

**О.А. Фелькина (Брест, Республика Беларусь)**

## **ЛЕКСИЧЕСКИЕ ПРОЦЕССЫ В РУССКОМ ЯЗЫКЕ В ОТРАЖЕНИИ ЧАСТОТНЫХ СЛОВАРЕЙ**

Частотные словари – относительно новый вид лексикографических источников, их появлению способствовало развитие информационных технологий: уже для составления первого русского частотного словаря (под ред. Л. Н. Засориной, 1977 [1]) использовались компьютеры. За последние десятилетия информационные технологии шагнули далеко вперед, мощность компьютеров существенно выросла, что позволяет создавать частотные словари с меньшими трудозатратами и на гораздо большем материале. Так, если словарь под ред. Л. Н. Засориной базируется на текстах объемом чуть более миллиона словоформ, то частотный список С. А. Шарова составлен на основе корпуса в 16 млн. словоупотреблений [2], а «Новый частотный словарь русской лексики» О. Н. Ляшевской и С. А. Шарова – 92 млн. словоупотреблений [3]. На первый взгляд, наличие частотных словарей, составленных в разные эпохи, дает возможность точно оценить происходящие в русском языке лексические процессы: выявить устаревающие слова и лексемы, становящиеся более употребительными и под. Однако существует несколько проблем, снижающих степень достоверности данных частотных словарей.

Во-первых, для выводов об изменении частотности слов необходимо сравнивать материал определенных десятилетий. Но в картотеку словаря Л. Н. Засориной включены труды В. И. Ленина, М. И. Калинина, выступления писателей и публицистов периода Великой Отечественной войны, научные труды (С. И. Вавилов, К. А. Тимирязев и др.). Причем первое издание книги К. А. Тимирязева «Жизнь растения» датируется еще 1878 г., работы С. И. Вавилова и А. Е. Ферсмана, вошедшие в базу словаря, относятся к 1927–1928 гг. Поэтому подкорпус публицистических и научных текстов (объединены в словаре) не может быть отражением русской речи 60-х гг. XX в. Объем газетных текстов составляет 125 тыс. словоупотреблений, это номера за 5 января 1968 года нескольких изданий. Таков же объем журнальных текстов, выбранных из журналов 1958–1962 гг. Таким образом, подкорпус газетных и журнальных текстов четко очерчен хронологически. Драматургия составляет самую большую часть корпуса словаря – 27,2 %. Она рассматривается как отражение разговорной речи, хотя именно разговорная речь меняется наиболее стремительно, а в картотеку вошли, например, пьесы «Враги» М. Горького (1906), «Дни Турбиных» М. Булгакова (1925), «Бронепоезд 14–69» В. В. Иванова (1927, опубл. в 1932) и под. Из 21 пьесы, включенной в источники словаря,

только 6 созданы после 1950 г., а 9 написаны еще до Великой Отечественной войны. Подкорпус художественной прозы состоит из 13 произведений, только 6 из которых созданы после войны.

Список С. А. Шарова построен на основе корпуса, включающего прозу, политические мемуары, газетные тексты и научно-популярную литературу 1970–2002 гг. Здесь четче временные рамки, но все же заметна разнородность: большинство произведений созданы в 1980–1995 гг., а газеты взяты за 1997–1999 гг. «Новый частотный словарь русской лексики» описывает материал 1950–2007 годов, это большой отрезок времени – почти шесть десятилетий, причем два из них охвачены и словарем Л. Н. Засориной. Известно, что в 90-е годы письменная и устная речь претерпели существенные изменения, поэтому объединение в одном корпусе текстов этого периода с более ранними затрудняет оценку лексических процессов второй половины и конца XX в.

