. P. 387–404.
21. *Zufferey S.* Introduction à la linguistique de corpus. Collection : Sciences. ISTE Group, 2020. 252 p.
22. *Добровольский Д.О., Зацман И.М.* Модель извлечения знания из параллельных текстов лексикографической информационной системы // Информатика и её применения. 2024. Т. 18, вып. 3. С. 97–105.
23. *Digitales* Wörterbuch der deutschen Sprache. URL: https://www.dwds.de (дата обращения: 06.04.2025).
24. *Klein W., Geyken A.* Das Digitale Wörterbuch der Deutschen Sprache (DWDS) // Lexicographica. 2010. Vol. 26, № 2010. P. 79–96.
25. *Geyken A., Wiegand F., Würzner K.-M.* On-the-fly Generation of Dictionary Articles for the DWDS Website // Electronic Lexicography in the 21st Century. Proceedings of eLex 2017 conference / ed. by I. Kosem, C. Tiberius. Leiden, the Netherlands : Lexical Computing, 2017. P. 560–570.

26. *Добровольский Д.О., Зацман И.М.* Интеграция электронного словаря с текстами параллельного корпуса: новый теоретический подход // Системы и средства информатики. 2025. Т 35, вып. 1. С. 111–124.
27. *Oxford* English Dictionary. URL: https://www.oed.com/ (дата обращения: 06.04.2025).
28. *Полухина П.А.* Oxford English Dictionary online: подготовка к третьему изданию словаря на примере Updates 2016 // Известия ВГПУ. 2018. № 8 (131). С. 136–144.
29. *Merriam* Webster Dictionary. URL: https://www.merriam-webster.com/ (дата обращения: 06.04.2025).
30. *Гончаров А.А., Добровольский Д.О., Зализняк А.А.* База данных конструкций с немецкими модальными глаголами и их русских соответствий // Труды междунар. конф. «Корпусная лингвистика-2023». СПб. : Изд-во СПбГУ, 2024. С. 51–60.

References

1. Dobrovol'skiy, D.O. (2025) *Nemetsko-russkiy slovar' aktual'noy leksiki: okolo 50 000 leksicheskikh edinitsz* [German-Russian Dictionary of Current Vocabulary: About 50,000 Lexical Units]. Moscow: Azbukovnik. [in print].
2. Leping, E.I. et al. (1980) *Bol'shoy nemetsko-russkiy slovar' = Das Grosse Deutsch-Russische Woerterbuch: v 2 t.* [Large German-Russian Dictionary = Das Grosse Deutsch-Russische Wörterbuch: in 2 vols]. Moscow: Russkiy yazyk.
3. Dobrovol'skiy, D.O. (2008–2010) *Novyy bol'shoy nemetsko-russkiy slovar'. V 3 t.: okolo 500 000 leksicheskikh edinits* [New Large German-Russian Dictionary: in 3 vols. About 500,000 Lexical Units]. Moscow: AST, Astrel'.
4. Duden. (1999) *Das große Wörterbuch der deutschen Sprache in zehn Bänden*. 3., völlig neu bearb. und erw. Aufl. Mannheim etc.: Dudenverl.
5. Duden online. (n.d.) *Duden online*. [Online] Available from: http://www.duden.de/ (Accessed: 18.06.2025).
6. DWDS-Wörterbuch. (n.d.) *DWDS-Wörterbuch*. [Online] Available from: https://www.dwds.de/d/woerterbuecher (Accessed: 18.06.2025).
7. Dobrovol'skiy, D.O., Kretov, A.A. & Sharov, S.A. (2005) Korpus parallel'nykh tekstov [Corpus of parallel texts]. *Nauchno-tekhnicheskaya informatsiya. Ser. 2: Informatsionnye protsessy i sistemy* [Scientific and Technical Information. Ser. 2: Information Processes and Systems]. 6. pp. 16–27.

