

УДК81'23

ОЦЕНКА ИДИОМАТИЧНОСТИ РУССКИХ БИНОМИАЛОВ С ПОМОЩЬЮ МЕТОДОВ КОРПУСНОЙ И ЭКСПЕРИМЕНТАЛЬНОЙ ЛИНГВИСТИКИ

Мария Максимовна Кочнева

магистрант Института когнитивных исследований

Санкт-Петербургский государственный университет

199034, Санкт-Петербург, Университетская наб., 7/9. st110207@student.spbu.ru

Татьяна Евгеньевна Петрова

к. филол. н., доцент кафедры теории и методики

преподавания искусств и гуманитарных наук

Санкт-Петербургский государственный университет

199034, Санкт-Петербург, Университетская наб., 7/9. t.e.petrova@spbu.ru

Статья посвящена оценке идиоматичности биномиалов как одного из видов коллокаций в русском языке. С помощью методов корпусной лингвистики получены данные о частотности, устойчивости, коллокативной силе высокочастотных и низкочастотных выражений. Применены статистические методы для определения показателей мер MI и log-likelihood. Сделан вывод о влиянии частотности коллокации, частотности входящих в ее состав слов и степени неразложимости компонентов на процесс идиоматизации биномиала. Психолингвистический эксперимент подтвердил предположение о том, что высокочастотные выражения обрабатываются носителями языка целиком.

Ключевые слова: биномиалы; идиоматизация; идиоматичность; ментальный лексикон; статистические методы.

1. Введение

Изучение биномиалов – координированных особым образом пар слов, относящихся к одной части речи, расположенных на одном уровне синтаксической иерархии и связанных сочинительным союзом [Malkiel, 1959], представляет собой перспективное направление современной экспериментальной лингвистики. На протяжении многих лет эта многочисленная, структурно целостная, но демонстрирующая разные уровни семантической спаянности группа словосочетаний служила предметом различных лингвистических исследований. Однако долгое время биномиалы не выделяли в отдельный класс. Их рассматривали в рамках других лингвистических категорий: парные слова, аллитеративы, параллелизмы, гендиадис, парные сочетания (*zwillingsformeln*), тавтологичные пары слов (*die tautologischen wortpaare*) и т. п. Разнообразие терминов указывает на то, что в фокусе внимания оказываются некоторые особенности биномиалов, их отдельные признаки, но не весь класс. Это порождало путаницу, поскольку один и тот же биномиал мог рассматриваться, например, и в группе аллитеративов, и в группе парных слов.

1.1. К вопросу об определении биномиала

Я. Малкиэль понимал необходимость избавиться от терминологической путаницы и первым выделил биномиалы в отдельный класс в статье “*Studies in Irreversible Binomials*” («Исследование необратимых биномиалов») [Malkiel 1959], отмечая, что эти парные выражения обладают общей структурой независимо от степени их фиксированности, идиоматичности, наличия аллитерации и т. п. Я. Малкиэль обращает внимание на то, что способ образования биномиалов универсален для разных языков: английского, немецкого, французского, испанского, итальянского, латинского, русского и польского [там же].

Однако определение, данное Я. Малкиэлем, хотя и является эталонным для современных зарубежных исследований, в отечественной лингвистике не получило широкого развития, несмотря на то что последние годы активно изучаются парные выражения – «замкнутые соединения словоформ, в основе которых лежит принцип “два в одном”» [Плотникова 2012: 183], среди которых можно найти биномиалы: *небо и земля, радость и веселье, отец и мать, день и ночь* и т. д.

А.С. Буб (одна из немногих отечественных лингвистов, кто использует понятие «биномиал» в том же значении, что и Я. Малкиэль) исследует процесс когнитивной обработки биномиалов носителями русского языка и билингвами [Буб 2019]. Она выделяет во всем классе биномиалов ядро, периферию и псевдобиномиалы – «сочетания слов, совпадающие по структуре с биномиалами, но не обладающие значительным потенциалом взаимной встречаемости» [там же: 66]. Биномиалы рассматриваются как разновидность коллокаций, которым характерно свойство воспроизводимости, что не позволяет автору «согласиться с утверждением о том, что биномиалы представляют собой окказиональные единицы» [там же: 55].

