

*А. Д. Козеренко, А. Н. Баранов, М. М. Вознесенская,
Д. О. Добровольский, К. Л. Киселева
A. D. Kozerenko, A. N. Baranov, M. M. Voznesenskaya,
D. O. Dobrovolsky, K. L. Kiselyova*

КВАНТИТАТИВНЫЕ ХАРАКТЕРИСТИКИ ИДИОМЫ КАК ПОКАЗАТЕЛЬ ЕЕ СТИЛИСТИЧЕСКИХ СВОЙСТВ¹

CORRELATION BETWEEN IDIOM FREQUENCY AND ITS STYLISTIC PROPERTIES

Аннотация. В статье рассмотрена связь между частотой употребления идиомы в жанрово-специфичных корпусах текстов и ее стилистической характеристикой. Проанализированы употребления ста идиом в художественной литературе, корпусе детективов и публицистике. Показано, что частотность идиом в разных корпусах может служить инструментом обоснования, проверки и коррекции стилистических помет.

Abstract. The paper deals with the use of 100 idioms in three text corpora representing fiction, detective stories and journalism. It is shown that correlations between stylistic properties of a given idiom and its frequency in different types of discourse may be used as a tool for correct usage labeling in dictionaries.

Стилистические пометы остаются одной из наименее формализованных сторон лексикографической деятельности. В существующих словарях – в том числе академических и нормативных – используются различные наборы стилистических помет, причем поскольку алгоритмов расстановки помет не существует, даже такие популярные пометы, как *разг.*, *прост.*, *спец.*, на практике используются составителями словарей совершенно по-разному. Задача объективации процедуры

¹ Работа выполнена при поддержке РГНФ, грант 12-34-10413, и Министерства образования и науки РФ, заявка 2012-1.1-12-000-3004-052 (Разработка интерактивной системы «Корпус текстов по русской фразеологии»).

расстановки помет, а также их внутренней системной организации представляется весьма актуальной.

В отделе экспериментальной лексикографии Института русского языка РАН авторы данной статьи работают над созданием частотного словаря русской идиоматики². Особенность проекта в том, что замеры частоты употребления идиом проводятся на различных по тематике и жанрам корпусах текстов. Анализ показывает, что многие фразеологизмы характерны лишь для некоторых типов дискурса и речевых жанров. Например, идиома *город на Неве* является типичным журнализмом: в корпусе публицистики она встречается 167 раз, а в корпусе русской прозы — всего 6. Возникает вполне правдоподобная гипотеза, что данные о частоте употребления могут указывать на стилистические характеристики идиомы, что, в свою очередь, может использоваться как инструмент для проверки и коррекции постановки стилистических помет в словаре. Назовем такой подход к определению стилистических характеристик идиом **корпусно-ориентированным**.

Разумеется, не все стилистические характеристики идиом (и слов в целом) можно установить, опираясь на корпусно-ориентированную методологию. Дело в том, что создание репрезентативных корпусов текстов для каждого жанра — весьма сложная и трудоемкая задача. Кроме того, далеко не все стилистические пометы в принципе носят жанровый характер. Так, пометы *истор.*, *устар.*, *ирон.*, *груб.*, *эвф.* («эвфемизм») вряд ли можно определять, опираясь только на корпусы текстов.

Система стилистических помет в ФОС³ включает, в частности, пометы *нейтр.*, *журн.*, *книжн.*, *высок.*, *жарг.*

² Баранов А.Н., Вознесенская М.М., Добровольский Д.О., Киселева К.Л., Козеренко А.Д. Проект частотного словаря русских идиом // Компьютерная лингвистика и интеллектуальные технологии: Труды междунар. конференции «Диалог–2012». М., 2012. С. 28–36.

³ Баранов А.Н., Вознесенская М.М., Добровольский Д.О., Киселева К.Л., Козеренко А.Д. Фразеологический объяснительный словарь русского языка. М., 2009.

Особенность этой системы состоит в том, что самая частотная стилистическая характеристика идиом – *разг.* – специально не маркируется пометой, в то время как менее частотная характеристика – *нейтр.* – получает эксплицитную маркировку. Это согласуется с принципом структурности описания (подробнее см. в ФОС) и отражает стилистическое своеобразие идиом.

Для проверки гипотезы о связи между частотой идиомы в жанрово-специфичном корпусе текстов и ее стилистической характеристикой из словника Тезауруса⁴ были выбраны 100 идиом с разными пометами и подсчитана частота их употребления в трех корпусах: Корпусе художественной литературы (КХЛ), Корпусе публицистики (КП) и Корпусе детективов (КД)⁵. Последний был задуман как отражение устной коммуникации. Однако использовать детективы как модель разговорной стилистики можно с оговорками: они содержат не только диалоги, но и повествовательные фрагменты, что влияет на частотность идиом разных стилей. Используемые корпуса можно считать репрезентативными для соответствующих периодов: КХЛ охватывает тексты с 1960-х годов и включает 36,88 млн словоупотреблений, КП – тексты периодических изданий с 1980-х до середины 2000-х годов (24,66 млн), КД – тексты детективов 1970-90-х годов, фрагментарно – произведения начала 2000-х годов (15,8 млн).

