

*А.С. Герд*

**НАЦИОНАЛЬНЫЙ КОРПУС РУССКОГО ЯЗЫКА –  
СЛОВАРНАЯ КАРТОТЕКА –  
АКАДЕМИЧЕСКИЙ СЛОВАРЬ**

Разработчики корпусов, как правило, пишут о том, что может делать их корпус, исходя из своих внутренних авторских идей, и совсем забывают о потребностях разных внешних пользователей. В последнее время, к сожалению, всегда актуальная проблема изучения информационной потребности лексикографов почти совсем исчезла со страниц печати.

Каковы основные типы запросов автора толкового академического словаря?

Найти:

- слово, найти новое слово по времени его появления,
- исходную форму слова;
- цитаты к уже известным значениям;
- цитаты к тем значениям, на которые в словаре нет цитат (чаще всего это грамматически обусловленные значения, например, страдательные формы русских глаголов или речевые употребления),
- дополнительные новые цитаты к тому или иному значению;
- новые типы лексической и синтаксической сочетаемости;
- новые фразеологизмы;
- новые современные научные толкования специальных терминов.

Может ли ответить РНК сегодня на эти запросы?

Как и в любой информационной работе и здесь, прежде всего, весьма важным остается вопрос об изучении информационной потребности пользователя, заказчика.

В настоящее время в ИЛИ РАН активно идет работа над новым Большим академическим словарем русского языка (БАС–3; см. предисловие и общие сведения о словаре. Т. 1, СПб., 2004)<sup>1</sup>.

Каковы принятые изменения в работе над новым БАС–3? Начиная с третьего, четвертого томов в последние годы наряду с пополнением словника основным является, прежде всего, увеличение числа цитат на значение, обновление цитат, поиск новой лексической и синтаксической сочетаемости для конкретных слов, подкрепленных цитатами.

Конечно, РНК очень полезен при поиске цитат на новые слова типа *кайтер*, *пирсинг*, *сноуборд*, *скутер*, *эмо*. Корпус помогает освежить иллюстрации на известное уже значение цитатами из новейшей литературы. Авторам толковых словарей не хватает цитат как раз не на основные, а на периферийные значения того или иного слова. Однако не только в толковый словарь, но и в тот лексический фонд, который является его основой, будь то бумажная картотека или электронная, включению подлежат далеко не все цитаты.

И в былые времена активного пополнения БСК, в 50–80-е годы XX века всегда был некоторый контроль за характером поступающих цитат, и, главное, за работой выборщиков, которым заранее выставлялись некоторые требования. Наиболее эффективно эта работа была поставлена в ИЛИ РАН (Ленинградском отделении Института языкознания АН) под руководством Р.П. Рогожниковой<sup>2</sup>.

Отбор цитат в БСК шел по представленности авторов, по хронологии, по жанрам, по типам текстов, по сочетаемости слов.

---

<sup>1</sup> Большой академический словарь русского языка. Т. 1–9, М.–СПб., 2004–2008.

<sup>2</sup> *Практическая лексикография: 100 лет словарной картотеке* / Отв. ред. Р.П. Рогожникова. М., 1989. См. также: *Захаров В.П.* Словарная картотека Института лингвистических исследований РАН как объект автоматизации // *Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог–2007»*. М., 2007.

Таким образом, один из первых вопросов, которые подлежат обсуждению в перспективе – это корпус (РНК) и лексический фонд – источник конкретного словаря. Очевидно, что это не одно и то же. На сегодняшний день при работе над новым Большим академическим словарем русского языка РАН пока БСК ИЛИ РАН в принципе дает по-прежнему больше материалов, чем РНК.

Основная антиномия в творческой работе словарника – это слово в его значениях, оттенках и употреблениях (многие из которых уже известны), с одной стороны, и море цитат в РНК, распределенных только по словоформам, – с другой.

Второй главный и сегодня неясный вопрос: в каком виде должно задаваться слово при поиске – словоформы или леммы.

Серьезным преимуществом БСК является и то, что сами цитаты в ней отобраны по необходимому размеру, смысловой достаточности и, как правило, хорошо документированы.

Как показывает опыт, каждая новая словарная тема, проект каждого нового словаря порождает свою картотеку, свой фонд исходных данных.

Долгие годы основным стимулом и регулятором пополнения БСК был БАС–1. Уже в 60-е годы XX века в ИЛИ РАН на базе БСК возникли специализированные картотеки Словаря русских народных говоров, Фразеологического словаря, Словаря русского языка XVIII века, наконец, недавно – Словаря русского языка XIX века, Словаря М.В. Ломоносова.

