

ИНТЕРНЕТ-ПРОЕКТЫ ИНСТИТУТА ЧЕШСКОГО ЯЗЫКА АН ЧР, ПОСВЯЩЕННЫЕ ИСТОРИЧЕСКОЙ ЛЕКСИКОГРАФИИ

Петр НЕЙЕДЛЫ

Институт чешского языка Академии наук ЧР
Ústav pro jazyk český Akademie věd ČR
(Прага, Чешская Республика)
<https://orcid.org/0000-0001-5012-4632>
nejedly@ujc.cas.cz

Штепан ШИМЕК

Институт чешского языка Академии наук ЧР
Ústav pro jazyk český Akademie věd ČR
(Прага, Чешская Республика)
<https://orcid.org/0000-0002-2963-3341>
simek@ujc.cas.cz

Милослава ВАЙДЛОВА

Институт чешского языка Академии наук ЧР
Ústav pro jazyk český Akademie věd ČR
(Прага, Чешская Республика)
<https://orcid.org/0000-0002-2446-1659>
vajdlova@ujc.cas.cz

Статья представляет три коллективных проекта, реализованных отделом развития языка Института чешского языка Академии наук ЧР и посвященных, в первую очередь, проблемам лексикографии: «Электронный словарь древнечешского языка», «Сетевой словотолковник» и «Лексическая база данных чешского языка эпохи Гуманизма и Барокко». Каждый из этих проектов позволяет широкой публике познакомиться с научной информацией об истории чешского языка и может быть полезен для славистов, занимающихся диахронической лингвистикой.

Ключевые слова: лексикология, лексикография, древнечешский язык, старочешский язык, словарь, база данных, интернет-портал

INTERNET LEXICOGRAPHIC PROJECTS OF THE CZECH LANGUAGE INSTITUTE CAS, FOCUSING ON HISTORICAL CZECH

Petr NEJEDLÝ

Czech Language Institute CAS
(Prague, Czech Republic)

Štěpán ŠIMEK

Czech Language Institute CAS
(Prague, Czech Republic)

Miloslava VAJDLOVÁ
Czech Language Institute CAS
(Prague, Czech Republic)

This article presents three collective projects implemented by the Department of Language Development of the Czech Language Institute of the Czech Academy of Sciences, focusing primarily on problems of lexicography: *Electronic Dictionary of Old Czech*, *Web Vocabular and Lexical Database of the Czech Language of the Humanist and Baroque Period*. Each of these projects allows the general public to get acquainted with scientific information about the history of the Czech language and may be useful for Slavists dealing with diachronic linguistics.

Keywords: lexicology, lexicography, Old Czech, Middle Czech, dictionary, database, web portal

Введение¹

Отдел развития языка Института чешского языка Академии наук Чешской Республики (Ústav pro jazyk český AV ČR; ÚJČ) был основан в 1948 году (Dvořáčková 2011: 160) с целью «комплексной научной реконструкции словарного состава чешского языка на основе памятников, возникших с первых лет письменности до конца XV века» (Navránek 1968: 7). Результатом этой работы стал большой академический «Словарь древнечешского языка» (*Staročeský slovník*, далее StČS 1968–2008). За более чем семьдесят лет существования отдела область его исследований расширилась за пределы словарного состава Средних веков. Тем не менее, изучение лексики и создание лексикографических трудов и сегодня остаются его приоритетами.

Наша статья знакомит читателей с тремя коллективными проектами, посвященными, в первую очередь, проблемам лексикографии. Первый проект посвящен «Электронному словарю древнечешского языка» (*Elektronický slovník staré češtiny*, далее ESSČ); второй — «Сетевому словотолковнику» (*Vokabulář webový*, далее VW), т.е. платформе, на которой представлен не только вышеупомянутый словарь, но и ряд других материалов, предназначенных для всех, кто занимается историей чешского языка; третий — «Лексической базе данных чешского языка эпохи Гуманизма и Барокко» (*Lexikální databáze humanistické a barokní češtiny*, далее LDHВČ). Каждый из этих проектов позволяет широкой публике ознакомиться с научной информацией об истории чешского языка и может быть полезен не только для богемистов, но и для славистов, занимающихся диахронической лингвистикой.

¹ Данная статья создана при поддержке Проекта долгосрочного концептуального развития Института чешского языка Академии наук ЧР RVO: 68378092 с использованием данных и инструментов, разработанных в рамках исследовательской инфраструктуры LINDAT/CLARIAH-CZ (<https://lindat.cz>), спонсируемой Министерством образования, дел молодежи и спорта Чешской Республики (проект № LM2018101).

Электронный словарь древнечешского языка — основная характеристика

ESSČ — это одноязычный исторический толковый интернет-словарь, отражающий словарный состав чешского языка от начала письменности до конца XV века. Призванный познакомить пользователей с научным описанием древнечешской лексики², он создан на базе материала, хранящегося в лексической картотеке StčS, которая содержит около полутора миллионов записей (Homolková 2007: 225–226).

Работа над ESSČ ведется с 2005 года, когда руководство ÚJČ по причине чрезмерных временных затрат остановило проект StčS. Главной задачей ESSČ было в кратчайшие сроки опубликовать статьи, посвященные лексемам алфавитного раздела от *při-* до *ž*, на тот момент не включенным в имеющиеся словари. Позже было принято решение отказаться от дифференцированного подхода и алфавитных ограничений в пользу более глубокого лексического анализа. В настоящее время в ESSČ представлены результаты трех моделей анализа и описания древнечешской лексики, применявшихся на каждом из этапов развития словаря и отразившихся в его макро- и микроструктуре (см. ниже, ср. Vajdllová 2018).