Результаты подсчетов в целом соответствуют ожиданиям.

Таблица 1. Частота употребления идиом с пометой *журн.*⁶

	КХЛ	КД	КП
--	-----	----	----

⁴ Баранов А.Н., Добровольский Д.О., Киселева К.Л., Козеренко А.Д. Словарь-тезаурус современной русской идиоматики. М., 2007.

⁵ Корпусы создавались в отделе экспериментальной лексикографии ИРЯ РАН во второй половине 1990-х гг.

⁶ Здесь и далее приводится относительная частота употребления идиом из расчета на 1 млн. словоупотреблений корпуса.

на злобу дня <i>журн.</i>	0,11	0,06	1,05
Город на Неве <i>журн.</i>	0,05	0,82	6,97
империя зла <i>журн.</i>	0,46	0,63	1,54
верхушка/вершина айсберга <i>журн.</i> ; надводная/видимая часть айсберга <i>журн.</i>	0,19	1,39	2,07
по гамбургскому счету <i>журн.</i> ; гамбургский счет <i>журн.</i>	0,24	0,25	1,18
священная корова <i>журн.</i>	0,30	0,06	1,30
утечка мозгов <i>журн.</i>	0,14	0,13	2,07
момент истины <i>журн.</i>	1,06	0,63	3,97
коридоры власти <i>журн.</i>	0,05	0,63	4,01

Из табл. 1 видно, что обычно идиомы с пометой *журн.* встречаются в КП чаще, чем в других корпусах. Обнаружились и очевидные ошибки в приписывании помет. Так, идиома *первая ласточка*, отнесенная к разговорным, представлена в КП заметно большим количеством вхождений: 0,35 (КХЛ), 0,82 (КД), 2,39 (КП). Это существенный аргумент в пользу изменения стилистической пометы. Однако в некоторых случаях высокая частота употребления в жанрово-специфичном корпусе противоречит интуиции и вряд ли может повлечь изменение стилистических характеристик. Так, идиома *последний из могикан*, получившая в Тезаурусе помету *книжн.*, существенно чаще встречается в КП, чем в других корпусах: 0,08 (КХЛ), 0,19 (КД), 1,54 (КП). Связь с соответствующим произведением Ф. Купера в данном случае оказывается существенно более важным фактором, чем реальное употребление в том или ином типе дискурса. Идиома *око за око, [зуб за зуб]*, также отнесенная к книжным, обнаруживает более ровное распределение по жанрово-специфичным корпусам: 0,52 (КХЛ), 0,51 (КД), 0,69 (КП). И в этом случае очевидная цитатность идиомы и

использование устаревшей лексической формы *око* оказываются более значимыми, чем данные по частоте употребления. Ср. аналогично идиомы с пометой *книжн.*: *эзопов язык* – 0,19 (КХЛ), 0,32 (КД), 0,57 (КП); *фиговый листок* – 0,33 (КХЛ), 0,44 (КД), 1,01 (КП). Близким образом ведут себя некоторые идиомы с пометой *высок.* Идиома *кануть в Лету* чаще встречается в КП: 1,03 (КХЛ), 0,76 (КД), 3,77 (КП), однако и в этом случае более важным оказывается имя собственное *Лета*, «перевешивающее» фактор частотности. Отметим, что результат конкуренции факторов «цитатности» и частотности не всегда предсказуем. Так, в противоположность только что рассмотренным примерам идиома *дистанция огромного размера*, помеченная как *журн.*, употребляется в КП существенно чаще: 0,05 (КХЛ), 0,19 (КД), 0,61 (КП), то есть «цитатное» происхождение не оказывается решающим аргументом для приписывания пометы *книжн.* Не исключено, что это связано с ее очевидной стилистической ущербностью, отражающей в первоисточнике солдафонную неуклюжесть речи Скалозуба.

Стилистические характеристики «сниженное», «грубое», «просторечное» и «жаргонное» должны отражаться в большей частоте употребления в КД.

Из табл. 2 видно, что гипотеза о связи стилистических характеристик «сниженное», «грубое», «просторечное» и «жаргонное» с жанром корпуса в целом подтверждается. Впрочем, и здесь возможны исключения.

