

Йоханна Виймаранта

Финляндия, Университет г. Турку

# Устный корпус Национального Корпуса Русского Языка как источник информации о концептуализации

**Ключевые слова:** концептуализация, русский язык, НКРЯ, методология, когнитивная лингвистика

**Key words:** conceptualization, Russian, Russian National Corpus, methodology, cognitive linguistics

## Abstract

This methodological article deals with how we can use the Corpus of Spoken Russian that is part of the Russian National Corpus as a source of language data for studying conceptualization. It is claimed that spoken corpora are better suited to this purpose than corpora of written texts or other sources of language data. The article describes the benefits of material that consists of units of spoken discourse rather than written sentences, even when we study questions that are not primarily related to spoken language. The notion of conceptualization refers to how humans categorize and classify their perceptions on the basis of how their bodies are built and what they know about the functioning and characteristics of different objects. The main benefit of material from the spoken corpus for studying these phenomena is the spontaneity of such material – in speech we do not have the time to make deliberate choices, so that spoken language reflects our conceptual processes in their purest form. Also, spoken language is not corrected and neutralized by others, as is often the case with the language of published texts. As examples of questions that can be studied with spoken corpus material I introduce the Russian onomatopoeic word *bac* ‘bang’ and the actant structure of the verb form *podderzhivaet* ‘supports’.

## 1. Введение

В когнитивной лингвистике языковой материал служит доказательством того, как человек воспринимает мир и классифицирует свои наблюдения о нем. Уже само существование определенной конструкции, словоформы

или грамматической категории в определенном языке может быть принято как доказательство некоего классификационного процесса в сознании говорящих на данном языке людей. С появлением электронных корпусов найти аутентичный языковой материал стало легко как никогда раньше, и придуманные исследователем примеры больше не считаются идеальным вариантом материала. Тем не менее, даже большие по объему электронные корпуса оказываются разными в отношении того, как они репрезентируют несознательную языковую продукцию говорящих.

Существует мнение [напр., Evans 2009], что языковым материалом при изучении концептуальных процессов должен служить спонтанный языковой материал в виде высказываний (*utterance*). В. Эванс считает, что предложения являются искусственными теоретическими конструкциями, и поэтому они не отражают настоящее употребление языка. Где же в таком случае можно найти языковой материал, который состоит из высказываний, а не из предложений – ведь именно в таком виде представлены корпуса письменных текстов? Сам Эванс – известная фигура в когнитивной лингвистике – не определяет хороший источник такого материала. С нашей точки зрения, таким материалом является как раз корпус устной речи.

В статье в центре внимания находится устный корпус Национального Корпуса Русского Языка (НКРЯ). Утверждается, что корпус такого типа имеет преимущество не только при изучении особенностей устной речи и закономерностей дискурса, но и тогда, когда языковой материал, доступный в электронном корпусе, должен служить примером или доказательством когнитивных процессов говорящего. Обсуждаются также ограничения этого корпуса и такого корпусного подхода.

Пока что когнитивная лингвистика является скорее теоретической, чем эмпирической наукой, и позиция языкового (в том числе корпусного) материала в ней неясна. Однако разные эмпирические и даже количественные методы в ней становятся все более популярными [см. Janda 2013]. Вопросы, связанные с различными устными корпусами, обсуждаются в [Wichmann 2009], а корпусная русистика в [Мустайоки и др. 2008].

## 2. Традиция изучения устной русской речи и корпус устного языка НКРЯ

Традиция исследования устной русской речи во многом связана с традицией изучения разговорной речи [напр., Земская 1973; 1983; 1987; Земская и др. 1981; Лаптева 2008; Китайгородская, Розанова 2010]. Из этой традиции усвоена и система представления материала, принимаемая в устном корпусе

НКРЯ. Определение разновидностей устной речи остается за пределами нашей работы. Заметим только, что устный корпус НКРЯ содержит разные формы спонтанной и квазиспонтанной речи. Он включает в себя расшифровки (стенограммы) магнитофонных записей публичной и частной устной речи, а также транскрипты кинофильмов. Корпус содержит примерно 10 миллионов словоупотреблений. Даже если корпус охватывает относительно длинный отрезок времени, большая часть его материалов датируется 2000-ми гг. Поиск в корпусе можно ограничить с помощью составления пользовательских подкорпусов, а также лексическими, грамматическими и/или семантическими параметрами поиска. Специфическим для этого подкорпуса является возможность ограничивающего поиска социолингвистическими параметрами (пол, возраст и год рождения говорящего). Такие ограничения дают нам уникальные возможности изучения различий в употреблении русского языка между полами и поколениями. Однако весь материал в устном корпусе не содержит информацию по всем этим параметрам, что влияет на достоверность результатов. Состав и структура корпуса (в стадии 2008 г.) описываются более подробно в [Гришина, Савчук 2009].