В данной статье вслед за другими учеными [Lambrecht 1984; Mollin 2014] мы рассматриваем все множество биномиалов как непрерывный континуум: от окказиональных единиц (*haph legomena*), которые относятся к индивидуальному языку человека – идиолекту, до идиом, которые хранятся в памяти носителей языка и воспроизводятся в готовом виде. Псевдобиномиалы (по классификации А.С. Буб) рассматриваются нами как один из типов биномиальных конструкций, в которых слабо выражена коллокативная связь.

К настоящему времени доказано, что в английском языке до 85% всех биномиалов составляют *haph legomena*, около 15% приходятся на коллокации и только 1–2% на идиомы [Mollin 2014]. Существующие исследования биномиалов в русском языке подтверждают наличие в классе биномиалов выражений разной степени устойчивости [Кочнева 2021].

Биномиалы представляют собой особую лингвистическую единицу, на примере которой можно изучать и получить комплексное представление о процессе идиоматизации в языке.

1.2. Процесс идиоматизации в языке

Одним из актуальных вопросов современной когнитивной науки является изучение процесса идиоматизации отдельных языковых единиц [Буб 2019; Слюсарь и др. 2017]. В определении идиоматичности ряд исследователей придерживается фразеологического подхода: сочетание считается идиоматичным, если его общее значение не равно сумме значений отдельных частей [Мельчук 1960].

В последнее время происходит расширение взглядов на определение идиоматичности, которая по мнению ряда ученых (см. например: [Молчкова 2012; Пастухова: электр. ресурс; Warren 2005 и др.]) берет начало в способности человека к запоминанию повторяющихся событий. Идиоматичность – это умение различать ситуации и явления, которые предполагают знание и ис-

пользование стандартных выражений, хотя возможны альтернативные варианты [Warren 2005]. Так, в ситуации телефонного разговора можно спрогнозировать, какие комбинации слов будут использованы, например: *добрый день, повторите, пожалуйста, плохая связь, я вас слушаю* и т. п. Идиоматичность здесь проявляется в знании семантических (комбинаторных) возможностей слов образовывать фразы.

Такое представление выходит за рамки фразеологического подхода и приводит к изучению идиоматизации как процесса, а значит, материалом исследования могут стать не только фразы с доказанной идиоматичностью, но и аддитивные словосочетания.

Л.В. Молчкова рассматривает идиоматизацию как процесс уплотнения, сжатия информации, который начинается с выражения индивидуальных смыслов, которые, переходя в категорию широкодоступных, становятся устойчивыми единицами. Эти переходы сопровождаются переосмыслением исходного выражения, что приводит к приращению смысла в плане содержания и лаконизации средств плана выражения. Таким образом, идиоматичность – это отображение процесса идиоматизации [Молчкова 2012].

Изучение идиоматичности И.А. Мельчук связывал с определением ее степени, понимая под этим количественные измерения частот [Мельчук 1960]. Количественные меры необходимы при установлении корреляции между устойчивостью словосочетания и идиоматичностью, а также сочетаемостью слов в паре и идиоматичностью. Чтобы получить такие данные, требуется проанализировать большое количество образцов. Корпусный подход является оптимальным методом для решения поставленной задачи.

К фундаментальным понятиям корпусной лингвистики относится понятие коллокация. В 1957 г. Дж.Р. Фёрс дал этому термину определение, которое легло в основу отдельного направления лингвистики. Под коллокациями он понимал характерные, часто повторяющиеся комбинации слов. Отправной точкой дальнейших исследований послужило изучение результатов совместной встречаемости слов. Дж. Р. Фёрс утверждал: “You shall know a word by the company it keeps” [Firth 1957: 179] («Вы узнаете слово по его окружению»; перевод наш – М.К., Т.П.). Его последователи (например, Дж. Синклер) нашли применение термину коллокация в корпусной и компьютерной лингвистике, что объясняет его дальнейшее широкое распространение.