Во-вторых, существует проблема больших текстов (длинный текст может содержать большое количество вхождений некоторого редкого слова). Так, С. А. Шаров первоначально отобрал корпус из 40 млн. словоупотреблений, в который вошел роман Ника Перумова объемом 250 тыс. слов (0,6 % от общего объема корпуса), написанный по мотивам «Властелина колец» Д. Р. Р. Толкина. Благодаря этому тексту слово *хоббит* оказалось в первой тысяче русских слов. В связи с этим С. А. Шаров ограничил выборку: из больших текстов брались отрывки до 10 тыс. слов (примерно 30 страниц), из текстов одного автора бралось не более 100 тыс. слов. В результате подмножество полного корпуса, использованное при подсчете частоты, составило около 16 млн. слов. Можно ли считать, что проблема устранена? Например, слово *анк* имеет в списке частотность 7,83, а в Национальном корпусе русского языка (далее НКРЯ) за 1950–2007 г. всего 19 вхождений (0,1), при этом только 2 раза не как имя собственное, и то не слово, а корень: *Древние греки называли железный якорь словом «апсига» – «анкура», происходившим от корня «анк», что по-русски означает «крюк», «кривой» или «изогнутый»* [Л.Н. Скрягин. Книга о якорях (1973)]. В остальных случаях это аббревиатуры *АНК* (Африканский национальный конгресс; АНК «Башнефть») [4]. В достаточно частотные попали имена героев различных произведений и отчасти исторических лиц: 10,83 *шмаков* (в «Чучеле» В. Железникова 140 вхождений; в романе «Живые и мертвые» К. Симонова – 175), 7,96 *решетнев* (в НКРЯ всего 6 в разных статьях; М. Ф. Решетнёв – учёный, конструктор, один из основоположников советской космонавтики) и т. п. И наоборот – минимальную частоту 1,04 в списке С. А. Шарова имеют слова *истец* (в НКРЯ 9,9 ipm), *реализовывать* (9,1) и т. п.

Третья проблема – различная стилистическая и жанровая принадлежность материала: в словаре 1977 г. художественная проза составляет лишь четверть картотеки (меньше, чем драматургия), а в более новом словаре – больше половины, и драматургия вообще не представлена. Сравнение материалов словаря Л. Н. Засориной и частотного списка С. А. Шарова показывает, что частотность ряда слов увеличилась в 1,4 и более раз: местоимения *тот* (в 4,2), *ничто* (в 2), *весь*, *никто*, *он*, *кто*, *какой*, существительные *ребенок* (в 2,2), *человек* (в 1,9), *начало*, *нога*, *ряд*, *голова*, *женщина*, *час*, *дом*, прилагательные *нужный* (1,9), *хороший*, *конечный*, глаголы *спросить*, *есть*, *оказаться*, *сидеть*, наречия *потом*, *где*, *тута*, *просто*, *сразу*, *вдруг*, служебные слова *чем* (в 2), *уж*. Однако, например, существительное *дом* в словаре Л. Н. Засориной при общей частотности 756 ipm в подкорпусе художественной литературы имеет показатель 1014, что не так существенно отличается от данных С. А. Шарова (1022). Частотность же этого слова в подкорпусе художественной литературы «Нового частотного словаря русской лексики» ниже, чем в словаре Л. Н. Засориной – 925,9. Для слов *нога*, *голова*, *глаз*, *женщина*, *вдруг*, *сразу* частотность в подкорпусе художественной литературы словаря Л. Н. Засориной даже выше, чем в списке С. А. Шарова. При этом первые три из перечисленных слов имеют более низкую частотность и в подкорпусе художественной литературы «Нового частотного словаря русской лексики», а три последние – напротив. Так, наречие *вдруг* в художественной прозе имеет частотность 775 в словаре Л. Н. Засориной и 916 в «Новом частотном словаре русской лексики», тогда как в списке С. А. Шарова – 713.