Основной проблемой при употреблении устного корпуса НКРЯ является его размер. 10 миллионов слов оказываются достаточной базой для разрешения многочисленных вопросов, но малочастотные лексические единицы в корпусе такого размера представлены случайно – они отсутствуют или единичны. При этом, если поиск ограничивается по социолингвистическим параметрам, то выбранный подкорпус, естественно, оказывается еще более скромным. В определенном смысле употребление корпуса приближается в таком случае к употреблению информантов – только материал становится намного более доступным для исследователя. Мы сталкивались с этой проблемой неоднократно, убеждаясь, например, что при изучении малочастотных приставочных глаголов с *под-* от устного корпуса НКРЯ было мало пользы.

### 3. Что такое концептуализация

Концептуализация означает то, как человек классифицирует и категоризирует информацию на основе того жизненного опыта, который у него есть о свойствах разных объектов. Согласно когнитивной лингвистике не существует полных синонимов, и, подбирая конкретную синтаксическую форму или слово, говорящий всегда делает выбор, который предопределяется процессами концептуализации.

Приведу первый пример. Поиск со словом *бац* дает 98 вхождений в устном корпусе НКРЯ. Сошлемся на следующие употребления слова *бац* в этом корпусе:

(1) [№ 6, муж, 22] И опять / **бац** / бомбежка / там все опять рухнуло [Беседа с социологом на общественно-политические темы (Новосибирск) // Фонд «Общественное мнение», 2003];

(2) [Светлана Г., жен, 45] Это не какое-то там облако было / а вдруг **бац** и исчезло [Разговор о видении // Из коллекции НКРЯ, 2008];

(3) [Ольга Александровна, жен] Просто я вот слушала-слушала я так поняла **бац** / у нас есть проект [Дискуссия по педагогике // Из коллекции НКРЯ, 2005].

В этих примерах звукоподражание *бац* выражает то эмоциональное состояние, с которым говорящий воспринимает происходящее. Употребление слова *бац* свидетельствует о связи в концептуальной системе говорящего между действиями, которые могут либо чисто физически вызвать звук определенного типа или на визуальном плане быть похожи на действия с таким звуком, либо вызвать чувства, которые имеют сходство или связь другого типа с такими действиями. Эти связи можно описать терминами концептуальные *метафоры* и *метонимия*. В когнитивной лингвистике метонимией называется перенос внутри одной области знания, т. е. употребление связанного по значению понятия (слова) вместо того понятия (слова), которое употребляется. В случае метафоры эти понятия принадлежат разным областям [см., напр.: Lakoff, Johnson 1980; 1999; Barcelona 2000; Dirven, Pörings 2003]. Поскольку толкование дефиниций разных когнитивных областей затруднительно, существуют и другие определения метонимии и метафоры, в том числе такое, по которому в метонимии существует и сохраняется ассоциативная связь между понятиями, а в метафоре – нет<sup>1</sup>.

Значения слова *бац* как обозначения звука и действия, которое вызывается этим звуком, явно являются частями одной и той же когнитивной области, т. е. их каузативное отношение можно интерпретировать как метонимическое. В примере (1) имеет место как раз такой случай. В примере (2) исчезновение облака, разумеется, не связано ни с каким звуком, или, по крайней мере, не с таким звуком, который было бы принято называть словом *бац* по-русски. Тем не менее то, что облака вдруг нет или оно тает перед глазами наблюдателя связывается с событиями, сопровождающимися таким звуком. Общим для звука и исчезновения облака можно в таком случае считать визуальное представление о действии, вызывающем такой звук. Абстракция тех реально существующих звуков, которые по-русски называются *бац*, связывается метонимически с визуальной абстракцией того, как выглядят действия с таким звуком (т. е. взрывы). В этой метонимии название звука репрезентирует звук, в концептуализации участвует также то, как в нашем представлении выглядит такое действие, которое называется *бац*. Данная метонимия имеет

<sup>1</sup> Более подробно разные – существующие в пределах когнитивной лингвистики – мнения о метафоре и метонимии и различиях между ними обсуждаются в [Barnden 2010] и [Viimaranta 2012].

