

М.Ю. МУХИН

*(Уральский федеральный университет,
г. Екатеринбург, Россия)*

УДК 81'42
ББК Ш105.51

НЕТИПИЧНАЯ ЛЕКСИЧЕСКАЯ СОЧЕТАЕМОСТЬ: ФОРМАЛИЗАЦИЯ ТЕРМИНА И АНАЛИЗ ТЕКСТА⁴

Аннотация: В статье анализируются возможности формализации индивидуальной лексико-семантической сочетаемости на базе корпуса художественных текстов. Представлена методика сопоставительного статистического анализа авторской синтагматики. Единицы анализа — лексические биграммы, которые характерны для текстов одного писателя в отличие от других авторов. Приведен пример контекстологического анализа лексической сочетаемости в романах М. Булгакова.

Ключевые слова: нетипичная лексическая сочетаемость, синтагматика, корпус текстов, лексическая биграмма, статистический анализ текста

1. Вводные замечания

Нетипичная (или нестандартная) лексико-семантическую сочетаемость считается признаком индивидуального стиля, а семантический эффект, возникающий вследствие специфичного словоупотребления, отражен в разных терминах: «актуализация» (Пражский кружок), «остранение» (В. Шкловский), «экспрессема» (В.П. Григорьев). Как писал В.Г. Адмони, «из дистантной постановки компонентов, непосредственно связанных друг с другом грамматически и семантически, из столкновения не сочетающихся по каким-либо признакам лексем и грамматических форм» в речевом ряду возникает «синтагматическое напряжение» [Адмони 1973: 117]. Иными словами, нетипичная сочетаемость приводит к различным приращениям смысла. Кон-

⁴ Исследование выполнено при поддержке Российского гуманитарного научного фонда — проект № 13-04-00322 «Русская лексика как межчастеречная система: полное идеографическое описание в лексикографических параметрах».

статация этого факта привела к множеству частных идиостилевых наблюдений. Однако формализация лексической синтагматики всегда была затрудненной по двум причинам: 1) множественность синтагматических связей лексики (количественный фактор), а также 2) невозможность полной автоматизации анализа семантического варьирования слов и их контекстных партнеров.

В то же время поиск методов измерения нетипичности (например, «коэффициентов креативности») давно занимает умы лингвистов — см. об этом в [Гридина 2009: 7]. Совсем недавно в Екатеринбурге, на защите докторской диссертации И. В. Ружицкого (октябрь 2015 г.), опять возник вопрос критериев нестандартности авторской сочетаемости, так как в работе признается интуитивность исследователя на начальном этапе выборки материала [Ружицкий 2015: 171]. В диссертации, как и во многих др. источниках, отмечается, что особая сочетаемость слов приводит к их игровому употреблению [Там же: 172]. Однако лингвистическую аргументацию, которую можно предъявить в отношении каждого нестандартного контекста, тем не менее трудно представить в строгой алгоритмизированной форме, пригодной для сплошной обработки множества контекстов. Классические представления об изучении лексической синтагматики получают новое развитие благодаря появлению текстовых корпусов и программ автоматической обработки речи. Неслучайно исследователи ставят вопрос о том, «в какой мере словари и корпуса текстов совместимы как носители синтагматической информации, в чем их преимущества и недостатки, каким источникам надо отдавать предпочтение» [Митрофанова, Крылов 2006].

2. Параметры исследования и понятие «лексическая биграмма»

Исследование строится на лемматизированном корпусе знаменательных слов (около 1,5 млн.), построенном на основании 16 текстов М. Булгакова, В. Набокова, А. Платонова и М. Шолохова. Объективировать индивидуально-авторскую модель сочетаемости на основании всех словоупотреблений в корпусе практически невозможно. Для этого следовало бы провести сплошную синтаксическую и семантическую разметку текстов

со снятием омонимии. Синтагматическая вариативность настолько велика, что в любом случае материал должен быть ограничен определенным набором слов или количеством контекстных партнеров. Возникает проблема и с выработкой общего основания для сопоставления синтагматических особенностей текстов разных авторов.

В данном исследовании модель индивидуально-авторской лексической сочетаемости основана на систематизации и анализе лексических биграмм — пар слов, употребляющихся в одном фразовом контексте и, вероятно, связанных семантическими отношениями. В таком лексическом понимании термин «биграмма» используется в работах [Захаров, Хохлова 2008; Митрофанова и др. 2008; Ягунова, Пивоварова 2010а, 2010б].