В ранних классификациях отечественных лингвистов термин коллокация не встречается, но угадывается по описанию признаков, отделя-

ющих эти выражения от других фразеологизмов. Так, В.В. Виноградов выделяет фразеологические единства – разновидность устойчивых выражений, в которых значения компонентов могут быть как равноправны, так и фразеологически связаны. «В живом употреблении они используются как готовые фразеологические единицы – воспроизводимые, а не вновь организуемые в процессе речи» [Виноградов 1977: 161]. От других фразеологических единиц единства отличаются тем, что могут допускать замену одного из компонентов выражения, например: *глубокая осень, глубока яма, глубокие чувства* и т. п.

В данной статье мы придерживаемся взгляда М.В. Хохловой, которая определяет коллокацию как «комбинацию двух или более слов, имеющих тенденцию к совместной встречаемости» [Хохлова 2008: 343]. М.В. Хохлова предлагает рассматривать термин «коллокация» как «родовое понятие для обозначения определенных типов устойчивых словосочетаний, фразеологических единиц и фразем» [там же: 345]. По нашему мнению, биномиалы являются одним из типов коллокативных выражений.

1.3. Экспериментальные методы исследования идиоматизации

Во многих исследованиях идиоматизации в языке изучаются факторы, влияющие на этот процесс.

Ряд авторов выделяет частотность выражения как одну из главных причин закрепления выражения в языке и предполагает, что высокочастотные сочетания хранятся в ментальном лексиконе целиком. Например, А. Сиянова-Чантурия и ее соавторы проводят исследование английских биномиалов с использованием методики регистрации движения глаз [Syanova-Chanturia, Conklin, van Heuven 2011]. Независимо от владения языком (носители vs. неносители языка) частотность фразы значительно влияет на окуломоторные характеристики. Результаты этого исследования выявили влияние частотности всего биномиала на время первого прочтения ($pMCMC = 0,0004$, $pr(>|t) = 0,0001$), общее время чтения ($pMCMC = 0,0001$, $pr(>|t) = 0,0000$) и количество фиксаций ($pMCMC = 0,0001$, $pr(>|t) = 0,0000$), при этом частотность слов, входящих в состав биномиала, не стала значимым предиктором скорости чтения.

Значимые результаты были также обнаружены также для порядка слов в биномиале. Время первого прочтения и общее время чтения были меньше для биномиалов с прямым порядком слов (т. е. предпочтительным, наиболее часто встречающимся в корпусе), по сравнению с биномиалами с обратным порядком (т. е. нарушен-

ным предпочтительным следованием слов, меной местами компонентов биномиала). А. Сиянова-Чантурия объясняет разницу в обработке прямого и обратного порядка частотностью всей фразы, а значит, репрезентацией более частотного варианта в ментальном лексиконе [там же]. Однако существуют работы, связывающие этот эффект с разницей в предсказуемости отдельных компонентов выражения [Reichle 1998; Engbert 2005].

Группа итальянских ученых в 2012 г. представила работу, посвященную обработке необратимых (т. е. с закрепленным порядком слов) биномиалов у людей с дислексией [Arcara et al. 2012]. Полученные результаты говорят в пользу того, что люди с дислексией лучше читают выражения, когда их компоненты представлены в привычном, закрепленном в языке порядке. Авторы рассматривают это как доказательство того факта, что необратимые биномиалы хранятся и извлекаются из ментального лексикона целиком.

На материале русского языка комплексного исследования идиоматизации биномиалов с использованием методов корпусной и экспериментальной лингвистики не проводилось. Исследование А.С. Буб, посвященное обработке биномиалов носителями русского языка и билингвами, – единственная известная нам на данный момент работа, описывающая процесс идиоматизации [Буб 2019]. Автор приходит к выводу о влиянии порядка слов и частотности всего выражения на время его прочтения и предполагает, что «информация о степени устойчивости биномиала может храниться в ментальном лексиконе носителей языка» [там же: 156].

2. Изучение идиоматизации русских биномиалов

Частотность является объективным показателем и влияет на закрепление биномиалов в языке, однако не хватает корпусных и эмпирических исследований того, на что именно опираются носители языка при создании и воспроизведении биномиалов. Цель настоящего исследования – оценить идиоматичность биномиалов русского языка разной степени частотности с помощью методов корпусной и экспериментальной лингвистики. Для решения поставленной задачи было проведено два исследования.