8. Dobrovol'skiy, D.O. (2020) Korpusnyy podkhod k issledovaniyu frazeologii: novye rezul'taty po dannym parallel'nykh korpusov [Corpus-based approach to phraseology research: new results from parallel corpora]. *Vestnik Sankt-Peterburgskogo universiteta. Yazyk i literatura*. 17 (3). pp. 398–411.
9. *Natsional'nyy korpus russkogo yazyka* [Russian National Corpus]. (2025) [Online] Available from: https://ruscorpora.ru/ (Accessed: 06.04.2025).
10. Kruzhkov, M.G. (2021) Kontseptsiya postroeniya nadkorpusnykh baz dannykh [Concept of building supracorpus databases]. *Sistemy i sredstva informatiki*. 31 (3). pp. 101–112.
11. Russian Federation. (2024) *Decree of the Government of the Russian Federation of December 25, 2024 No. 1892 "On the National Dictionary Fund"*. [Online] Available from: http://government.ru/docs/53894/ (Accessed: 06.04.2025). (In Russian).
12. Plungyan, V.A. & Rakhilina, E.V. (2025) O tsifrovoy leksikografii [On digital lexicography]. *Trudy Instituta russkogo yazyka im. V.V. Vinogradova*. 1 (43). pp. 360–366.
13. Ol'khovskaya, A.I. (2023) Korpusa na sluzhbe u leksikografii: primenenie korpusnykh tekhnologiy pri sostavlenii slovarey [Corpora in the service of lexicography: Application of corpus technologies in dictionary compilation]. *Russkiy yazyk za rubezhom*. 1. pp. 77–83.
14. Apresyan, Yu.D. (ed.) (2014) *Aktivnyy slovar' russkogo yazyka* [Active Dictionary of the Russian Language]. Vol. 1. Moscow: Yazyki slavyanskoy kul'tury.
15. Krysin, L.P. (ed.) (2014–2021) *Tolkovyy slovar' russkoy razgovornoy rechi* [Explanatory Dictionary of Russian Colloquial Speech]. Iss. 1–4. Moscow: Yazyki slavyanskoy kul'tury.
16. Lyashevskaya, O.N. & Sharov, S.A. (2009) *Chastotnyy slovar' sovremennogo russkogo yazyka (na materialakh Natsional'nogo korpusa russkogo yazyka)* [Frequency Dictionary of Modern Russian (based on the Russian National Corpus)]. Moscow: Azbukovnik.
17. *Collocations, Colligations, and Corpora (CoCoCo)*. (2025). [Online] Available from: https://cosyco.ru/cococo/ (Accessed: 06.04.2025).
18. Ooi, V.B.Y. (1998) *Computer corpus lexicography*. Edinburgh: Edinburgh University Press.
19. Frankenberg-Garcia, A., Rees, G. & Lew, R. (2021) Slipping through the Cracks in e-Lexicography. *International Journal of Lexicography*. 34 (2). pp. 206–234.
20. Rees, G. (2022) Using corpora to write dictionaries. In: O'Keeffe, A. & McCarthy, M. (eds) *The Routledge handbook of corpus linguistics*. London and New York: Routledge. pp. 387–404.

21. Zufferey, S. (2020) *Introduction à la linguistique de corpus.* Collection: Sciences. ISTE Group.
22. Dobrovol'skiy, D.O. & Zatsman, I.M. (2024) Model' izvlecheniya znaniya iz parallel'nykh tekstov leksikograficheskoy informatsionnoy sistemy [A model for knowledge extraction from parallel texts for a lexicographic information system]. *Informatika i eyo primeneniya*. 18 (3). pp. 97–105.
23. DWDS. (2025) *Digitales Wörterbuch der deutschen Sprache*. [Online] Available from: https://www.dwds.de (Accessed: 06.04.2025).
24. Klein, W. & Geyken, A. (2010) Das Digitale Wörterbuch der Deutschen Sprache (DWDS). *Lexicographica*. 26 (2010). pp. 79–96.
25. Geyken, A., Wiegand, F. & Würzner, K.-M. (2017) On-the-fly Generation of Dictionary Articles for the DWDS Website. In: Kosem, I. & Tiberius, C. (eds) *Electronic Lexicography in the 21st Century. Proceedings of eLex 2017 conference*. Leiden, the Netherlands: Lexical Computing. pp. 560–570.
26. Dobrovol'skiy, D.O. & Zatsman, I.M. (2025) Integratsiya elektronnogo slovarya s tekstami parallel'nogo korpusa: novyy teoreticheskiy podkhod [Integration of an electronic dictionary with parallel corpus texts: a new theoretical approach]. *Sistemy i sredstva informatiki*. 35 (1). pp. 111–124.
27. *Oxford English Dictionary*. (2025). [Online] Available from: https://www.oed.com/ (Accessed: 06.04.2025).
28. Polukhina, P.A. (2018) Oxford English Dictionary online: podgotovka k tret'emu izdaniyu slovarya na primere Updates 2016 [Oxford English Dictionary online: preparation for the third edition of the dictionary using the example of Updates 2016]. *Izvestiya VGPU*. 8 (131). pp. 136–144.
29. *Merriam Webster Dictionary*. (2025). [Online] Available from: https://www.merriam-webster.com/ (Accessed: 06.04.2025).
30. Goncharov, A.A., Dobrovol'skiy, D.O. & Zaliznyak, A.A. (2024) [Database of constructions with German modal verbs and their Russian equivalents]. *Korpusnaya lingvistika-2023* ["Corpus Linguistics-2023"]. Proceedings of the International Conference. Saint Petersburg: St. Petersburg State University. pp. 51–60. (In Russian).

Сведения об авторе:

Егорова Анна Юрьевна – научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук (Москва, Россия). E-mail: anna.yu.egorova@yandex.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Anna Yu. Egorova, researcher, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (Moscow, Russian Federation).
E-mail: anna.yu.egorova@yandex.ru

The author declares no conflicts of interests.

Статья поступила в редакцию 29.04.2025;
одобрена после рецензирования 13.08.2025; принята к публикации 04.09.2025

The article was submitted 29.04.2025;
approved after reviewing 13.08.2025; accepted for publication 04.09.2025