метафорическую связь с исчезновением облака, поскольку звук и облаку явно принадлежат разным областям знания.

В примере (3), в свою очередь, говорящая вдруг понимает что-то, и это осознание неожиданного действия похоже в ее восприятии на громкий звук от взрыва. С позиций концептуализации опять важную роль играет несуществующий звук. Однако в этом случае метонимическая связь между звуком и словом метафорически связывается с неожиданным появлением проекта, т. е. вторым концептом является здесь ощущение, вызванное звуком. Таким образом, здесь существует двойная метонимия в соединении с метафорой, в которой звук, его название и ощущение, связанное со звуком, являются понятиями одной когнитивной области, а другое действие, которое вызывает подобное ощущение, принадлежит иной когнитивной области.

Отмечаем, что назвать звук именно *бац*, т. е. дать звуку на фонетическом уровне такую реализацию, призвана именно концептуализация, свойственная конкретному языку. В других языках те же самые звуки, которые категоризируются в русском языке как *бац*, воспринимаются и называются по-разному, и обязательно их все можно описать одним и тем же словом.

Поверхностный анализ такого типа возможен был бы, несомненно, и без корпуса устного языка. Примеры употребления *бац* можно найти и в словарях и в других корпусах. Однако, возможно, нигде кроме устного корпуса нет такого уникального источника спонтанных употреблений этого слова. Они ведут нас прямо к концептуальным процессам.

Следует заметить, что предложенное выше описание некоторых употреблений слова *бац* с помощью понятий *метафора* и *метонимия* не является единственно возможным вариантом их интерпретации с помощью аппаратуры когнитивной лингвистики.

#### 4. Вопрос об аутентичности и спонтанности

Если вся языковая продукция может считаться отражением концептуализации, то в спонтанной устной речи эта связь прослеживается еще в более чистом виде. В потоке речи каждое словоупотребление отражает показательный вариант концептуализации. Даже если у говорящего для каждой выражаемой мысли существуют разные варианты (которые отражают разные виды концептуализации и выбор между которыми тоже рассказывает о концептуализации говорящего), то в потоке речи говорящий не может делать намеренный выбор между ними, времени для этого не хватает. Это свойство спонтанной устной речи, т. е. необходимость реагировать очень быстро, на подсознательном уровне, делает ее идеальным языковым материалом для изучения процессов

концептуализации. Более того, неподготовленная устная речь является подчеркнута авторской, она не скорректирована другими.

Многие электронные корпуса составлены из газетных текстов и/или художественной литературы. В современных корпусах представлены и более свободные виды письменной коммуникации, например тексты из блогов. Такие корпуса, например основной и газетный корпуса НКРЯ, способны ответить на вопрос о том, каков современный – связанный с нормой – узус в письменном языке. Устный же корпус рассказывает нам о том, какие модели принимаются именно в спонтанной, неподготовленной устной речи. Спонтанность и аутентичность языкового материала – явный плюс такого типа корпуса.

В то время как поиск определенного словосочетания в корпусе с письменными текстами (например, в основном корпусе НКРЯ) говорит о частотности и возможных контекстах словосочетания, такой же поиск в устном корпусе свидетельствует о том, как данное словосочетание действует как конструкция или синтаксическая единица.

В устном корпусе НКРЯ есть и такие части, которые являются в некоторой степени или полностью подготовленными. Эти элементы существенным образом отличаются от спонтанной речи. Их (в том числе транскрипции кинофильмов) можно при желании отключить от поиска. Они, однако, тоже демонстрируют те свойства, благодаря которым устная речь в фильмах звучит как настоящая устная речь.

Проиллюстрируем сказанное с помощью примера употребления глагола *поддерживать*. Подчеркнем, что важным тут является не учет конкретных значений этого слова, а методологический аспект: направленность на выявление того, как можно изучать концептуальные процессы с помощью корпуса устной речи. Глагол исследуется в активной конструкции, где перед ним находится существительное или местоимение в именительном падеже, и на расстоянии от 1 до 2 сам глагол в форме *поддерживает* (т. е. между подлежащим и сказуемым может быть максимум одно слово) и после глагола сразу следует прямое дополнение. Чтобы все условия поиска были идентичными, между словами не допускаются знаки препинания.