Для решения поставленной задачи сформирован массив знаменательных лексических пар в пределах одного предложения, в который входят как левосторонние, так и правосторонние контекстные партнеры. Например, в предложении *Состоялась защита докторской диссертации* выделяется 12 знаменательных лексических биграмм:

Лемма-1	Лемма-2	Лемма-1	Лемма-2	Лемма-1	Лемма-2	Лемма-1	Лемма-2
состояться	защита	защита	состояться	докторский	состояться	диссертация	состояться
состояться	докторский	защита	докторский	докторский	защита	диссертация	защита
состояться	диссертация	защита	диссертация	докторский	диссертация	диссертация	докторский

Полученный массив содержит более 10,5 млн. сочетаний и в чистом виде не пригоден из-за большого количества пар лексем, не связанных семантически (т. е. лишнего материала, «шума»). Количество таких «незначимых» биграмм зависит от длины фразы, расположения членов пары в разных и частично автономных фрагментах конструкции (например, во вставной конструкции и основной части предложения, авторской и персонажной речи). Кроме того, существует множество биграмм, встречающихся один раз и в одном тексте, т. е. претендующих на статус случайных или обусловленных исключительно сюжетом произведения.

В связи с этим естественно возникает вопрос об ограничении материала, которое выражается: 1) в сужении расстояния между словами внутри фразы (т. е. определении «окна»), 2) в количественном ограничении лексики, 3) в определении состава

значимых для автора биграмм — материала дальнейшего контекстологического анализа.

1-е ограничение определено опытным путем. Исходя из специфики русского словосочетания (состоящего из двух, трех и более слов), возможности дистантного расположения его элементов и вероятного наличия между членами биграммы служебных слов и местоимений расстояние между словами было принято равным 5. Таким образом, максимальное количество слов, с которыми данное слово, находящееся не в начальной и не в конечной части предложения, может составлять бигramму, равно 10 (5 с левой стороны и 5 — с правой). Количество биграмм, извлеченных из корпуса, сокращается в таком случае до 2,9 млн., употребленных 3,6 млн. раз.

Второе ограничение связано с возможностями дальнейшей семантической обработки материала, который слишком велик по объему. В то же время, для того, чтобы судить об индивидуально-авторских различиях, необходимо единое основание для сопоставления идиостилей. В качестве такого основания в работе избрана специфика контекстного окружения слов, часто встречающихся у всех авторов и во всех 16 текстах. Обязательным членом биграммы, таким образом, должно быть слово типа *человек, глаз, говорить, видеть, один* и т. п. Теоретически так можно выявить авторские приоритеты, выраженные в работе с «универсальным» языковым материалом. Количество таких биграмм около 400 тыс. в 700 тыс. контекстов.

Третье ограничение напрямую связано с выработкой методики оценки значимости той или иной биграммы для идиостиля писателя.

3. Критерии авторской сочетаемости, «синтагматическая активность» лексики.

Итак, нам необходим критерий отбора материала (значимых для идиостиля лексических пар), который бы выявил особенности лексической сочетаемости в текстах каждого автора и в то же время не испытывал сильной зависимости от объема и содержания конкретного произведения.

Чтобы уравнивать идиостилевой статус текстов одного автора, имеющих разный объем, и исключить биграммы, которые

встречаются единично и/или только в одном произведении, выработана сопоставительная частотная методика. В материал включены биграммы, которые одновременно:

- 1) встречаются в одном тексте не менее двух раз,
- 2) встречаются не менее, чем в двух текстах одного автора,
- 3) вообще не попадают в список избранных биграмм других авторов.

Такая установка позволяет отобрать пары слов, часто встречающиеся в разных произведениях одного автора и совсем не характерные для других писателей. После применения описанных ограничений отобрано 2400 лексических пар, употребленных 21350 раз.

Чтобы выявить синтагматические приоритеты четырех авторов, определим набор слов, имеющих в творчестве каждого из них наибольшее количество употреблений с оригинальными контекстными партнерами. Например, оригинальные, не встречающиеся у других авторов биграммы со словом *жизнь* в произведениях А. Платонова употреблены более чем в 200 контекстах, и среди ближайших частотных контекстных партнеров этого слова — *счастье, чувство, истина* и др. У М. Шолохова, при всем объеме произведений, контекстов с оригинальными биграммами (например, *проклятая жизнь, недолгая жизнь, сложилась жизнь* и др.) около 50, еще меньше у В. Набокова — 13 (со словами *сторона жизни* и *душа, воспринимающая жизнь*) и у М. Булгакова — 4 (только употребление словосочетания *театральная жизнь*).