Принципы составления

На всех этапах развития ESSČ подход к древнечешскому словарному составу оставался неизменным. В соответствии с теорией Пражской лингвистической школы и практикой StčS авторы ESSČ понимают словарный запас как структурированную систему с центром и периферией, в которой одни элементы образуют связи с другими (TLP 1966), а лексему как внутренне структурированную единицу, обладающую уникальными формальными и содержательными свойствами (StčS 1968: 20–35). Концепция ESSČ в общих чертах и в упрощенном виде повторяет концепцию StčS (StčS 1968: 16–49, Nejedlý et al. 2002) и опирается на труды ее ведущего автора Игоря Немца (см. прежде всего Němec 1965; 1968; 1975; 1980; 2009).

Словарные статьи ESSČ в большинстве случаев содержат лексемы, не являющиеся именами собственными. Из проприальной лексики в словаре выборочно представлены лишь историко-культурные имена; остальные имена собственные фиксируются, но не сопровождаются разъясняющей информацией. Встречающиеся в памятниках иноязычные вкрапления в ESSČ не включаются.

Многозначное слово рассматривается как единая лексическая единица и характеризуется одной статьей; омонимы анализируются по отдельности.

Как единая лексема воспринимаются также морфологические варианты слова³, основообразующие варианты глаголов и существительных⁴, а также семейства глаголов с приставками *s-*, *z-*, *vz-*, у которых на основании значения невозможно

² См. <https://vokabular.ujc.cas.cz/informace.aspx?t=ESSC&o=slovniky>.

³ Например, *kadeřě f.*, *kadeř m.*

⁴ Например, *bósti*, позже *bodnúti*; *rámě*, позже *rameno*.

однозначно идентифицировать приставку и/или определить, идет ли речь о разных лексемах с отличной структурой значения или нет⁵.

В отличие от некоторых других диахронических словарей, ESSČ не объединяет в одной статье словообразовательные варианты или словообразовательно близкие лексемы одного семейства. Исключением из этого правила являются прилагательные, заканчивающиеся на *-ní*, *-ný*, *-ovní*, *-ovný*⁶, которые, ввиду особенностей чешской рукописной орфографии, нельзя с уверенностью отличить друг от друга. Кроме того, в словаре не применяется гнездовой принцип расположения материала.

Заглавные слова в соответствии с практикой современной чешской исторической лексикографии (*Slovník staročeský* Яна Гебауэра (далее GbSlov), StČS 1968–2008) реконструированы в предполагаемой фонетической форме начала XIV века (около 1300 года). В заглавии статьи фиксируются также фонетические (но не орфографические) варианты лексемы, а также приводится ее основная морфологическая характеристика. Кроме того, ESSČ отслеживает развитие формы и морфологических характеристик описываемых слов.

При толковании лексемы учитывается предполагаемое развитие ее значения. Отдельные значения разъясняются как современными чешскими эквивалентами, так и с помощью описательного толкования, в котором подчеркиваются важнейшие смысловые аспекты понятия. При определении значения слова внимание уделяется его компонентному анализу, а также валентности и сочетаемости. Словарное определение в обусловленных случаях включает в себя стилистическую характеристику, морфологические ограничения и возможные изменения части речи. Лексикализованные словосочетания рассматриваются отдельно.

При необходимости словарные статьи завершаются примечаниями, содержащими ссылки на системно близкие лексемы (прежде всего синонимы и антонимы), комментарии к приведенным толкованиям или релевантную информацию в специальной литературе.

Характер структурных элементов ESSČ зависит от того, на каком из этапов работы они возникли.

До 2019 года ESSČ создавался под руководством Милославы Вайдловой; с 2019 года проект ведет Штепан Шимек.

ESSČ как авторский словарь без примеров, создаваемый по дифференциальному принципу в сопоставлении с *Малым древнечешским словарем*⁷ (2005–2010)

Возникший в 2005 году ESSČ продолжал алфавитный порядок StČS, последним заглавным словом которого было *pří*. Перед коллективом авторов встала задача в течение пяти лет составить описание лексических единиц от *pří*- до *ž*. Учитывая

⁵ Например, *zčěžený/scěžený* ‘счеженный, прожеженный, счеживанием избавленный от примесей и нечистот’.

⁶ Например, *lesní, lesný; duchovní, duchovný*.

⁷ *Malý staročeský slovník*, далее MSS. MSS — дифференциальный лексикографический справочник, составленный в первую очередь для студентов вузов.

масштабы работы, было принято решение исключить из будущего ESSČ лексемы, содержащиеся в MSS⁸. Поскольку основным приоритетом стало выполнение работы в установленный срок, составители были вынуждены значительно упростить описание лексических единиц и сократить объем статей по сравнению с StčS: были опущены мотивации слов, перечисление их валентностей и иллюстрации отдельных значений примерами из средневековых памятников. Кроме того, помимо имен собственных из словаря были исключены и образованные от них лексемы. Каждая статья считалась авторской работой, ее содержательная часть не подвергалась редактуре или унификации по установленному образцу; публикации предшествовала лишь проверка формальных характеристик и грамматической информации.

Различие задач и подходов, которыми руководствовались составители ESSČ и MSS, неизбежно привело к несоответствиям и противоречиям в лексикографическом описании упомянутого раздела древнечешского словарного запаса. В MSS форма заглавного слова не всегда реконструирована к началу XIV века, поскольку лексемы приведены в форме, зафиксированной в древнечешских памятниках (MSS: 6). В результате некоторые слова или их семейства, встречающиеся в разных, не связанных друг с другом формах, оказались включены как в MSS, так и в ESSČ⁹. Помимо этого, MSS нередко рассматривает основообразующие варианты лексем с одним значением или других родственных слов в рамках одной словарной статьи.

Однако важнейшее отличие между обоими словарями заключалось в дифференциальном подходе MSS к выбору лексем и их толкований. Словарные статьи MSS, как правило, указывают лишь некоторые значения слов — и главным образом (а порой и исключительно) лексем, значение которых в современном чешском изменилось (MSS: 6)¹⁰. Словарные определения при этом не всегда точны, а в некоторых случаях и вовсе отсутствуют. ESSČ, напротив, подает подробную информацию о лексических единицах, вследствие чего бросается в глаза диспропорция в описаниях отдельных семейств, включенных в оба словаря.