2.1. Эксперимент-1: корпусное исследование

Исследовательский вопрос заключался в проверке влияния частоты встречаемости биномиала в корпусе на коллокативную силу входящих в биномиал слов, т. е. степень спаянности, неразложимости его компонентов.

Материал исследования. Для анализа была выбрана структура биномиала «А и Б» (где А, В – существительные, «и» – сочинительный союз).

Всего было изучено 147 выражений. Мы анализировали пары словоформ, а не леммы. Необходимость такого решения была продиктована тем, что «различные формы слов одной и той же лексемы приводят к различным значениям частоты встречаемости пары» [Mollin 2014].

Поиск биномиалов осуществлялся по подкорпусу триграмм Национального корпуса русского языка (НКРЯ – <https://ruscorpora.ru>, дата обращения: 10.02.2022). В строке поиска были указаны следующие параметры: первое слово (N) + второе слово (союз «и») + третье слово (N). Основываясь на данных ipm, подкорпус сформировал список 100 наиболее часто встречающихся в русском языке биномиалов. Из анализа были исключены 16 выражений; к ним относятся случаи неснятой омонимии, например, наречие *чаще* и прилагательное *малого и среднего* и т.п. В результате для исследования было отобрано 84 выражения.

Используемый подкорпус триграмм не дает возможность получить список самых низкочастотных выражений. Было принято решение обратиться к подкорпусу текстов Всемирной организации здравоохранения, созданного в ходе исследования «Биномиалы в контексте пандемического дискурса» в 2021 г. [Кочнева 2021]. Выражения для данной работы были выбраны из текстов о пандемии COVID-19, опубликованных в период с февраля по октябрь 2020 г. в разделе «Новости». Отбор осуществлялся по вышеперечисленным параметрам. Из списка 132 биномиалов для данного исследования были отобраны 63 самых низкочастотных выражения.

Список 147 биномиалов можно найти по ссылке: <https://docs.google.com/spreadsheets/d/15t2s9kDPulmJuwPwIX79PTI4wd0JWZSIvSPATGfTVeE/edit?usp=sharing>.

Методы исследования. Количественные характеристики являются универсальным критерием для анализа всего класса биномиалов. Они позволяют выйти за рамки фразеологического подхода и дают новые возможности для одновременного исследования всех этапов процесса идиоматизации. По мнению М.В. Хохловой, изучению взаимосвязи между сочетаемостью и устойчивостью выражения «может способствовать анализ коллокаций, полученных из корпусов на основании статистического метода» [Хохлова 2008: 343].

В лингвистике известно большое количество тестов, которые направлены на количественную оценку того, насколько сильно слова в паре притягиваются друг к другу [Evert 2007]. Стандартом в корпусных исследованиях стало использование мер MI, t-score, log-likelihood – все они «служат

для вычисления степени близости между компонентами в корпусе» [Хохлова 2008: 345].

В данной работе использована мера взаимной информации (Mutual Information):

$$MI = \log_2 \frac{O}{E}$$

Значения MI соответствуют битам:

- 0 бит соответствует $O = E$ (O – наблюдаемая частота пары слов в корпусе; E – ожидаемая частота пары слов) и означает, что слова встречаются так часто, как и ожидалось;
- 1 бит соответствует $O = 2E$ и означает, что частота встречи в два раза превышает ожидания;
- 2 бита соответствует $O = 4E$;
- 3 бита соответствуют $O = 8E$;
- 10 бит соответствуют примерно $O = 1000E$ и т.д. в геометрической прогрессии. Отрицательное значение MI указывает на то, что пара слов встречается реже, чем ожидалось.

Для получения информации о частоте совместного вхождения слов в биномиале вычислялась мера логарифмического правдоподобия (log-likelihood)

$$\text{log-likelihood} = 2 \sum_{ij} O_{ij} \times \log \frac{O_{ij}}{E_{ij}}$$

В ходе анализа было проведено корпусное исследование, которое позволило составить рейтинг частотности отобранных биномиалов. По частоте вхождений в НКРЯ все выражения были разделены на три группы:

- высокочастотные – 89 биномиалов – более 50 вхождений;
- среднечастотные – 30 биномиалов – от трех до 50 вхождений;
- низкочастотные – 28 биномиалов – менее трех вхождений.