Таблица 2. Частота употребления идиоом с пометой *снижен.*, *груб.*, *-прост.* и *жарг.*

	КХЛ	КД	КП
на хрен <i>груб. эвф.</i>	2,55	6,33	0,28
до зарезу нужен... <i>прост.</i> ; позарез нужен... <i>прост.</i>	0,62	7,41	1,34
на фиг <i>снижен.</i>	3,20	3,67	1,09

выпасть в осадок <i>снижен.</i>	0,49	1,08	0,53
гнилой базар <i>груб. жарг.</i>	0,11	0,13	0,08
чтоб я сдох <i>снижен.</i> ; чтоб ты сдох <i>снижен.</i>	0,05	0,63	0,00
братъ на понт <i>жарг.</i>	0,57	1,27	0,12
без утайки <i>прост.</i>	0,54	1,90	0,45
без балды <i>жарг.</i>	0,11	0,19	0,00
всю плешь проесть <i>снижен.</i>	0,05	0,95	0,08
держи карман [шире] <i>снижен.</i>	0,57	0,95	0,12
ни в какую <i>прост.</i>	2,33	2,91	1,58
нашел дурака <i>груб.</i>	0,24	0,57	0,04
послать... в задницу <i>груб. эвф.</i>	0,14	0,38	0,12
японский городской <i>снижен. эвф.</i>	0,16	0,25	0,00
как пить дать <i>прост.</i>	1,60	3,48	0,45
того [и] гляди <i>прост.</i>	3,12	4,49	2,60
спешу и падаю <i>снижен.</i>	0,03	0,06	0,04
ваньку валять <i>жарг.</i>	0,65	1,46	0,28

Так, идиома *забить стрелку/стрелу* будучи жаргонной, практически одинаково часто встречается в КД и в КХЛ: 0,54 (КХЛ), 0,51 (КД), 0,20 (КП). Идиома на *полном серьезе*, опять-таки будучи жаргонной, более частотна в КП: 0,70 (КХЛ), 2,15 (КД), 2,35 (КП). Идиома *Вася Пушкин*, скорее, ощущается как сниженная, однако она чаще всего встречается в КП: 0,03 (КХЛ), 0,06 (КД), 0,28 (КП). Идиома *навострить лыжи* отнесена в Тезаурусе к жаргонным, однако наиболее частотна она в КХЛ: 0,24 (КХЛ), 0,13 (КД), 0,04 (КП). Наличие грубых и жаргонных идиом в КХЛ объясняется неоднородностью художественного произведения, включающей в себя наряду с собственно-

авторским повествованием языковые структуры, ориентированные на речь героев, и саму прямую речь персонажей. Что касается такого рода идиом в КП, оно связано, по всей видимости, со спецификой современных СМИ, ориентированных на устную речь.

Наименьшую корреляцию с жанрово-специфичными корпусами обнаруживают идиомы со значимым отсутствием пометы («разговорное») и идиомы, помеченные в Тезаурусе как *нейтр.* («нейтральное»). По идее разговорные идиомы должны быть более частотны в КД, а нейтральные – обнаруживать относительно равномерное распределение по корпусам. Однако это встречается довольно редко.

Из табл. 3 видно, что идиомы *без году неделя, направо и налево, вдоль и поперек* (предположительно разговорные) действительно чаще всего обнаруживаются в КД, однако очевидно разговорные идиомы *на днях, для галочки* существенно чаще встречаются в КП.

Столь же неопределенна ситуация со стилистически нейтральными идиомами. Так, идиома *рано или поздно* ощутимо чаще обнаруживается в КП. Частота употребления нейтральной идиомы *до сих пор* в КП во много раз выше, чем в двух других корпусах. В то же время нейтральная идиома *в конце концов* в два с лишним раза чаще встречается в КХЛ и в КД, чем в КП. Для нейтральных идиом существенным фактором, влияющим на частотность, оказывается семантика. Так, *в конце концов* служит для изложения последовательности событий и поэтому более частотна в нарративных текстах, чем в СМИ. Аналогичными частотными характеристиками обладает и нейтральная идиома *то и дело*, описывающая повторяющиеся ситуации.

Таблица 3. Частота употребления идиом с пометой *нейтр.* и со значимым отсутствием пометы («разговорное»)

	КХЛ	КД	КП
без году неделя	0,52	1,14	0,24

вдоль и поперек	2,28	2,34	1,62
направо и налево; налево и направо	2,96	4,18	3,53
рано или поздно <i>нейтр.</i>	12,47	19,62	21,29
на днях	10,28	14,87	67,48
до сих пор <i>нейтр.</i>	75,30	99,56	207,62
в конце концов <i>нейтр.</i>	109,90	106,77	48,58
то и дело <i>нейтр.</i>	28,96	32,97	13,50
для галочки	0,27	0,25	0,61

Выявленные частотные характеристики идиом объясняются различными факторами. Очевидно, что здесь может играть роль неоднородность корпусов. При создании частотного словаря русских идиом существующие корпусы будут переработаны и уточнены. С другой стороны, к дискурсивным в ФОС и Тезаурусе отнесены пометы *книжн.*, *журн.*, *жарг.*, *прост.*, тогда как характеристики «разговорное» и «нейтральное» не являются исключительно дискурсивными. Помета *нейтр.* используется в Тезаурусе омонимично, как для идиом, нечувствительных к типу дискурса, так и для маркировки стилистически нейтральных идиом. Разговорный характер идиом (значимое отсутствие пометы в словнике ФОС и Тезауруса) также может интерпретироваться двояко: дискурсивно (как принадлежность к устной речи) и стилистически. Такая специфика использования пометы *нейтр.* и значимого отсутствия пометы в определенной степени объясняет разброс в частотных характеристиках идиом по разным корпусам.