На первом этапе работы ESSČ было проанализировано 51 500 статей рабочего словника и опубликовано 17 500 полных словарных статей (Vajdlová 2011: 29).

ESSČ как авторский словарь без примеров, описывающий весь словарный состав языка от А до М (2010–2016)

Вышеперечисленные проблемы и диспропорция между толкованиями лексем в MSS и ESSČ вызвали необходимость изменить подход к форме и объему статей при описании следующего алфавитного раздела. На втором этапе работы были в полной мере характеризованы все идентифицированные древнечешские лек-

⁸ MSS также опубликован на портале VW (см. ниже).

⁹ Ср., например, статьи MSS *strávený*, *vezda/vežda*, *vezdajší/vezdajší* и статьи ESSČ *ztrávený*², *vešda*, *vešdajší*.

¹⁰ Например, статьи *psáti*, *řezaný*, *váha*, *vzdáti*.

сические единицы, начинающиеся на буквы от *A* до *M*¹¹. Опираясь на предыдущий опыт, авторы словаря отказались от идеи комбинирования концептуально несовместимых ESSČ и MSS. Кроме того, были составлены новые описания всех лексем, представленных в «Древнечешском словаре» Яна Гебауэра (GbSlov), созданного по методологии своего времени, значительно отличающейся от сегодняшней, на ограниченном материале и недостаточно репрезентативной выборке; параллельные лексемы в GbSlov не всегда описаны единообразно, некоторые содержащиеся в нем слова впервые встречаются в XVI веке, а в статьях, посвященных определенному типу слов, отсутствуют современные чешские эквиваленты и/или толкования значений.

В эти годы ESSČ по-прежнему развивался как авторский словарь без примеров и редактуры, обеспечивающей единообразную подачу материала. Тем не менее, был последовательно реализован стандартный лексический анализ материала и проведено цельное толкование лексических единиц. Обязательной частью словарной статьи у мотивированных лексем стала информация о мотивирующей лексической единице или ссылка на ближайшее родственное слово.

Отсутствующие иллюстративные цитаты словарных статей должны были компенсироваться примерами GbSlov. Поскольку многие лексемы в GbSlov отсутствуют, авторы старались восполнять недостаток примеров иначе: например, в толковании значения приводили факультативную сочетаемость лексической единицы, а в примечании систематически обращали внимание на частое использование лексемы в данном конкретном памятнике и типе текстов или у определенного автора.

На данном этапе работы стало очевидно, что для создания качественного, избавленного от противоречий, единообразного описания словарного состава невозможно обойтись без редактуры и примеров, иллюстрирующих значения лексем, включенных непосредственно в статьи ESSČ. Упрощенная унифицированная редакция была опробована на разделе, содержащем слова, начинающиеся на букву *M*, который, впрочем, был представлен широкой публике все еще без примеров.

На втором этапе работы ESSČ было проанализировано более 47 400 статей рабочего словника и опубликовано около 14 000 полных словарных статей.

ESSČ как редактируемый словарь с выборочными иллюстративными примерами (с 2017 года)

Третий этап работы над ESSČ, начавшийся в 2017 году, сосредоточен на лексикографическом разборе слов от *Řib* до *Ž*, ранее описанных лишь в MSS. Слишком короткая и упрощенная форма словарных статей, ставшая результатом подхода, применявшегося на первом этапе, равно как и потребность в унифицированной характеристике отдельных лексических семейств, ономаσιологических

¹¹ Слова, начинающиеся на букву *M*, были выбраны по той причине, что лексемы на *N* уже в полной мере были включены в StčS.

категорий, семантических и понятийных полей и т. д., содержащихся в этой части древнечешского словарного состава, вызвали необходимость ревизии исходного дифференциального описания лексики. В результате все словарные статьи данного раздела подверглись серьезной переработке (Vajdllová 2019).

От предыдущих эта фаза отличается, в первую очередь, отказом от авторского характера статей, не содержащих примеров. Теперь в статьях последовательно приводится подробное толкование значения (или значений) лексемы, словообразовательные данные, а также информация о формальном и грамматическом развитии слова. Каждое значение в обязательном порядке иллюстрируется двумя примерами из древнечешских текстов: во-первых, одним из наиболее ранних, во-вторых — репрезентативным, имеющим лексикографическую ценность и наглядно отражающим значение, а в тех случаях, когда это возможно — основную валентность и сочетаемость, контекст и жанр (Fuková et al. 2019: 7–9). Словарные статьи корректируются при помощи унифицирующей семантической и графематической редактуры.

Основная цель третьего этапа работы над ESSĎ — доведение вышеупомянутого алфавитного раздела ESSĎ до уровня дополнения к незавершенному StĎS, пусть и более сжатого, но при этом соизмеримого и качественного как в лексикологическом, так и в лексикографическом аспекте.

