

*Н.И. Кулакова*

### **ИСПОЛЬЗОВАНИЕ КОРПУСНЫХ МАТЕРИАЛОВ ДЛЯ ИССЛЕДОВАНИЯ НЕКОТОРЫХ ЛЕКСИКОГРАФИ- ЧЕСКИХ ВОПРОСОВ**

В лексикологии и лексикографии уже с античных времен принято рассматривать синонимию, антонимию, омонимию в качестве важнейших отношений между лексическими единицами. Вполне понятно, что наличие таких отношений способствует упорядоченности словаря как сложной системы. Однако до сих пор почти не изучался вопрос о том, в чем именно проявляется структурированность словаря «за счет» наличия синонимов, антонимов, омонимов. Так, обычно остается за рамками рассмотрения, какой из группы омонимов или синонимов занимает доминантное положение в паре (тройке) – и, шире, в лексиконе как таковом, относится к ядру, а какой оказывается на периферии, редко встречается в речевом опыте, связан минимальным числом связей с другими единицами. Между тем кажется ясным, что это важно, если обращаться к рассмотрению реальных механизмов функционирования ментального лексикона. Моделировать более точно существующие между единицами связи, уточнять между ними «расстояние» («психолингвистическое» и «текстовое») позволяет соединение корпусных и экспериментальных методик.

Нами были исследованы следующие пары омонимов: *брак* (2 омонима), *дисциплина* (2), *ключ* (2), *коса* (3), *ласка* (2), *мир* (2), *норка* (2), *пол* (2), *пост* (2), *предложение* (2), *причастие* (2), *стан* (3), *вид* (2), *следствие* (2), *пост* (2), *лук* (2).

Изучались также синонимические пары: *труд* – *работа*, *враг* – *недруг*, *ложь* – *неправда*, *дар* – *подарок*, *печаль* – *тоска*, *метель* – *буря*, *враг* – *неприятель*, *ребенок* – *кроха*, *путь* – *тропа*, *мошенник* – *вор*, *болезнь* – *недуг*, *врач* – *доктор*, *разговор* – *беседа*, *вьюга* – *пурга*, *друг* – *приятель*, *путь* – *дорога*, *ложь* – *вранье*, *ребенок* – *мальш*, *враг* – *противник*, *болезнь* – *хворь*, *ме-*

*тель – выюга, сила – мощь, пища – еда, печаль – грусть, работа – занятие, судьба – фортуна, труд – дело, талант – дар, мошенник – плут, крик – вопль.*

Омонимичные пары отбирались по Словарю современного русского языка<sup>1</sup>. Синонимичные пары – по Словарю синонимов русского языка под ред. А.А. Александровой<sup>2</sup>. Намеренно некоторые из единиц повторялись два-три раза, например, *болезнь – недуг, болезнь – хворь*, или *труд – работа, труд – дело, работа – занятие*, что позволило моделировать некоторые семантические поля в лексиконе.

Синонимичные единицы принято считать близкими по смыслу, тесно связанными, а омонимичные – несвязанными. Однако, ссылаясь на ряд экспериментальных методик (семантического шкалирования, ассоциативный эксперимент и др.) можно утверждать, что «как связанность, так и не связанность не являются абсолютными, они характеризуются определенной мерой, и, соответственно, возможными цепочками переходов»<sup>3</sup>.

Мы исходили в исследовании из того, что степень связанности или несвязанности единиц может быть различной, а их соотношение – симметрично или асимметрично. В первом случае «вес» обеих единиц равен или соотносим, связь в такой паре имеет зеркальную направленность, во втором – одна из единиц занимает доминантное, лидирующее положение, а связь в такой паре имеет ярко выраженное направление.

---

<sup>1</sup> Словарь русского языка: В 4-х т. / РАН, Ин-т лингвистических исследований / Под ред. А.П. Евгеньевой. 4-е изд., стер. М.: Рус. Яз., Полиграф. Ресурсы, 1999.

<sup>2</sup> Александрова А.А. Словарь синонимов русского языка. 1989.

<sup>3</sup> Касевич В.Б. Лексикон и лексикология. Основные аспекты и проблемы лексикологии // Типологические обоснования в грамматике: К 70-летию проф. В.С. Храковского / Отв.ред. А.П. Володин. – М.: Знак, 2004. С. 201–215.

Что позволяет судить о симметричности/ асимметричности отношений в паре?

**Во-первых**, данные частотных словарей и Корпусов: различная частотность единиц в паре, когда слова относятся к разным частотным классам, и одна из единиц широко употребляется, частотна, а другая мало употребляется, предполагает асимметрию отношений. Нами оценивалось распределение частот встречаемости единиц по материалам Национального корпуса русского языка (НКРЯ), <http://ruscorpora.ru>, на момент оценки ок. 35 млн. с/у).