Отдельно была определена частотность слов, входящих в состав биномиалов.

Результаты исследования. Два биномиала получили отрицательные значения MI – *правительства и компании* (-0,225) и *люди и организации* (-0,95). Это означает, что встреча двух слов в паре произошла вопреки ожиданию. Остальные 145 значений MI являются положительными. Поскольку наблюдаемая в корпусе частота биномиала превысила ожидаемую, то можно говорить о неслучайности встречи слов в паре.

Рейтинг биномиалов по MI существенно отличается от рейтинга частотности: биномиалы, состоящие из высокочастотных слов-компонентов, оказались в конце рейтинга по показателям MI (см. Табл. 1). Анализируемые компоненты находятся между собой в прямой зависимости и чем выше показатели частотности слов, тем меньше показатель MI. Например, биномиал *день и ночь*

**Кочнева М.М., Петрова Т.Е. ОЦЕНКА ИДИОМАТИЧНОСТИ РУССКИХ БИНОМИАЛОВ
С ПОМОЩЬЮ МЕТОДОВ КОРПУСНОЙ И ЭКСПЕРИМЕНТАЛЬНОЙ ЛИНГВИСТИКИ**

(с частотой слов *день* – 183 047 вхождений в НКРЯ; *ночь* – 56 938 вхождений в НКРЯ) занимает 113 позицию в рейтинге по МІ, против первой

позиции в рейтинге частотности. Это указывает на то, что встреча двух очень частотных слов ожидаемое, но не уникальное событие.

Таблица 1

Биномиалы с высокими показателями частотности слов, входящих в их состав: фрагмент

№	Биномиал	МІ	Частотность в НКРЯ	Рейтинг МІ	Рейтинг частотности
1	<i>день и ночь</i>	5,88	2166	113	1
2	<i>руки и ноги</i>	6,19	1452	108	2
3	<i>наука и жизнь</i>	6,89	950	93	7
4	<i>отец и мать</i>	5,78	915	114	8
5	<i>война и мир</i>	5,43	520	121	23

Анализ полученных данных показал, что мера взаимной информации МІ присваивает высокие показатели низкочастотным парам. Например, *Билла и Мелинды* (1 вхождение в НКРЯ) занимает 4 место в рейтинге МІ; *этиология и патогенез*

(2 вхождения в НКРЯ) – 3 место в рейтинге МІ (см. Табл. 2). В таких случаях мера МІ определяет «размер эффекта», т. е. подчеркивает уникальность или специфичность сочетания [Evert 2007: 28].

Таблица 2

Биномиалы с самыми высокими показателями МІ: фрагмент

№	Биномиал	МІ	Частотность в НКРЯ	Рейтинг МІ	Рейтинг частотности
1	<i>налогам и сборам</i>	18,16	290	1	78
2	<i>эскизы и кроки</i>	17,74	342	2	61
3	<i>этиология и патогенез</i>	17,48	2	3	130
4	<i>Билла и Мелинды</i>	17,46	1	4	131
5	<i>образу и подобию</i>	15,97	484	5	33

Таким образом, можно сделать вывод о том, что мера МІ позволяет провести различие между биномиалами, ставшими в русском языке общеупотребительными, и биномиалами, обладающими специфическими особенностями.

Вычисление меры логарифмического правдоподобия log-likelihood позволило сравнить частоты

совместного вхождения отдельных слов каждого биномиала в НКРЯ с рейтингом частотности этого биномиала и его рейтингом по показателям меры МІ. Анализ полученных данных показал, что рейтинги меры log-likelihood оказались близки к рейтингам частотности биномиалов и контрастировали с рейтингами меры МІ (см. Табл. 3).