Описание словарного состава древнечешского языка

по завершении третьего этапа работы над ЭССЯ

(за основу взята таблица в публикации (Fuková, Pytlíková 2021: 268–269))

	<i>A–M</i>	<i>N–při</i>	<i>Přib–Ž</i>
StĎS (1968–2008)	–	полное описание, иллюстрированное примерами	–
ESSĎ (2006–)	ESSĎ, II этап подробное описание, без примеров	–	ESSĎ, III этап подробное описание, с базовыми примерами

Сетевой словотолковник — история и задачи

VW изначально предназначался для публикации ESSĎ: к разработке функций поиска и интернет-среды сотрудники ÚJĎ приступили вскоре после составления первых статей. Сайт был запущен в 2006 году под названием «Сетевой словотолковник — интернет-портал исторических словарей», что отражало исходный план использовать приложение исключительно для публикации текстовых словарей. Постепенно на платформе VW стало появляться большое количество источников разных типов, вследствие чего она получила новое название — «Сетевой словотолковник — интернет-портал источников для изучения истории чешского языка» — подробнее о создании, развитии и использовании VW см. (Ďerná 2007;

Černá 2011; Černá, Lehečka 2016; Vajdlová 2007; Dušek 2017). Помимо современных диахронических толковых словарей в VW можно найти электронные издания и цифровые копии старых словарей, электронные издания памятников чешской письменности, языковые базы данных, научную литературу, цифровые копии старопечатных грамматических справочников, аудиокниги, а также веб-инструменты, которые могут быть полезны для подготовки электронных изданий или языкового анализа текстов. Многие источники представлены не только в текстовой, но и в звуковой форме или в виде изображений. Условием размещения того или иного материала на портале является его соответствие высоким научным требованиям.

Структура и содержание VW сегодня

VW предоставляет широкой публике доступ к различным материалам, для презентации которых предусмотрены соответствующие функции. Каждая группа источников размещена на отдельной интернет-странице; для выбора одной из них достаточно открыть соответствующую закладку на главной странице VW.

На странице *Древнечешские словари* пользователь может ознакомиться с составом современных диахронических словарей: ESSČ, StčS, GbSlov, MSS и «Древнечешского словарика» Франтишека Шимека (Šimek 1947). Кроме того, в этот раздел помещен и «Словоуказатель Словаря древнечешских имен собственных» Яна Свободы¹². Требуемое слово или сочетание знаков можно искать в самостоятельно выбранных источниках или же просматривать подряд словарные статьи, расположенные по алфавиту.

На странице *Банки текстов* находится постоянно расширяющийся корпус древнечешских и старочешских памятников¹³, процент лемматизации которых постоянно растет. В основу корпуса легли электронные издания рукописей, инкунабул и старопечатных книг конца XIII — начала XIX века, подготовленные сотрудниками Отдела развития языка. Пользователь может работать с банками текстов по отдельности (в программе, созданной специально для VW) или с помощью корпусного интерфейса Института Чешского национального корпуса *Kontext*¹⁴, который позволяет анализировать все тексты одновременно, создавать подкорпусы и искать лексемы по лемме. По состоянию на январь 2022 года *Банк древнечешских текстов* содержит более 6,5 миллионов токенов, а *Банк старочешских текстов* — более одного миллиона токенов. Таким образом, корпус VW является самым крупным диахроническим корпусом чешского языка¹⁵.

¹² Подробнее о Словоуказателе см.: <https://vokabular.tjic.cas.cz/informace.aspx?t=IndexSvob&=slovníky>.

¹³ Старочешским языком считается язык памятников, созданных в 1500–1780 годы.

¹⁴ См. <https://www.korpus.cz/kontext/query?corpname=syn2020>.

¹⁵ Например, диахронический корпус Института Чешского национального корпуса в настоящее время содержит 3,5 миллиона токенов (см. <https://wiki.korpus.cz/doku.php/cnk:uvod>), при этом 54 % из них представляют тексты, написанные в XIX и XX веках (см. <https://wiki.korpus.cz/doku.php/en:cnk:diakorp>).

Большая часть текстов, представленных в банках текстов (и, соответственно, в корпусе), также доступна в форме полных транскрипций, опубликованных на странице *Издания*. В настоящий момент раздел содержит 309 памятников, каждый из которых сопровождается издательским примечанием и критическим комментарием.

Важным источником познания чешского ушедших эпох служат грамматические справочники. На странице *Грамматики* опубликовано 113 цифровых копий таких трудов, написанных в период с XV по XIX век. Их оригиналы хранятся как в фондах ÚJČ, так и в других чешских и зарубежных библиотеках. Пользователь может не только просматривать цифровые копии по отдельности, но и осуществлять в них поиск по метаязыковой информации, в первую очередь — по лингвистическим терминам, аннотация которых проведена в каждом памятнике. Все публикации сопровождаются библиографической и книговедческой характеристикой.

Наиболее значимым текстом для исследователя словарного состава чешского языка эпохи раннего национального возрождения, а также старочешского периода, служит «Немецко-чешский словарь» Йозефа Юнгманна (JgSlov). В модуле *Цифровые словари* опубликована цифровая копия экземпляра этого словаря, некогда принадлежавшего непосредственно Юнгманну (о чем свидетельствуют и его собственноручные дополнения и исправления), а сегодня хранящегося в библиотеке ÚJČ. Словарь, как и другие материалы VW, можно просматривать постранично — или искать в нем статьи по транскрипции и транслитерации заглавных слов.

Страница *Специальная литература* призвана сделать доступными для всех интересующихся ключевые научные труды, в центре которых стоят вопросы исторического развития чешского. На сегодня в разделе представлены все четыре тома «Исторической грамматики чешского языка» Яна Гебауэра (Gebauer 1929; 1958; 1960; 1963) и две монографии Игоря Немца, основные идеи которых легли в основу концепции StčS и ESSč (Němec 1968; 1980). Помимо просмотра транскрибированных версий книг, раздел дает возможность полнотекстового поиска в них. Электронное издание трудов Гебауэра дополнено цифровой копией оригинала: фотография соответствующей страницы отображается при нажатии на ее номер в транскрипции.

Страница *Аудиокниги*, предназначенная для образовательных целей и популяризации деятельности ÚJČ, предлагает для бесплатного использования аудиозаписи (в форматах WAV, MP3 и OGG) *Избранных произведений древнечешской прозы* (в том числе фрагментов *Оломоуцкой Библии*, *Книжек о шахматах* Томаша Штитного, *Ткача* и *Пассионала*), а также сатиры 1518 года *Право Франты*.