**Во-вторых**, материалы психолингвистических экспериментов. В нашем исследовании испытуемым предлагалось придумать по 3 предложения с каждым из омонимов, затем оценивалось, какой из омонимов оказывался более частотным и «актуальным» по всем контекстам и по первому. Отдельно испытуемые оценивали субъективную частотность каждого из синонимов и их близость в паре (по пятибалльной шкале).

Обширный материал экспериментов с носителями языка представляет также Русский ассоциативный словарь<sup>1</sup> (РАС). Данные именно этого Словаря использовались нами при классификации пар. Каждый синоним из пары при этом рассматривался нами в качестве СТИМУЛА и оценивалось его ассоциативное поле (в частности, встретился ли второй синоним из пары в качестве реакции, была ли она частотной или периферийной). Когда наиболее частотной реакцией к одной из единиц в паре являлась вторая единица, а та в свою очередь, была наиболее частотной (или одной из частотных) реакцией к первой, единицы считались сопоставленными (имеющими зеркальную направленность в словаре), а соотношение симметричным.

При оценке ассоциативного поля омонимичных единиц и ассоциативной значимости каждого из омонимов подсчитывалось

---

<sup>1</sup> Русский ассоциативный словарь. В 2 т. / Ю.Н. Караулов и др. М.: ООО «Издательство Астрель»: ООО «Издательство АСТ», 2002.

процентное соотношение реакций, относящихся к двум разным значениям (омонимам). Кроме того, учитывались наиболее частотные реакции (к какому из омонимов они относились). Если ассоциативно более значимым оказывался один из омонимов, соотношение считалось асимметричным, если число реакций к каждому из омонимов было примерно равным, – симметричным.

На корпусе русских текстов нами было получено распределение частот встречаемости рассматриваемых единиц; эти корпусные данные были сопоставлены с теми, что были в психолингвистических экспериментах, а также по материалам Русского ассоциативного словаря.

В результате, каждый из источников в отдельности (Корпус и экспериментальный материал, наши эксперименты и данные РАС) позволил пары омонимов и пары синонимов разбить на четыре группы:

- с абсолютно симметричным соотношением в паре;
- с симметричным соотношением;
- тяготеющие к асимметричным;
- с асимметричным соотношением.

**По материалам Корпуса:** рассматриваемые единицы были разделены на три класса встречаемости на основании частоты встречаемости в НКРЯ. Найденные частоты встречаемости нанесли на логарифмическую шкалу: минимальное значение соответствовало нулевой, максимальное – максимальной частоте встречаемости по словарю; полученная таким образом шкала делилась на три равные части, а слова по нахождению в той или иной части – на три класса встречаемости.

Соотношение в парах (по данным Корпуса) абсолютно симметрично в том случае, когда у двух единиц близкая или одинаковая частотность, например, *талант* (6000 с/у) – *дар* (6000 с/у), симметричное, когда единицы по частотности попадают в один класс, например, частые: *ложь* (3000 с/у) – *неправда* (2000), тяготеет к асимметричному (единицы по частотности оказываются в

двух соседних классах, например, *ложь* (3000 с/у) – *вранье* (800 с/у), асимметричное (единицы относятся к разным частотным классам, которые не соседствуют, например, *ребенок* (60000 с/у, очень частое) – *кроха* (400 с/у, редкое).

Результаты группировки пар по материалам корпуса представлены в табл. 1.

Таблица 1. Симметричность соотношения в парах по материалам НКРЯ

По Корпусу			
Абс. симм. соотношение	Симм. соотношение	Тяготеют к асимметричным	Асимметричное соотношение
<i>Стан 1,2,3</i>	<i>Следствие 1,2</i> <i>Норка 1,2</i> <i>Вид 1,2</i> <i>Пол 1,2</i> <i>Предложение 1,2</i> <i>Коса 1,2</i>	<i>Пост 1,2</i> <i>Мир 1,2</i> <i>Лук 1,2</i> <i>Дисциплина 1,2</i> <i>Брак 1,2</i> <i>Коса 1,3</i> <i>Причастие 1,2</i>	<i>Ласка 1,2</i> <i>Ключ 1,2</i>
<i>Талант – дар</i>	<i>Ложь – неправда</i> <i>Мошенник – плут</i> <i>Труд – работа</i> <i>Дар – подарок</i> <i>Печаль – тоска</i> <i>Метель – буря</i> <i>Враг – неприятель</i> <i>Крик – вопль</i> <i>Врач – доктор</i> <i>Разговор – беседа</i> <i>Вьюга – пурга</i> <i>Путь – дорога</i> <i>Пицца – еда</i> <i>Печаль – грусть</i> <i>Труд – дело</i> <i>Талант – дар</i>	<i>Ложь – вранье</i> <i>Болезнь – недуг</i> <i>Мошенник – вор</i> <i>Враг – недруг</i> <i>Ребенок – малыши</i> <i>Путь – тропа</i> <i>Друг – приятель</i> <i>Болезнь – хворь</i> <i>Сила – мощь</i> <i>Работа – занятие</i> <i>Судьба – фортуна</i>	<i>Ребенок – кроха</i>