Способность слов иметь особое контекстное окружение мы называем синтагматической активностью. Так, если исходить из сопоставления авторских показателей, наибольшую синтагматическую активность в творчестве М. Булгакова, проявляют, в частности, глаголы речевой деятельности *говорить, сказать, отвечать*, а также существительное *голос*.

Просто перечислив лексемы, невозможно отразить специфику их семантических отношений с контекстными партнерами. Поэтому оригинальные для синтагматики каждого автора контексты этих доминантных слов были подвергнуты контекстологическому анализу, методика которого строится на выявлении

семантических отношений между исходным словом и его контекстными партнерами. При семантической разметке контекстов мы ориентировались на классификации, принятые в идеографических словарях, созданных под руководством проф. Л.Г. Бабенко лексикографической группой «Русский глагол» — см., например [Словарь-тезаурус синонимов 2007].

Рассмотрим авторские синтагматические приоритеты на частных примерах анализа произведений Михаила Булгакова. Их названия в дальнейшем сокращены следующим образом: *БГ* — «Белая гвардия», *ЖМ* — «Жизнь господина де Мольера», *ТР* — «Театральный роман», *ММ* — «Мастер и Маргарита».

4. Примеры контекстологических обобщений: лексика речевой деятельности в романах М. Булгакова

Художественное повествование, как правило, не обходится без фрагментов внешней и внутренней речи, в которых употребляются глаголы речевой деятельности. Специфика их сочетаемости определяется идиостилевыми предпочтениями. Приведем набор типовых контекстных партнеров синтагматически активной лексики в булгаковских текстах:

- *говорить*: пьеса, более, телефон, роман и др.;
- *сказать*: пьеса, врач, робко, прибавить, достоинство (т. е. с достоинством) и др.;
- *отвечать*: очень, господин, вежливо, совершенно и др.;
- *голос*: прокричать, трубка, неприятный, показаться, совершенно, телефонный, детский и др.

По результатам контекстологического анализа можно сформулировать следующие семантические соответствия избранных слов и их контекстных партнеров.

● Речь (*говорить, сказать*) ↔ тема — произведение искусства (существительные *пьеса, роман*)

«Искусство» — важнейшая денотативная сфера, реализованная в булгаковском творчестве, и тематика разговоров персонажей часто связана с оценкой произведений: *Я хотел вам сказать о вашей пьесе, Мольер, — начал король (ЖМ); Что же говорили о новой пьесе литераторы и парижские знатоки театра?; А что говорил Ликоспастанов о моем романе? (ТР); Не говоря ничего по существу романа, он*

спрашивал меня... (ММ) и многие другие контексты. Переживание творца, связанное с восприятием его произведения (критическим или, наоборот, благожелательным), — лейтмотивная тема трех булгаковских романов.

• Речь (*сказать, отвечать/ответить*) ↔ субъект или адресат речи (существительные *врач, господин*)

Врач как субъект речи из рассмотренных авторов встречается только у Булгакова (ср. с субъектами и адресатами речи у др. писателей, приведенными далее): *«Ага, — сказал врач [Ивану], — а почему так спешили?»* (ММ); *«О том, как сохранить здоровье, вам скажет любой хороший врач»* (ЖМ). Уважительное отношение к собеседнику, страх перед ним или требования субординации предполагают использование слова *господин*: *господин полковник, господин директор и т. п.: Помилуйте, господин офицер, — трясясь в ужасе, ответил голос, — я ничего не говорю* (БГ).

• Речь (*говорить, голос*) ↔ телефонная связь (*телефон, телефонный, трубка*)

Зная булгаковский текст, можно заведомо предположить, что разговор персонажей по телефону в нем встречается чаще, чем у других писателей. Действительно, слово *телефон* писатель употребляет 89 раз (Набоков — 53, Шолохов — 11, Платонов — 8): *«Если дверь приоткрывалась, слышно было, как он говорил по телефону...»* (ТР); *«Да, — говорила горничная в телефон...»* (ММ); *«Тот же голос возник в трубке телефона в шести верстах от сторожки...»* (БГ). По понятным причинам телефон как реалия отсутствует только в «Жизни господина де Мольера».