Научную ценность, не в последнюю очередь с точки зрения исторической лексикографии, представляют материалы, размещенные на странице *Источники*. В нем с помощью списка всех вышеупомянутых словарей и трудов можно перейти как к самим источникам, так и к паратекстам (например, к предисловиям или спискам сокращений). Помимо этого, в модуле представлены цифровые копии картотеки для GbSlov. Наибольший же интерес представляют электронные издания

одиннадцати словарей XVI–XIX веков — транслитерированные (а в редких случаях и транскрибированные) рукописи и старопечатные книги, среди которых — «Nomenclator tribus linguis» Даниэла Адама из Велеславина (1586), «Thesaurus Linguae Bohemicae» Вацлава Яна Росы (вторая половина XVII века) и «Deutsch-böhmisches Wörterbuch» Йозефа Добровского (1821). Поиск осуществляется исключительно в рамках отдельных словарей; *Nomenclator* Велеславина дополнен цифровыми копиями страниц оригинала.

Страница *Инструменты* содержит программы и приложения, которые могут быть полезны при подготовке изданий исторических текстов или при их анализе: электронную среду *Transcriptorium*, проверочную линейку, нестандартные знаки или приложение для проверки древнечешской орфографии в *MS Word*. Для лексикологических и лексикографических целей предназначен инструмент анализа токенов, а также приложения *Древнечешские образцы* и *Древнечешские формы слов*. Программа *DocxToXml* помогает перевести текст в формат xml. Все инструменты доступны для скачивания с исходными кодами, что позволяет приспособить их к различным языковым средам и потребностям пользователей.

VW постоянно растет и развивается; в настоящее время ведется подготовка новой версии портала, которая будет отличаться устройством и структурой, обеспечивающей пользователям и авторам не только более быстрый поиск, но и возможность одновременно работать с источниками, которые ранее были доступны лишь по отдельности. Так, например, современные диахронические словари планируется объединить с историческими, благодаря чему станет возможным более эффективное изучение развития словарного состава языка на основе описаний прошлых веков, сопоставленных с научной рефлексией Новейшего времени.

Лексическая база данных чешского языка эпохи Гуманизма и Барокко. Основная характеристика

LDHВЃ отражает словарный состав чешского языка в эпоху раннего Нового времени. База данных содержит лексические единицы, зафиксированные в памятниках письменности и представленные в аутентичном контексте их употребления¹⁶.

Мотивы создания базы данных

Основным импульсом к появлению LDHВЃ стало отсутствие репрезентативного словаря чешского языка XVI–XVIII веков. Словарный состав этого периода частично описан в JgSlov, включающем в себя материал текстов, возникших с первых лет чешской письменности до начала XIX века, среди которых

¹⁶ См. <https://madla.ujc.cas.cz>. — Часть информации о LDHВЃ, включенной в нашу статью, черпнута с вводной страницы портала.

в незначительном количестве и далеко не всегда в достоверной форме встречается старочешская лексика. Однако, учитывая характер источников, которыми пользовался Юнгманн, этот материал очень неоднороден и ограничен, и потому не может быть полноценно использован лингвистами. В сокращенной форме словарный состав этой эпохи (прежде всего XVI века) включен в GbSlov. Данная ситуация резко контрастирует с состоянием изучения лексики древнечешского языка или языка XIX–XXI веков.

Созданию словаря старочешского препятствует недостаток лексического материала, который в должной мере отражал бы язык трех веков, в соответствии с требованиями эпох Ренессанса и Барокко. Важные общественные и культурные перемены — главным образом, уход от схоластического образа мысли, расцвет гуманистической науки и протестантской теологии — дали толчок к выработке гораздо более сложного мировоззрения, отличного от средневекового и, безусловно, нашедшего свое отражение в языке. Следствием активного развития сети школ стал рост грамотности, а вместе с ней — увеличение количества авторов и реципиентов письменных текстов; прогресс в области книгопечатания привел к умножению книжной продукции; усиление мобильности в Европе и на Ближнем Востоке углубило межкультурные контакты, а войны и политические (равно как и конфессиональные) коллизии привели к необходимости сосуществования и коммуникации носителей разных языков (Nejedlý 2018a). В отличие от предыдущей (средневековой) фазы своего развития словарный состав малого языка, каким на тот момент являлся чешский, значительно расширился. Согласно нашим предположениям, в XVI–XVIII веках в нем насчитывалось от 120 000 до 150 000 уникальных лексических единиц (Nejedlý 2018a: 429), для полного описания которых требуется не менее трех миллионов записей в базе данных (Nejedlý 2007: 127).

Разработка и развитие базы данных

База данных возникла в результате сотрудничества группы научных работников Отдела развития языка ÚJČ под руководством Петра Нейедлы при активном участии студентов и студенток богемистики (Lehečka, Martínek, Vajdlová 2007). Непосредственной «предшественницей» базы данных LDHVBČ является картотека с эксцерптами из памятников XVI–XVIII веков, содержащая около 550 000 записей, которую в 1970–1990-е годы XX века собрал сотрудник ÚJČ Зденек Тыл (1913–2008; см. (Тыл 1984; 1994)). В 2004 году, когда картотека перешла в собственность ÚJČ, были созданы концепция и структура LDHVBČ, с тех пор претерпевшие лишь незначительные изменения. Непосредственная работа над базой данных началась в 2005 году. За дигитализацией картотеки Зденека Тыла последовало постепенное обогащение LDHVBČ материалом меньших по объему картотек ÚJČ сходного характера, проходившее в 2006–2009 годах; в итоге количество записей выросло до 750 000.

Второй этап развития LDHВЃ, начавшийся в 2009 году, заключается в добавлении новых записей — эксцерптов из ранее не привлекавшихся исторических памятников. К концу 2022 года LDHВЃ насчитывала почти 1 300 000 записей и 110 000 зафиксированных лексических единиц. Одним из принципов пополнения базы данных в настоящее время является такой выбор источников, который охватил бы все типы и жанры текстов исследуемого периода, начиная от памятников художественной и специальной литературы различных областей и заканчивая личными дневниками, корреспонденцией, административными бумагами и другими произведениями прагматической письменности.