Когда в конце 80-х годов XX века в ИЛИ РАН начал активно обсуждать проект Нового академического словаря русского языка сразу возник вопрос о пополнении БСК и новыми источниками<sup>1</sup>. Возникновение картотек, ориентированных на отдельные новые академические словари не случайно, таким образом, практически каждая большая новая словарная тема порождает свою специали-

---

<sup>1</sup> Вопрос был поставлен гораздо раньше А.М. Бабкиным, Ф.П. Филиным и Ф.П. Сороколетовым, см. также: *Скляревская Г.Н.* Новый академический словарь. Проспект. СПб., 1994.

зированной картотеку. Так было, например, и в Межкафедральном словарном кабинете имени Б.А. Ларина в Петербургском университете. Из картотеки печорских говоров – родился Словарь говоров Низовой Печоры, Псковский областной словарь – потребовал создания своей грандиозной картотеки, под Словарь русских говоров Карелии – была организована новая картотека<sup>1</sup>.

Сегодня отношения между РНК и автором оригинального толкового словаря подобны поведению естествоиспытателя в природе.

Вот естествоиспытатель вошел в определенный лес, состоящий из разных пород деревьев и ищет какие-то формы насекомых, растений. Словарник вошел в лес разных текстов в РНК и ищет определенную словоформу (РНК – лес; тексты – виды деревьев). Естествоиспытатель нашел некую форму растения, словарник нашел словоформу. И насекомое в лесу, и словоформа в текстах представлены множеством своих форм и вариаций. Естествоиспытатель узнал и предварительно определил род насекомого, растения, словарник – словоформу (слово).

Но и у естествоиспытателя, и у словарника все впереди. Самое трудное естествоиспытателю – точно определить вид, подвид и поместить его на свое место в систематику; словарнику – определить значение (оттенок, употребление) и поместить его в систему значения слова в словаре.

Все сказанное свидетельствует о том, что между РНК и практической академической лексикографией должно быть еще одно звено, которое перерабатывает материалы РНК в данные пертинентные для автора словарной статьи.

Опыт такого фильтра есть. Как сейчас организована работа над БАС–3 в ИЛИ РАН? У каждой буквы – есть опытный ответственный редактор. У редактора – есть помощник. Редактор, ра-

---

<sup>1</sup> Из более редких публикаций см.: *Беликов В.И.* Оцифрованные тексты как материал для словаря русских регионализмов // Труды международной конференции «Корпусная лингвистика – 2006». СПб., 2006.

ботая над отрезком, определяет, какие новые слова включить, где, к какому значению не хватает цитат, какие намечаются новые значения и насколько они подкреплены иллюстрациями. Помощник ищет новые цитаты на отдельные значения конкретного слова как в БСК, в РНК, так и в других сетевых электронных фондах. Редактор отбирает, корректирует и решает, какие из цитат включать в словарь<sup>1</sup>.

Главная задача такого звена – фильтр цитат для словарной статьи. Сегодня, эта работа автоматизирована лишь до некоторой степени<sup>2</sup>.

Один из самых сложных вопросов здесь – где и как должно быть организовано такое звено?

Если учесть, что в будущем все новые толковые словари будут создаваться преимущественно на базе корпусов, то роль такой службы многократно возрастает.

Не за горами в ИЛИ РАН в Академии наук обсуждение принципов Словаря русского языка первой половины XX века. Как известно, именно язык источников этого периода недостаточно представлен в существующих толковых словарях.

Параллельно нуждается в обсуждениях актуальная идея электронного тезауруса как сокровищницы русской лексики, родных языках.

Насколько готов РНК к фундаментальным запросам современной академической лексикографии?

---

<sup>1</sup> Герд А.С. Новый академический словарь – исходные позиции и ориентиры // Актуальные проблемы разработки нового академического словаря русского языка (тезисы). Л., 1990.

<sup>2</sup> См. постановку этого вопроса также: Савчук С.О., Пискунова С.В. Опыт создания корпуса текстов первой половины XX века // Труды Международной конференции «Корпусная лингвистика – 2006». СПб., 2006.

### **Принятые сокращения**

*БАС-1* – Словарь современного русского литературного языка АН СССР. Т. 1–17. М.–Л., 1948–1965.

*БАС-3* – Большой академический словарь русского языка ИЛИ РАН. Т. 1–9. СПб., 2004–2008; издание продолжается.

*БСК* – Большая словарная картотека Института лингвистических исследований РАН (Санкт-Петербург).

*ИЛИ* – Институт лингвистических исследований РАН, Санкт-Петербург.

*РАН* – Российская Академия наук.

*РНК* – Национальный корпус русского языка.