Таблица 3

Частотные характеристики биномиалов: фрагмент

№	Биномиал	Частотность в НКРЯ	МІ	log-likelihood	Рейтинг частотности	Рейтинг МІ	Рейтинг log
1	<i>день и ночь</i>	2166	5,9	12771,8	1	113	1
2	<i>руки и ноги</i>	1452	6,2	9140,9	2	108	2
3	<i>днем и ночью</i>	1338	8,8	9103,9	3	59	3
4	<i>мужчины и женщины</i>	957	8,9	6697,6	6	57	4
5	<i>добра и зла</i>	1003	11,8	6630,3	5	22	5
6	<i>мужчин и женщин</i>	868	9,8	5779,6	9	45	6
7	<i>науки и техники</i>	1021	10,1	5731,3	4	42	7
8	<i>муж и жена</i>	817	7,6	5027,8	10	76	8
9	<i>руками и ногами</i>	697	7,8	5012,8	13	72	9
10	<i>наука и жизнь</i>	950	6,9	4992,7	7	93	11
11	<i>отец и мать</i>	915	5,8	4967,6	8	114	12
12	<i>дни и ночи</i>	669	7,1	4424,6	15	86	13
13	<i>рабочих и крестьян</i>	732	9,3	4277,5	12	52	14

Дополнительно был проведен анализ коллокативной силы биномиалов, т. е. степень спаянности, неразложимости их компонентов. Сильным мы считаем коллокат (второе слово биномиала),

частота которого составляет более 50% от общего количества всех коллокатов в конструкции «слово1+и+коллокат». Например, слово *день* является в русском языке высокочастотным и общеупотреб-

бительным. Известно, что по формуле «день + и + ...» оно образовало с разными словами группу биномиалов, общее количество вхождений которых составило 3 248 (т. е. 100%). В это число вхо-

дят все случаи сочетания слова *день* со словом *ночь* – 2 166 вхождений, что составляет 66% от всех вхождений. Частотность биномиала является предиктором коллокативной силы (см. Табл. 4).

Таблица 4

Коллокативная сила биномиалов: фрагмент

№	Биномиал	Коллокативная сила, %	Частотность в НКРЯ	Частотность слова 1 со всеми коллокатами в НКРЯ
1	<i>образу и подобию</i>	96	484	503
2	<i>налогам и сборам</i>	95	290	304
3	<i>мужчины и женщины</i>	93	957	1033
4	<i>эскизы и кроки</i>	93	342	366
5	<i>пух и прах</i>	91	310	339
6	<i>мальчики и девочки</i>	90	351	389
7	<i>днем и ночью</i>	89	1338	1497
8	<i>министерств и ведомств</i>	88	334	379
9	<i>мужчин и женщин</i>	87	868	995
10	<i>изменений и дополнений</i>	87	388	446
11	<i>оперы и балета</i>	87	278	321
12	<i>братья и сестры</i>	86	511	594
13	<i>фабрик и заводов</i>	86	319	371
14	<i>мужчина и женщина</i>	85	447	524

2.2. Эксперимент-2: психолингвистическое исследование

Исходная гипотеза второго эксперимента состояла в том, что частотность биномиала влияет на его разложимость и обработку.

Материал исследования. Стимулы для исследования были выбраны из списка биномиалов, анализируемого в Эксперименте-1. Всего было отобрано 40 биномиалов: 20 высокочастотных – более 500 вхождений в корпус (*день и ночь* – 2166; *братья и сестры* – 511), 20 низкочастотных – менее 30 вхождений (*прав и возможностей* – 24; *артисты и исполнители* – 0).

Респонденты. В эксперименте участвовал 71 человек в возрасте от 18 до 24 лет, все респонденты являлись носителями русского языка.

Процедура эксперимента. Каждому респонденту предъявлялась первая часть каждого из 40 биномиалов (по форме «слово1+и+...»), которую он должен был продолжить одним словом. Стимулы предъявлялись в случайном порядке. В результате было получено 2 840 реакций.

Результаты исследования. Выделено несколько типов реакций:

- 1) совпадение с исходным стимулом-биномиалом;
- 2) отказ от ответа;
- 3) реакция в форме словосочетания;
- 4) новый вариант (синонимы, квазисинонимы, дополнения).

