

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

II МЕЖДУНАРОДНАЯ  
НАУЧНАЯ ФИЛОЛОГИЧЕСКАЯ  
КОНФЕРЕНЦИЯ ИМЕНИ  
ЛЮДМИЛЫ АЛЕКСЕЕВНЫ  
ВЕРБИЦКОЙ

19–26 марта 2024, Санкт-Петербург

*Сборник тезисов*

САНКТ-ПЕТЕРБУРГ, 2024

© Санкт-Петербургский государственный университет, 2024

Ольховская Александра Игоревна<sup>1</sup>, Зельникова А. А.<sup>2</sup>

<sup>1</sup>канд. филол. наук, заведующий лабораторией филологических исследований,

Государственный институт русского языка им. А. С. Пушкина

<sup>2</sup>магистрант, Государственный институт русского языка им. А. С. Пушкина

## К ВОПРОСУ ОБ УСОВЕРШЕНСТВОВАНИИ МЕТОДИКИ КОРПУСНОГО ОТБОРА НОВОЙ ФРАЗЕОЛОГИИ

*Ключевые слова:* новая фразеология, коллокация, мера устойчивости, корпусная лингвистика, корпус текстов.

В [Зельникова, Ольховская 2023] была предложена методика, позволяющая производить эффективный поиск и верификацию новой фразеологии с опорой на корпусные данные. В её основе лежит использование т.н. маркеров новизны — речевых оборотов, с помощью которых говорящий маркирует новую, с его точки зрения, языковую единицу. В ходе обсуждения и апробации методики были намечены пути её совершенствования. Цель настоящей заметки состоит в описании способов повышения валидности корпусной методики отбора современных фразеологизмов.

Прежде всего, было увеличено число маркеров новизны с 15 до 56. Помимо оборотов, указывающих на современность и актуальность выражения (как сейчас говорят, как в нынешние дни выражаются, модное выражение и т.п.), были учтены маркеры, акцентирующие среду его распространения (как в сети говорят, как в интернете выражаются, как говорит молодёжь). Учёт новых маркеров позволил дополнить список фразеологических единиц, в том числе выражениями, характерными для речи молодёжи (чики-пуки, респект и уважуха, взрыв мозга).

Во-вторых, был существенно расширен корпусный материал исследования. Кроме газетного корпуса НКРЯ, были использованы основной корпус НКРЯ, корпус «Социальные сети» НКРЯ, Генеральный интернет-корпус русского языка, RuTenTen. Итоговый список фразеологизмов получен путем наложения данных из разных источников и включает 383 единицы. Следует отметить, что новый список значительно превосходит предыдущий, в котором насчитывалось лишь 241 словосочетание.

В-третьих, была разработана более четкая система фильтрации найденного материала, включающая три ступени. На первом этапе мы обратились к использованию хронологического фильтра. Хронологические рамки современной фразеологии были несколько изменены: если раньше мы причисляли к современным обороты, которые начали употребляться в языке в 80-е гг., то теперь — такие, которые входят в обиход начиная с 90-х гг. Представляется, что скорректированные рамки в большей степени соответствуют понятию «современная фразеология». На этом этапе из списка было исключено около 150 фразеологизмов, среди них: держать руку на пульсе (1929), утечка мозгов (1969), американская улыбка (1977), без галстуков (1987), работать на дядю (1969) и мн. др.

На втором этапе фильтрации был привлечен такой важный для фразеологии параметр, как идиоматичность оборота. Под идиоматичностью мы понимаем принципиальную «невыводимость» значения целого выражения из значений слов, входящих в его состав. Здесь из списка было изъято около 70 единиц, например, информационное поле, вести бизнес, медийная персона, альтернативно одаренный, утрата доверия, дешевые понты, гламурная киса и др.

Наконец, на последнем этапе были отсеяны все обороты, состоящие более чем из двух компонентов (всего около 60 единиц). Это обусловлено сугубо формальными ограничениями тематических инструментов измерения силы коллокаций и, в частности, тем, что имеющиеся в корпусной лингвистике формулы вычисления мер устойчивости предназначены лишь для оценки двухкомпонентных словосочетаний.

При этом следует оговориться, что там, где это было возможно, мы попытались свести трехкомпонентные единицы к двумкомпонентным. Например, в *шаговой доступности* → *шаго-*

*вая доступность, за чертой бедности* → *черта бедности, лицо кавказской национальности* → *кавказская национальность*.

Наконец, в-четвертых, была произведена попытка формализовать оценку силы словосочетаний (меры устойчивости). Проведя практическое сравнение различных мер устойчивости на конкретном фразеологическом материале, авторы обозначили приблизительные пороги значений, отделяющие фразеологизмы от коллокаций. «Можно предположить, что для причисления устойчивого оборота к фразеологизму его MI должна быть выше 5, MI3 — выше 25, t-score — выше 40, LogDice — выше 3» [Зельникова, Ольховская 2023а: 422].

Показатели мер устойчивости извлекались из корпусного менеджера Sketch Engine. Единица проходила отбор в том случае, если как минимум три из четырех показателей соответствовали установленным пороговым значениям. На основании данного параметра из списка было исключено 12 единиц, например, по феншую, карьерный лифт, взорвать интернет. Итоговый список составили 64 фразеологических выражения. К новым выражениям, найденным по скорректированной методике, можно отнести обороты *ниже плинтуса, офисный планктон, черный пиар, учить матчасть, сто пудов*.

По результатам исследования была существенно доработана методика корпусного отбора современной фразеологии и получен список новых фразеологизмов русского языка. Материалы могут быть использованы в практической лексикографии, а также в сфере преподавания русского языка, в том числе как иностранного.

## Литература

- Зельникова А. А., Ольховская А. И. Методика корпусного отбора современной фразеологии // Сборник тезисов LI Международной научной филологической конференции им. Л. А. Вербицкой. СПб., 2023. С. 624–625.
- Зельникова А. А., Ольховская А. И. Математические инструменты измерения коллокаций в аспекте изучения фразеологии // III Костомаровский форум. Славянская культура: истоки, традиции, взаимодействие. Материалы Международной научно-практической конференции. М., 2023а. С. 418–423.