Цели и задачи базы данных

LDHВЃ выполняет две основные функции:

1) помогает составить общее представление о словарном составе чешского языка в период с 1500 по 1780 год, расширяя и углубляя его образ, закрепленный прежде всего словарем JgSlov;

2) создает материальную базу, которая в будущем станет основой словаря чешского языка раннего Нового времени и/или других языковых справочников.

База данных доступна всем зарегистрированным пользователям: как специалистам, так и студентам богемистики и исторических специальностей. LDHВЃ будет открыта для широкой публики после проведения ряда структурных и содержательных изменений, завершение которых запланировано на 2023 год.

Описание базы данных

Помимо общей информации, размещенной в разделах *О проекте* и *Подсказка*, а также рекомендаций *Как цитировать базу данных*, портал LDHВЃ содержит главные модули *Примеры* и *Источники*, которые предоставляют возможность поиска конкретной лексики во всех источниках и знакомят с информацией о памятниках, использованных при создании базы данных. Кроме того, на странице *Комментарии* пользователь портала может напрямую обратиться к авторам со своим вопросом или предложением.

В режиме *Поиск* модуля *Примеры* пользователь может отобразить зафиксированные случаи употребления искомого слова — или же, наоборот, все лексические единицы, выписанные из того или иного источника. Более того, портал позволяет комбинировать оба запроса. В случае библейских текстов поиск может осуществляться в выбранном фрагменте Библии. Поле поиска можно расширить (или же, наоборот, ограничить) параметрами *Лемма* и *Слово* (см. ниже).

Модуль *Источники* содержит ключевую информацию о каждом тексте, из которого производилась эксцерпция: его название, имя автора (или авторов), аббревиатуру источника в базе данных, а там, где это возможно — сведения о редакторе, переводчике, издателе, дате и времени публикации произведения, его объеме, библиотечном инвентарном номере, а также о других подробностях. В описании

печатной книги, помимо прочего, приводится регистрационный номер, присвоенный ей в цифровом Каталоге чешских и словацких печатных книг, изданных в 1476–1800 годы (Knižopis). Кроме того, каждая запись содержит данные о количестве лексем, выписанных из источника.

С материалами, содержащимися в разделах *Примеры* и *Источники*, можно также ознакомиться в режиме постраничного просмотра, то есть просматривая все записи подряд по алфавиту, как в бумажных словарях.

Структура записи в базе данных

Каждая запись содержит цитату из исторического памятника, в которой зафиксировано исходное слово, отображенную в виде цифровой копии карточки с эксцерптом или набранного на компьютере фрагмента текста.

Такая цитата сопровождается дополнительными сведениями, обязательными из которых являются:

а) *Лемма*, или заглавное слово — системная (словарная) форма данной конкретной лексической единицы в предполагаемой фонетической и морфологической форме, реконструированной на момент около 1500 года; исключениями являются лексемы, вошедшие в словарный состав позже. Учитывая продолжительность периода, в который происходили отдельные фонетические изменения, при реконструкции системной формы отдельных слов авторы базы данных исходят из сведений исторических грамматических справочников, а также из собственных знаний, полученных при изучении лексического материала;

б) *Слово* — форма лексической единицы, непосредственно представленная в источнике. База данных сохраняет исходную фонетическую форму слова, представленного в источнике, которая может отличаться от леммы (так, например, лемме *stržen* соответствует слово *sřěn*);

в) *Источник* — представленный аббревиатурой текст, из которого был выписан пример употребления лексической единицы.

д) Иногда приводится дополнительная информация о лексеме: год создания (прежде всего у юридических документов) и страница источника, на которой расположен пример, а в случае современных изданий — номер соответствующего тома или журнала; у библейских текстов — название книги, номер главы и стиха. При необходимости база данных комментирует зафиксированную форму и/или меру достоверности записи, приводит этимологию заимствованных слов, а также толкование значения слова.

Принципы включения лексических единиц в базу данных

Понимание лексической единицы как части лексической системы соответствует концепции StčS (см. выше): лексическая единица — выражение, имеющее самобытные форму и содержание (StčS 1968: 20–22). Таким образом, заглавное

слово мы понимаем как самостоятельную форму лексемы, которая обладает, по меньшей мере, одним самостоятельным лексическим значением и состоит в структурных отношениях с другими лексемами, от которых отличается исключительным сочетанием формы и значения.

Поскольку результатом нашей работы является база данных, а не собственно словарь, приведенное выше определение применяется в общих чертах: лексическая единица включается в базу данных как единое заглавное слово; значения многозначных слов не рассматриваются отдельно, поскольку подробный семантический анализ релевантен уже для комплексного лексического анализа всего словарного состава, который, в свою очередь, должен исходить из большого количества языкового материала.

В отличие от стандартных словарей, LDHВЃ не рассматривает по отдельности ни предполагаемые, ни несомненные омонимы (*hýl* ‘снегирь; тележка землекопа’, *plýtvati* ‘плавать; расточать’). Лексемы, состоящие из нескольких слов, сопровождают лемму семантически ключевого слова (так, например, выражение *dáti na pamětnou* приводится под леммой *pamětná*; а выражение *nestáti za balatku* — под леммой *balatka*), включая и те случаи, когда ключевых слов несколько (так, выражение *z komára dělati velblouda* приводится под леммами *komár* и *velbloud*).

Для облегчения работы пользователя в LDHВЃ имена собственные приводятся в качестве самостоятельных лексических единиц даже в тех случаях, когда с формальной точки зрения являются идентичными апеллятивам (например, фамилия *Pekárek* и профессионализм *pekárek*)¹⁷. Прилагательные, являющиеся частью имен собственных, в многословных выражениях всегда приводятся под соответствующей апеллятивной леммой (так, сочетание *Harfové moře* приводится под леммой *harfový*).