Среди высокочастотных биномиалов все кроме одного (*сил и средств*) были восстановлены

респондентами. У девяти стимулов 60–94% реакций совпадало с исходными биномиалами, у семи стимулов – 20–40% реакций совпадало с исходными биномиалами, у трех стимулов – меньше 10%.

Среди низкочастотных выражений у 13 стимулов количество ответов, совпавших с исходным биномиалом, не превышает 35%. У семи стимулов совпадений не было совсем.

Сравнив проценты правильного восстановления биномиалов обеих групп с помощью *t*-теста Стьюдента, мы подтвердили, что различия в восстановлении низко- и высокочастотных стимулов статистически значимы ($p < 0,01$).

Для высокочастотных стимулов респонденты предложили 332 новых варианта второго слова. Из них 152 случая зафиксировано в НКРЯ как коллокации. Для низкочастотных стимулов было предложено 552 варианта – 150 являются коллокациями. Анализ количества новых вариантов в каждой из групп показал, что вариативность ответов при обработке низкочастотных биномиалов значимо выше, чем при восстановлении высокочастотных ($p < 0,0003$).

Таким образом, результаты исследования показывают, что носители языка по-разному обрабатывают высокочастотные и низкочастотные биномиалы. Высокочастотные конструкции восстанавливаются лучше, что является дополнительным доказательством их целостного хранения в ментальном лексиконе.

3. Выводы

Полученные с помощью методов корпусной лингвистики данные (Эксперимент-1) и результаты психолингвистического исследования (Эксперимент-2) позволяют сделать вывод о том, что на процесс идиоматизации оказывают влияние как частотность всей коллокации, так и частотность слов, входящих в состав биномиала. Высокие показатели частотности биномиала, частотность его слов, высокая коллокативная сила пары являются статистическими признаками идиоматичности и говорят в пользу того, что данная пара уже включена в процесс идиоматизации. Именно частотность биномиала является главным критерием, определяющим начало развития идиоматичности.

Результаты корпусного исследования русских биномиалов совпадают с результатами эксперимента, направленного на проверку того, как носители русского языка восстанавливают высокочастотные и низкочастотные биномиалы, а также подтверждают полученные рядом авторов (см., например [Слюсарь и др. 2017]) выводы о том, что высокочастотные сочетания хранятся в ментальном лексиконе целиком.

Общие результаты исследования согласуются с положением Дж. Синклера о двух типах языковой обработки: принципом открытого выбора (языковые составляющие текста – это результат большого количества сложных выборов) и принципом идиоматичности (целостное хранение фрагментов языка) [Sinclair 1991].

Примечание

¹ Исследование выполнено при поддержке Санкт-Петербургского государственного университета, проект ID 94034584 «Механизмы чтения и интерпретации текста на родном и неродном языках: междисциплинарное экспериментальное исследование с использованием методов регистрации движения глаз, визуальной аналитики и технологий виртуальной реальности».

Список литературы

Буб А.С. Когнитивная обработка коллокаций-биномиалов русского языка (экспериментальное исследование) // Вестник Томского государственного университета. 2019. № 442. С. 5–13.

Виноградов В.В. Об основных типах фразеологических единиц в русском языке // Виноградов В.В. Избранные труды: Лексикология и лексикография. М.: Наука, 1977. С. 140–161.

Кочнева М.М. Биномиалы в контексте пандемического дискурса // Мир и пандемии: трансформации, коммуникации, стратегии: матер. Всеросс. науч. конф. студентов-стипендиатов

Оxfordского Российского фонда. Екатеринбург: Изд-во Урал. ун-та, 2021. С. 136–139.

Мельчук И.А. О терминах «устойчивость» и «идиоматичность» // Вопросы языкознания. 1960. № 4. С. 72–80.

Молчкова Л.В. Идиоматизация как способ компрессии информации // Вестник Волгоградского государственного университета. Языкознание. 2012. № 2(16). С. 124–128.

Пастухова Е.А. Идиоматичность как особая форма подачи информации в речи (на примере конструкции пути). [Электронный ресурс]. URL: https://upload.pgu.ru/iblock/95d/statya-_13_.pdf (дата обращения: 25.02.2022)

Плотникова А.М. Парные сочетания в современном русском языке и их креативный потенциал // Уральский филологический вестник. 2012. № 2. С. 183–187.