**По данным эксперимента:** в эксперименте с омонимичными единицами приняли участие 85 испытуемых, а в экспериментах с синонимичными единицами – 34 испытуемых от 18 до 75 лет.

У испытуемых в экспериментальной ситуации проявилась некая «презумпция связанности» омонимичных единиц: они пытались в ответах объединить соответствующие омонимы, сталкивали их в одном контексте (*брак в браке, мир в мире, норка спряталась в норке, следствие из следствия было следующим*), рассуждали («*Норка в норке прячется, потому так и называется, Ласка пушистая, нежная, ласковая*»).

При оценке близости пар синонимичных единиц согласие по всем парам слов в среднем для языкового материала получилось весьма высокое: 90% всех пар получили единодушное, хорошее или среднее согласие испытуемых. Полученные данные можно считать достоверными.

Признак асимметричность vs. симметричность соотношения единиц в паре омонимов предопределяет стратегию **выбора** испытуемых в эксперименте.

Построенная нами структура вероятностей выбора испытуемыми того или иного омонима для групп асимметричных единиц выглядит следующим образом: в 71% случаев из 100% для своего первого речевого контекста испытуемые выбирают «наиболее частотный, ассоциативно значимый» омоним. После чего этот же омоним во втором речевом контексте повторяется в 57% случаев, а в 24% случаев – другой омоним. 23% испытуемых из тех, кто дважды употребил ОМОНИМ1, и в третий раз используют ОМОНИМ1.

Для групп симметричных омонимов картина вероятностей другая. Вероятность выбора той или иной единицы в первом контексте практически равная, сопоставимая (56% и 44%), во втором контексте – 48% и 52%.

При исследовании синонимичных единиц была обнаружена отчетливая тенденция: среди наиболее близких (по результатам шкалирования) пар нет пар с асимметричным соотношением, векторной направленностью, и наоборот, чем меньше близость в паре, тем с большей вероятностью соотношение в такой паре асимметричное или тяготеющее к асимметричному.

Результаты корпусных и экспериментальных исследований, а также данных словаря сопоставимы. Ассоциативная частотность (ассоциативная сила) соотносится с частотой встречаемости (по корпусу текстов) данных единиц, а также с субъективной частотностью. Для слов с высокой субъективной частотностью, в целом, характерно иметь высокую частотность языковую, и наоборот, чем ниже мы спускаемся по субъективной шкале частот, тем ниже показатели объективной частотности.

Различная частотность единиц в паре (субъективная и по Корпусу) предполагает асимметрию отношений в паре в том числе с точки зрения представления в ментальном лексиконе.

В асимметричных парах значительно более частотная (по Корпусу) единица в паре сравнительно с другой единицей более актуальна для носителей языка и быстрее извлекается из словаря. А когда соотношение в паре симметрично и частотность двух единиц (как по Корпусу, так и субъективная) сопоставима, тогда данные единицы «равно актуальны» и для носителей языка, близость слов в такой паре оценивается испытуемыми высоко.

При сопоставлении результатов были выявлены и расхождения, например, расходятся данные по распределению языковой и субъективной (в данном случае ассоциативной) частотностями для омонимов *СТАН*. Для преимущественно студенческой группы испытуемых омоним «корпус тела» оказался актуализирован в большей степени, чем «стоянка» (43% и 19% соответственно, а по первой реакции 69% и 16%), тогда как в Корпусе (27% и 43%).

Для единиц, связанных отношениями синонимии, также имеются расхождения, например, пары *ложь – вранье*, *болезнь – хворь*, *ребенок – малыши* по данным Корпуса тяготеют к асимметричны, по данным же РАС – имеют относительно симметричное соотношение.

Причины расхождений требуют дополнительного обсуждения, в том числе, необходима стратификация рассматриваемых единиц по числу и типу значений, а корпусных данных по жанровому типу и времени создания текста.