Иноязычные вкрапления, встречающиеся в источниках, в базу данных не включены.

Принцип записи лексических единиц в базе данных

Лексические единицы в LDHВЃ приводятся в словарной форме. Как лемма, так и слово транскрибируются в соответствии с узусом современного чешского языка. Исключением из этого правила отчасти являются заимствованные слова: при определении их лемм особое внимание уделяется предполагаемой мере их включенности в чешскую лексическую систему. В поле *Лемма* мы приводим стандартную форму слова (*kavalér, exekutor, sofistický*), а в поле *Слово* — форму, встречающуюся непосредственно в источнике (*cavaglier, exequutor, sophistycský*). Кроме того, в слове мы сохраняем подтвержденное употребление диграфа *au*, сигнализирующего переходный этап до возникновения дифтонга *ou* (*dlauhost*).

¹⁷ База данных предоставляет возможность поиска слов, начинающихся с большой или с маленькой буквы.

Проблемы, возникающие при определении леммы и слова, часто связаны с долготой гласных, не всегда последовательно обозначенной в источниках и создающей трудности даже для современных издателей исторических текстов. По этой причине в случаях некоторых семейств или словообразовательных типов в учет берется продолжительность существования слова в языке, особенности его развития, роль в лексической системе, возможное влияние диалектов и т. д. Впрочем, поскольку мы ставим перед собой задачу представить словарный состав не конкретного текста, а всего языка; в вопросах долготы гласных решающую роль играют этимология и словообразовательные характеристики лексемы; при необходимости авторы LDHВЃ обращаются к работе Милады Недведовой *Kvantita samohlásek ve starší češtině* (Nedvěďová s. a.).

Перспективы дальнейшего развития базы данных

Уже в первые годы своего существования LDHВЃ обогатила наши знания о языке: благодаря ей удалось обнаружить доселе неизвестные лексемы (*hltoun*, см. (Fuková, Nejedlý, Šimek 2019: 81)), объяснить форму и значение слов, долгое время остававшиеся неясными (*kloryť*, см. (Nejedlý 2018b); *nasáďista*, см. (Nejedlý 2011)), найти самые ранние примеры употребления слов, знакомых нам по более молодым текстам (*kytara*), подтвердить предполагаемое происхождение некоторых лексических единиц (*he(j)ský* — современное *hezký*, см. (Rejzek 2012; 2015: s. v. *hezký*)), лучше понять некоторые грамматические и словообразовательные процессы (образование прилагательных цвета с помощью приставок *na-/ná-*, см. (Nejedlý 2010b)), а также пролить свет на взаимоотношения лексических единиц в системе и их развитие, равно как и развитие подсистем как целого (Nejedlý 2010a).

При этом мы не забываем о нашей главной цели — превратить базу данных LDHВЃ в полноценный фундамент для стандартного словаря чешского языка раннего Нового времени, созданного на основе методологии, соответствующей требованиям современной лексикологии. Именно поэтому объем словарного материала должен быть таким, чтобы на его основе можно было проследить как внутрисистемные отношения лексических единиц, так и отношения внесистемные, т. е. их связь с рефлексией реальности в сознании носителей языка и с самой реальностью¹⁸. Этой цели мы планируем достичь в ближайшие годы.

ЛИТЕРАТУРА

Černá 2007: Černá A.M. Zpráva o internetové aplikaci Vokabulář webový // Naše řeč. 2007. 90, 164.

Černá 2011: Černá A.M. Vokabulář webový — internetová platforma ESSČ (jeho současnost a budoucnost) // Jazykovědné aktuality. 2011. 48, 40–43.

Černá, Lehečka 2016: Černá A.M., Lehečka B. Vokabulář webový — webové hnízdo pramenů k poznání historické češtiny // Piorecký K. Ostrovy spolehlivého poznání. Elektronické informační zdroje Akademie věd ČR o společnosti a kultuře. Praha, 2016, 21–24.

Dvořáčková 2011: Dvořáčková V. Osudy Ústavu pro jazyk český. Praha, 2011.

¹⁸ О лингвистических аспектах будущего словаря подробнее см. (Fuková, Nejedlý, Šimek 2019).