Слюсарь Н.А. и др. Экспериментальные исследования грамматики: словосочетания с буквальным и небуквальным значением / Н.А. Слюсарь, Т.Е. Петрова, Е.В. Михайловская, Н.В. Череповская, В.К. Прокопья, Д.А. Чернова, Т.В. Черниговская // Вопросы языкознания. 2017. № 3. С. 83–98.

Хохлова М.В. Экспериментальная проверка методов выделения коллокаций // Инструментарий русистики: корпусные подходы / ред. А. Мустайоки, М.В. Копотев, Л.А. Бирюлин, Е.Ю. Протасова. Helsinki: Helsinki University Press, 2008. С. 343–357.

Arcara G. et al. Is 'hit and run' a single word? The Processing of Irreversible Binomials in Neglect Dyslexia / G. Arcara, G. Lacaíta, E. Mattaloni, L. Passarini, S. Mondini, P. Benincà, C. Semenza // Frontiers in Psychology. 2012. № 3(11). Pp. 1–11.

Engbert R. et al. A Dynamical Model of Saccade Generation During Reading/ R. Engbert, A. Nuthmann, E.M. Richter, R. Kliegl // Psychological Review. 2005. Vol. 112(4). Pp. 777–813.

Evert S. Corpora and Collocations: Extended Manuscript/ Institute of Cognitive Science, University of Osnabrück, 2007. 53 p.

Firth J.R. A Synopsis of Linguistic Theory 1930–1955 // Special Volume of the Philological Society. Oxford: Oxford University Press, 1957. Pp. 1–32.

Lambrecht K. Formulaicity, Frame Semantics and Pragmatics in German Binomial Expressions // Language. 1984. Vol. 60(4). Pp. 753–796.

Malkiel Y. Studies in Irreversible Binomials // Lingua. 1959. Vol. 8. Pp. 113–160.

Mollin S. The (Ir)reversibility of English Binomials: Corpus, Constraints, Developments. Philadelphia: John Benjamins Publishing Company, 2014. 254 p.

Reichle E.D. et al. Toward a Model of Eye Movement Control in Reading / E.D. Reichle, A. Pollatsek, D.L. Fisher, K. Rayner // *Psychological Review*. 1998. Vol. 105(1). Pp. 125–157.

Sinclair J. Corpus, Concordance and Collocation. Oxford: Oxford University Press. 1991. 200 p.

Sivanova-Chanturia A., Conklin K., van Heuven W.J.B. Seeing a Phrase “Time and Again” Matters: the Role of Phrasal Frequency in the Processing

of Multiword Sequences // *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2011. Vol. 37(3). Pp. 776–784.

Warren B. A model of idiomaticity // *Nordic Journal of English Studies*. 2005. Vol. 4(1). Pp. 35–52. [Электронный ресурс]. URL: https://www.researchgate.net/publication/262842806_A_Model_of_Idiomaticity (дата обращения: 09.03.2022).

ASSESSMENT OF IDIOMATICITY OF RUSSIAN BINOMIALS USING METHODS OF CORPUS AND EXPERIMENTAL LINGUISTICS

Maria M. Kochneva

Master Student, Institute of Cognitive Studies
Saint Petersburg State University

Tatiana E. Petrova

Associate Professor, Department of Theory and Methodology of Teaching Arts and Humanities
Saint Petersburg State University

This paper is devoted to the evaluation of idiomaticity of binomials as one of the types of collocations in the Russian language. Using the methods of corpus linguistics, data on the frequency, fixity and collocative power of high-frequency and low-frequency expressions were obtained. Statistical methods were used to determine the MI and log-likelihood measures. We revealed that the process of binomial idiomaticity is influenced by the frequency of the whole collocation, the frequency of the words included in it, and the degree of decomposability of the components. A psycholinguistic experiment confirmed the assumption that high-frequency binomials are processed by native speakers as coherent syntactic units.

Key words: binomials; ideomatisation; idiomaticity; mental lexicon; statistical measures.