МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Филологический факультет
Кафедра русского языка
Лаборатория общей и сибирской лексикографии
ИНСТИТУТ РУССКОГО ЯЗЫКА ИМ. В.В. ВИНОГРАДОВА РАН
ГОСУДАРСТВЕННЫЙ ИНСТИТУТ РУССКОГО ЯЗЫКА им. А.С. ПУШКИНА
ПОД ЭГИДОЙ РОССИЙСКОГО ОБЩЕСТВА ПРЕПОДАВАТЕЛЕЙ
РУССКОГО ЯЗЫКА И ЛИТЕРАТУРЫ (РОПРЯЛ)

ЛЕКСИКОГРАФИЯ ЦИФРОВОЙ ЭПОХИ

Сборник материалов Международного симпозиума
(24–25 сентября 2021 г.)

Томск
Издательство Томского государственного университета
2021

DOI: 10.17223/978-5-907442-19-1-2021-122

М.М. Угрюмова
Maria M. Ugryumova

**Лексикографический компонент
Томского диалектного корпуса***
The lexicographical component of the Tomsk dialect corpus

Томский государственный университет, Томск – Tomsk State University, Tomsk
maria_ugr@mail.ru

Аннотация. Рассматриваются особенности разработки лексикографического модуля Томского диалектного корпуса: принципы отбора и подготовки словарных единиц для внедрения в корпус, подходы к представлению материала.

Summary. The article considers the features of the development of the lexicographic module of the Tomsk dialect corpus: the principles of selecting and preparing vocabulary units for implementation in the corpus, approaches to the presentation of the material.

Ключевые слова: русские говоры, электронный словарь, диалектный корпус, лексикографический компонент

Keywords: Russian dialects, electronic dictionary, dialect corpus, lexicographic component

Корпусное направление является одним из наиболее актуальных для современной диалектологии. Создание корпуса на материале говоров Среднего Приобья началось в 2010 г. группой томских диалектологов и программистов под руководством профессора Е.А. Юриной [7]. Данный ресурс проектируется как универсальная информационно-поисковая система, включающая текстовый, собственно корпусный (грамматическая разметка) и лексикографический модули [2. С. 101].

Цель разработки лексикографического модуля – выдача толкования собственно диалектных лексем, что дает возможность пользователю прояснить значение слова без обращения к диалектному словарю, упрощает и ускоряет работу с диалектными текстами, а также делает доступным их

* Исследование выполнено за счёт гранта Российского научного фонда «Разработка электронных ресурсов для исследования народно-речевой культуры Среднего Приобья» (проект № 19-78-10015).

понимание широкому кругу лиц. Идея интеграции Томского диалектного корпуса с диалектным словарём как первого шага семантической разметки впервые была предложена Е.В. Иванцовой [3].

В настоящий момент перед разработчиками стоит задача – подготовить электронный словарь, т.е. реестр диалектных единиц, к которым дано словарное толкование и ряд помет, для интеграции с корпусом. Источником единиц электронного словаря стал дифференциальный толковый «Словарь русских старожильческих говоров средней части бассейна р. Оби» под ред. В.В. Палагиной [4] и два дополнения к нему [5, 6] (далее – СРСГ).

Принципы перевода материала в электронную форму для внедрения в корпус определяются тем, что лексикографический компонент рассматривается как часть универсальной электронной системы, при этом автоматическая обработка осуществляется на уровне отдельного слова, а каждая словарная статья функционирует как автономная.

Одной из центральных задач является обработка единиц словаря-источника:

1) определение случаев многозначности и омонимии (слова находились в разных томах): ДВОЙНИ'К¹, м. Одна часть расколотого пополам полена. ДВОЙНИ'К², м. Устар. Вид сошника.

2) уточнение толкования словарной единицы. В соответствии с принципом автономного существования словарной статьи отсылочные толкования («**то же, что...**», «**см.**»), данные в СРСГ, заменялись на содержательные. Кроме того, в некоторых случаях, если толкование лексемы казалось неясным или спорным, проводилась проверка по другим словарям. Так, например, в СРСГ: ЖЕЛТУНИ'ЦА 'лекарственное растение (**какое?**)'. Уточнение по «Словарю фитонимов Среднего Приобья» [1] позволяет дать более точное толкование: **Растение пижма. Tanacetum vulgare L.** Глистник, девятельник <...> дикая рябинка, желтоголовник, **желтуница**, желтуха <...> [1. Т. 2. С. 79].

В некоторых случаях словарное толкование было не только уточнено, но и дополнено, расширено (например, к словам, связанным со старинным бытом, названиям религиозных праздников, промысловой лексике и т.п.). Так, к словарной статье ДЁР³ 'кострика' дается пояснение: «жёсткая кожица и деревянистые волокна льна и конопли, остающиеся, как отброс, после их трепания и чесания».

3) Исключение диалектных единиц, совпадающих по форме с общерусскими словами (душа 'ямочка в нижней части шеи спереди', дым 'туман'). Предполагается, что толкование слова будет выдаваться во всех

текстах, включающих его, что может привести к ошибкам в случае совпадения с общерусскими словами. Однако в некоторых случаях востребованным (частотным) является диалектное значение: КОКЕ'ТКА 'небольшой платок на голову' (не 'женщина, стремящаяся нравиться и привлекать внимание мужчин'), такие слова оставлены в итоговой версии словаря.

4) Исключение диалектных фразеологических и нефразеологических оборотов, состоящие из общерусских слов (НЕ ДАТЬ ВЕ'ТРУ ДУ'НУТЬ 'нежить, растить в неге', ДВОЙНАЯ РАЗЛУКА 'игра горелки с большим количеством участников' и др.). Причина состоит в техническом ограничении создаваемого ресурса, к настоящему времени возможно распознавание отдельного слова, а не оборота. Обороты и устойчивые выражения, включающие диалектные слова, даются при них, в одной словарной статье: ЕДО'ВЫЙ 'предназначенный для еды'. □ ЕДО'ВЫЙ ХЛЕБ. Зерно для помола.

Ожидается, что внедрение лексикографического компонента в корпус значительно упростит работу с диалектными текстами, сделает возможным их понимание неспециалистами.

Литература

1. *Арьянова В.Г.* Словарь фитонимов Среднего Приобья [в 3 т.]. Томск: Изд-во Том. госуд. пед. ун-та, 2006–2008.
2. *Земичева С.С.* От *абарма* до *яичишика*: разработка лексикографического компонента Томского диалектного корпуса // Вопросы лексикографии. № 18. 2020. С. 98–117.
3. *Иванцова Е.В.* Томский диалектный корпус: обоснование концепции и перспективы развития // Вопросы лексикографии. 2017. № 11. С. 54–70.
4. Словарь русских старожильческих говоров средней части бассейна р. Оби / ред. В.В. Палагина. Томск: Изд-во Том. ун-та, 1964–1967. Т. 1–3.
5. Словарь русских старожильческих говоров средней части бассейна р. Оби (Дополнение) / ред. О.И. Блинова, В.В. Палагина. Томск: Изд-во Том. ун-та, 1975. Ч. 1–2.
6. Среднеобский словарь: (Дополнение) / ред. В.В. Палагина. Томск: Изд-во Том. ун-та, 1983–1986. Ч. 1–2.
7. *Юрина Е.А.* Томский диалектный корпус: в начале пути // Вестник Томского государственного университета. Филология. 2011. № 2 (14). С. 58–63.