- Dušek 2017:** Dušek J. Vokabulář webový a škola // Jazyk — Literatura — Komunikace. 2017. 6, 27–35.
- ESSČ:** Elektronický slovník staré češtiny. Praha, 2006—. <<https://vokabular.ujc.cas.cz/hledani.aspx>>
- Fuková et al. 2019:** Fuková I. et al. Návod ke zpracování ESSČ. Praha, 2019. Материал для внутреннего пользования.
- Fuková, Nejedlý, Šimek 2019:** Fuková I., Nejedlý P., Šimek Š. K připravovanému Slovníku raněnovověké češtiny // Naše řeč. 2019. 102, 76–87.
- Fuková, Pytlíková 2021:** Fuková I., Pytlíková M. Historical Online Dictionary as a Dynamic Process: Fourteen Years and Three Stages of the Electronic Dictionary of Old Czech // Van de Velde H., Dolezal F.T. Broadening Perspectives in the History of Dictionaries and Word Studies. Newcastle upon Tyne, 2021, 262–275.
- GbSlov:** Gebauer J. Slovník staročeský 1–2 (*a–netbalivost*).² Praha, 1970.
- Gebauer 1929:** Gebauer J. Historická mluvnice jazyka českého 4. Skladba. Praha, 1929.
- Gebauer 1960:** Gebauer J. Historická mluvnice jazyka českého 3. Tvarosloví 1. Skloňování.² Praha, 1960.
- Gebauer 1963:** Gebauer J. Historická mluvnice jazyka českého 1. Hláskosloví.² Praha, 1963.
- Gebauer 1963a:** Gebauer J. Historická mluvnice jazyka českého 3. Tvarosloví 2. Časování.² Praha, 1958.
- Havránek 1968:** Havránek B. Předmluva // Staročeský slovník. Úvodní stati, soupis pramenů a zkratek. Praha, 1968.
- Homolková 2007:** Homolková M. Historická lexikografie // Pleskalová J., Krčmová M., Večerka R., Karlík P. Kapitoly z dějin české jazykovědné bohemistiky. Praha, 2007, 199–242.
- JgSlov:** Jungmann J. Slovník česko-německý. Praha, 1835–1839.
- Knihopis:** KPS-Databáze Knihopis. <<https://www.knihopis.cz>>
- Lehečka, Martínek, Vajdlová 2007:** Lehečka B., Martínek F., Vajdlová M. Příspěvek k poznání české slovní zásoby doby střední // Kamper-Warejko J., Kaproň-Charzyńska I., Kulwicka-Kamińska J. Studia nad słownictwem dawnym i współczesnym języków słowiańskich. Toruń, 2007, 97–102.
- LDHBC:** Nejedlý P. et al. Lexikální databáze humanistické a barokní češtiny. Praha, 2010. <<https://madla.ujc.cas.cz>>
- MSS:** Bělič J., Kamiš A., Kučera K. Malý staročeský slovník. Praha, 1979.
- Nedvěďová s. a.:** Nedvěďová M. Kvantita samohlásek ve staré češtině. Неопубликованная диссертация.
- Nejedlý 2007:** Nejedlý P. Rekonstrukce sémantického vývoje lexikálních jednotek v období humanismu a baroka // Kamper-Warejko J., Kaproň-Charzyńska I. Z zagadnień leksykologii i leksykografii języków słowiańskich. Toruń, 2007, 127–134.
- Nejedlý 2010a:** Nejedlý P. Diskontinuita jazykového vývoje a lexikální systém // Čmejrková S., Hoffmannová J., Havlová E. Užívání a prožívání jazyka. Praha, 2010, 183–187.
- Nejedlý 2010b:** Nejedlý P. Slovní zásoba humanistické a barokní češtiny: mezi kontinuitou a diskontinuitou // Janečková M., Alexová J., Pospíšilová V. et al. Slovesné baroko ve středoevropském prostoru. Praha, 2010, 105–117.
- Nejedlý 2011:** Nejedlý P. Nazádisté a násadisté // Naše řeč. 2011. 94, 159–161.
- Nejedlý 2018a:** Nejedlý P. Cesty k slovníku raněnovověké češtiny // Niševa B., Blažek D., Krejčířová I., Skwarska K., Šlaufová E., Vašíček M. Slovanská lexikografie počátkem 21. století. Praha, 2018, 489–494.
- Nejedlý 2018b:** Nejedlý P. Ze slovníku raněnovověké češtiny: *kloryt* // Naše řeč. 2018. 101, 185–188.
- Nejedlý et al. 2002:** Nejedlý P. et al. Výstavba hesla v diachronním slovníku. Praha, 2002. Материал для внутреннего пользования.
- Němec 1965:** Němec I. Slovo a význam v historickém slovníku // Slovo a slovesnost. 1965. 26, 136–150.

Němec 1968: Němec I. Vývojové postupy české slovní zásoby. Praha, 1968.

Němec 1975: Němec I. Principy studia slovní zásoby a jejího vývoje // Slovo a slovesnost. 1975. 36, 92–103.

Němec 1980: Němec I. Rekonstrukce lexikálního vývoje. Praha, 1980.

Němec 2009: Němec I. Práce z historické jazykovědy / Dittmann R., Sobalíková H., Vajdlová M. Praha, 2009.

Rejzek 2012: Rejzek J. K původu adjektiva *hezky* // Nejedlý P., Vajdlová M. Cesty slov. Praha, 2012, 177–182.

Rejzek 2015: Rejzek J. Český etymologický slovník.² Voznice, 2015.

StČS 1968: Staročeský slovník. Úvodní stati, soupis pramenů a zkratk. Praha, 1968.

StČS 1968–2008: Staročeský slovník 1–26. Praha, 1968–2008.

Šimek 1947: Šimek F. Slovníček staré češtiny. Praha, 1947.

TLP 1966: Travaux linguistique de Prague, 2. Les problèmes du Centre et de la périphérie du système de la langue. Prague; Paris, 1966.

Tyl 1984: Tyl Z. Česká slovní zásoba tzv. doby střední a některé otázky jejího dalšího lexikografického zpracování // Jazykovědné aktuality. 1984. 21, 28–30.

Tyl 1994: Tyl Z. Unum necessarium české historické lexikografie // Miscellanea oddělení rukopisů a starých tisků. 1994. 10, 429–439.

Vajdlová 2007: Vajdlová M. Internetová aplikace Vokabulář webový — nový zdroj poznání staročeské slovní zásoby // Kamper-Warejko J., Kaproň-Charzyńska I. Z zagadnień leksykologii i leksykografii języków słowiańskich. Toruń, 2007, 55–58.

Vajdlová 2011: Vajdlová M. Minulost a budoucnost Elektronického slovníku staré češtiny // Jazykovědné aktuality. 2011. 48, 29–34.

Vajdlová 2018: Vajdlová M. Elektronický slovník staré češtiny v kontextu slavistickém // Niševa B., Blažek D., Krejčířová I., Skwarska K., Šlaufová E., Vašíček M. Slovanská lexikografie počátkem 21. století. Praha, 2018, 523–530.

Vajdlová 2019: Vajdlová M. Klady a zápory elektronického zpracování a zveřejnění slovníku (na příkladu Elektronického slovníku staré češtiny) // Naše řeč. 2019. 102, 88–101.

VW: Vokabulář webový: webové hnízdo pramenů k poznání historické češtiny. Praha, 2006—. <https://vokabular.ujc.cas.cz>