Телефонная связь для автора не просто техническое средство общения на расстоянии. Невозможность видеть собеседника позволяет сосредоточить внимание на слуховом восприятии, что приводит, например, к использованию множественных характеристик голоса: *тихий, вкрадчивый и развратный; резкий, неприятный; дребезжащий; мягкий, но грустный; препротивный гнусавый* и т. д. Характеризация голоса приводит к специфическим драматургическим решениям: *«Иван Савельевич?» — осведомилась трубка препротивным гнусавым голосом. — «Его*

нет в театре!» — крикнул было Варенуха, но трубка сейчас же перебила: «Не валяйте дурака, Иван Савельевич, а слушайте» (ММ). Ср. также пример из романа «Белая гвардия»: *Г о л о с, не слушая, заквакал в трубке в ответ: «Беглым по урочищу... по коннице...»* (БГ). Активно используется эффект театрализации телефонной речи в связи с образами свиты Воланда (ММ): *Они заняты, — ответила трубка дребезжащим голосом, — а кто спрашивает?; Тихий, в то же время вкрадчивый и развратный женский голос шепнул в трубку: «Не звони, Римский, никуда, худо будет»* (ММ). Особую отстраненность собеседника подчеркивают метонимические переносы (*человек* → *трубка*, *человек* → *голос*), из-за чего возникает впечатление, что голос существует как бы сам по себе, независимо от того, кому он принадлежит. Так, в романе «Белая гвардия» некий штабной голос по телефону осуществляет командование: *Повинуясь телефонному голосу, унтер-офицер Турбин Николай вывел двадцать восемь человек юнкеров и через весь Город провел их согласно маршруту.*

- Речь (*сказать, отвечать/ответить*) ↔ эмоция и поведение персонажа (слова *робко, вежливо, с достоинством*)

Характеризация персонажа через особенности его речи — привычный литературный прием. М. Булгаков во всех романах использует, например, наречие *робко*: *Но вы можете выздороветь... — робко сказал Иван; «Три бубны», — робко сказал Лариосик* (БГ). С одной стороны, оно подчеркивает нерешительность, кроткость, ненавязчивость человека (т. е. его внутренние качества и эмоциональное состояние). С другой стороны, оно дополняет внешний портрет персонажей.

Наречие *вежливо* вообще входит в состав слов, которые, в сравнении с другими тремя авторами, часто употребляет только Булгаков: *Не разрешаю, — вежливо и спокойно ответил ему Мальшев* (БГ); *На это господин Ратабон ответил вежливо, что это ему неизвестно* (ЖМ). Еще раз подчеркнем, что при помощи этих слов персонажи получают не только речевую, но и портретную и психологическую характеристику: *«Позвольте», — сказал я с достоинством* (ГР). В использовании перечисленных наречий мы видим драматургиче-

скую установку писателя на речевое выражение внешних и внутренних качеств человека. При этом часто выражается авторское отношение — например, ироничное: *«Я беру эту пару, мосье», — сказала брюнетка с достоинством, надевая и вторую туфлю* (ММ). Словами главного героя «Театрального романа» Булгаков формулирует этический принцип: *Держаться вежливо, но с достоинством и, боже сохрани, без намека на угодливость* (ТР).

Речевая сфера, кроме романов М. Булгакова, оригинально разработана и в произведениях М. Шолохова. Ее «подробная» реализация в его текстах проявляется, например, во множественных характеристиках человеческого голоса: *дурной, чужой* (в значении качественного признака: «отчужденный, отрешенный»), *командный, плачущий, надтреснутый, спокойный, басовитый, сдержанный, теплый, расслабленный*. Естественно, что в шолоховских произведениях встречается оригинальный набор субъектов и адресатов речи, как и вообще номинаций человека: *казак, старуха, командир, хозяйка* и др.

У других писателей примеров оригинальной синтагматики отмеченных слов гораздо меньше. Для В. Набокова характерны сочетания *сказать бодро* и *следует сказать*, среди субъектов/адресатов речи обнаруживаем слово *шофер*, а голос человека часто раздается откуда-то *сзади*: *«Неслыханное бесстыдство»*, — *комментировал чей-то толстый голос сзади* (Н:Д); *«Перерыв, маэстро»*, — *сказал голос сзади* (Н:ЗЛ). У Платонова можно отметить только несколько особых субъектов или адресатов речи: *машинист, комиссар, мужик*.