По данным словаря Л. Н. Засориной и частотного списка С. А. Шарова, в 1,4 и более раз упала частотность существительных *война*, *работа*, *земля*, *книга*, *сила*, *вопрос*, *мир*, *путь* (в 1,5–1,9 раз), *свет*, *страна* (в 2,1–2,2), *товарищ*, *народ*, *армия*, *камень*, *партия*, *труд* (в 2,4–2,7), *развитие* (в 3,2), *наука* (в 3,6), *стекло*, *завод*, *производство*, *рабочий* (в 4,3–5,8), *борьба* (в 6,4), глаголов *слышать*, *дать*, *пойти*, *идти*, *уйти*, *выходить* (в 1,6–1,7), *начать* (в 2,2), *уходить* (в 2,5), *входить* (в 5), *видать* (в 14,4), прилагательных *высокий*, *молодой* (в 1,7–1,8), *новый* (в 1,9), *великий* (2,6), *советский* (в 2,9), *важный* (в 3,4), *старший*, *американский* (в 4,1–4,8), наречия *более*, числительного *тысяча*, местоимения *вы*, *такое*, *свое*, *этот* (в 1,6–1,7), *свой* (в 2,6), *всякий* (в 5), *что* (в 13,7), предлога *против*. Падение частотности слов *война*, *сила*, *вопрос*, *мир*, *путь*, *товарищ*, *народ*, *армия*, *партия*, *труд*, *развитие*, *завод*, *производство*, *рабочий*, *борьба*, *советский*, *молодой*, *новый*, *великий* вполне предсказуемо, поскольку они тесно связаны с советской пропагандой и обилием текстов военной тематики в словаре Л. Н. Засориной. Так,

частотность слова *развитие* в газетно-журнальном подкорпусе составляет 589 ipm, тогда как в художественной прозе всего 22; аналогично для слов *вопрос* (1321 и 358), *мир* (1659 и 384), *новый* (2919 и 883), *путь* (815 и 324), *против* (581 и 216). 451 из 575 употреблений слова *камень* и 546 из 625 случаев использования слова *стекло* относятся к научным и публицистическим текстам. Высокая частотность данных слов объясняется тем, что среди источников словаря есть книги «Занимательная минералогия» А. Е. Ферсмана и «Стекло» Н. Н. Качалова. Высокая частотность глаголов движения (*уйти*, *выходить* и др.) отчасти объясняется большим объемом драматургического подкорпуса в словаре Л. Н. Засориной.

Более убедительно выглядит сравнение однотипных подкорпусов. Так, слово *путь* в публицистике во второй половине XX в. стало употребляться вдвое реже, тогда как его частотность в художественной прозе снизилась только на 20 %. Похожая ситуация с прилагательным *новый*. Частотность слова *мир* в публицистике тоже упала почти вдвое, но при этом в художественной литературе почти на 20 % возросла. Заметно нарастание «человечности» публицистической речи: прилагательное *хороший*, соматизмы *голова*, *глаз*, *нога* стали в публицистике заметно частотнее, хотя в художественной прозе не изменили или снизили частотность.

На основании проведенного исследования можно сделать следующие выводы: 1) для достоверной оценки лексических процессов необходимы данные по частотности, полученные на основе стилистически однотипных текстов, которые относятся к четко ограниченному периоду (желательно не более 10 лет); 2) наиболее достоверные оценки в настоящее время можно получить на основе НКРЯ, который позволяет задать и хронологические, и жанрово-стилистические рамки; 3) поскольку в НКРЯ входят и довольно объемные тексты, достоверность оценок может в отдельных случаях снижаться, однако интерфейс корпуса позволяет откорректировать полученные данные.

#### Список литературы

1. Частотный словарь русского языка : ок. 40 000 слов / Под ред. Л. Н. Засориной. – М. : Русский язык, 1977. – 936 .
2. Шаров, С. А. Частотный словарь [Электронный ресурс] / С. А. Шаров. – Режим доступа: <http://www.artint.ru/projects/frqlist.php>. – Дата доступа: 24.09.2019.
3. Ляшевская, О. Н. Новый частотный словарь русской лексики [Электронный ресурс] / О. Н. Ляшевская, С. А. Шаров. – Режим доступа: <http://dict.ruslang.ru/freq.php?> – Дата доступа: 07.11.2019.
4. Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru>. – Дата доступа: 07.11.2019.