Точно так же, в сравнительном ключе, произведено обобщение контекстов других синтагматически активных слов в творчестве четырех писателей.

Таким образом, нетипичная лексико-семантическая сочетаемость может быть определена как свойственная конкретному автору синтагматическая активность слова, наличие у него такого круга контекстных партнеров, который не имеет аналогов в другом идиостиле. При таком понимании нетипичность становится измеримой, а не просто понимаемой интуитивно. С другой стороны, мы отдаем себе отчет в том, что далеко не все контек-

сты синтагматически активных слов интуитивно воспринимаются как нестандартные. В лингвистике существуют подобные противоречия. Например, признание слова неологизмом осуществляется по как минимум двум критериям: интуитивному и формально-хронологическому, — которые в разных случаях расходятся.

Развитие исследования авторской синтагматики перспективно связано с будущими семантически размеченными корпусами, но уже на основании проведенного анализа можно говорить о формализации признаков «нетипичность» и «нестандартность» по отношению к лексическому окружению. Кроме того, среди авторских контекстов, выделенных с применением предложенной методики, есть много случаев, которые и по традиционным лингвистическим критериям можно, без сомнения, отнести к фактам языковой игры.

ЛИТЕРАТУРА

Автоматическая обработка текста. Поиск по биграммам. URL: <http://www.aot.ru/demo/bigrams.html>

Адмони В.Г. Особенности синтаксической структуры в художественной прозе XX в. на Западе / В. Г. Адмони // *Phylologia. Исследования по языку и литературе. Памяти акад. В. М. Жирмунского.* – Л., 1973. – С. 115–124.

Гридина Т.А. К истокам вербальной креативности: творческие эвристики детской речи // *Лингвистика креатива.* – Екатеринбург, 2009. – С. 5-58.

Захаров В.П., Хохлова М.В. Статистический метод выявления коллокаций // *Языковая инженерия: в поиске смыслов: докл. сем. «Лингвистические информационные технологии в Интернете»: XI Всерос. объединенная конф. «Интернет и современное общество».* – СПб. : СПбГУ, 2008. – С. 40–54.

Митрофанова О.А., Белик В.В., Кадина В.В. Корпусное исследование сочетаемостных предпочтений частотных лексем русского языка // *Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной междунар. конф. «Диалог».* – М., 2008. URL: <http://www.dialog-21.ru/dialog2008/materials/html/56.htm> (дата обращения: 01.11.2015).

Митрофанова О.А., Крылов С.А. «Типовой» контекст: слу-

чайность или закономерность? // Компьютерная лингвистика и интеллектуальные технологии: Тр. междунар. конф. «Диалог–2006». М., 2006. URL: <http://www.dialog-21.ru/dialog2006/materials/html/Mitrofanova.htm> (дата обращения: 01.11.2015).

Национальный корпус русского языка. URL: <http://www.ruscorpora.ru> (дата обращения: 01.11.2015).

Ружицкий И.В. Языковая личность Ф.М. Достоевского: лексикографическое представление. Дис. ... д-ра филол. наук. – М., 2015. URL: <http://lib.urfu.ru/mod/data/view.php?id=51&rid=242648&filter=1> (дата обращения: 01.11.2015).

Словарь-тезаурус синонимов русской речи / Под ред. Л.Г. Бабенко. – М.: АСТ-ПРЕСС КНИГА, 2007. – 512 с.

Ягунова Е.В., Пивоварова Л.М. Извлечение и классификация коллокаций на материале научных текстов: предварительные наблюдения // Мат-лы V междунар. конф. «Прикладная лингвистика в науке и образовании». – СПб., 2010а. URL: http://www.webground.su/services.php?param=get_class_collac&part=get_class_collac.htm (дата обращения: 01.11.2015).

Ягунова Е.В., Пивоварова Л.М. Природа коллокаций в русском языке. Опыт автоматического извлечения и классификации на материале новостных текстов // Сб. НТИ. Сер. 2, № 6. – М., 2010б. URL: http://www.webground.su/services.php?param=priroda_collac&part=priroda_collac.htm (дата обращения: 01.11.2015).

©Мухин М.Ю., 2015