Данный поиск в корпусе был преднамеренно ограничен таким образом, что вхождений будет не очень много, но те, которые будут, совершенно точно соответствуют желаемой конструкции. Если бы мы хотели изучить данный вопрос шире, можно было бы выбрать и другие похожие по семантическому наполнению конструкции, как, например, *способен/способна/способно поддерживать* или *может/могут поддерживать*. Такие поиски дали бы дополнительную информацию о том, как человек воспринимает процесс поддержания и свойства возможных актантов при нем. Подобным образом изучение употребления глагола совершенного вида *поддержать* могло бы раскрыть новые свойства. Следует, однако, различать здесь две позиции. Если хотим



слова *идеи, муж, наука, тематика, политика, президент, правительство, учитель* и т. д. Таким образом, употребление данной глагольной формы не является семантически ограниченным. При анализе с позиций концептуализации небезынтересно заметить, что употребление формы третьего лица единственного числа в форме настоящего времени (нсв) свидетельствует о концептуализации. В процессе ее реализации существенны такие факторы, как активное участие действующего актанта (субъекта-агенса), положительный оттенок глагола, отвлеченность всех употреблений (ведь *поддерживать* может иметь и совершенно конкретное значение опоры, даже если оно в современном мире, возможно, менее центрально, чем разные виды абстрактной поддержки). Все это, – что как говорящий, так и его языковая среда предположительно знают о поддержании и его проявлении в реальном мире, – части энциклопедической информации о данном слове, представленной в общей концептуальной базе.

## 5. Заключение

В статье – с помощью двух поисков в устном корпусе НКРЯ – обсуждались возможности и как следствие преимущества корпуса при изучении концептуализации. По мере того, как данный корпус будет увеличиваться и развиваться, будут решены и затронутые в статье проблемы. Мы убеждены, что с помощью таких материалов сможем не только избежать возможного при неудачном стечении обстоятельств методологического срыва, но и предсказываем данной области исследования методологически блестящее будущее.

## Литература

- Гришина Е. А., Савчук С. О., 2009, Корпус устных текстов в Национальном корпусе русского языка: состав и структура [в:] *Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы*, Санкт-Петербург: Нестор-История, с. 129–148.
- Земская Е. А. (ред.), 1973, *Русская разговорная речь*, Москва: Наука.
- Земская Е. А. (ред.), 1983, *Русская разговорная речь. Фонетика. Морфология. Лексика. Жест*, Москва: Наука.
- Земская Е. А., 1987, *Русская разговорная речь: лингвистический анализ и проблемы обучения*, Москва: Русский язык.
- Земская Е. А., Китайгородская М. В., Ширяев Е. Н., 1981, *Русская разговорная речь. Общие вопросы. Словообразование. Синтаксис*, Москва: Наука.
- Китайгородская М. В., Розанова Н. Н., 2010, *Языковое существование современного горожанина. На материале языка Москвы*, Москва: Языки славянских культур.
- Лаптева О. А., 2008, *Русский разговорный синтаксис*, Москва: Издательство ЛКИ.



- Мустайоки А., Копотев М. В., Бирюлин Л. А., Протасова Е. Ю. (ред.), 2008, *Инструментарий русистики: корпусные подходы*, (*Slavica Helsingiensia*, Vol. 34), Хельсинки: Yliopistopaino.
- Barcelona A. (ed.), 2000, *Metaphor and Metonymy at the Crossroads. A Cognitive Perspective*, (*Topics in English Linguistics*, 30), Berlin: Mouton de Gruyter.
- Barnden J. A., 2010, Metaphor and metonymy: making their connections more slippery, *Cognitive Linguistics*, Vol. 21–1, с. 1–34.
- Dirven R., Pörings R. (eds.), 2003, *Metaphor and Metonymy in Comparison and Contrast*, Berlin: Mouton de Gruyter.
- Evans V., 2009, *How Words Mean. Lexical Concepts, Cognitive Models, and Meaning Construction*, Oxford: Oxford University Press.
- Janda L. (ed.), 2013, *Cognitive Linguistics: the Quantitative Turn: the Essential Reader*, Berlin: Mouton de Gruyter.
- Lakoff G., Johnson M., 1980, *Metaphors We Live By*, Chicago (IL): University of Chicago Press.
- Lakoff G., Johnson M., 1999, *Philosophy in the Flesh. The Embodied Mind and Its Challenge to Western Thought*, New York: Basic Books.
- Viimaranta J., 2012, The metaphors and metonymies of domination: explaining the different meanings of the Russian prefix *нод-*, *Russian Linguistics*, Vol. 36 (2), с. 157–174.
- Wichmann A., 2009, Speech corpora and spoken corpora [в:] A. Lüdeling, M. Kytö (eds.), *Corpus linguistics. An international handbook*, Handbücher zur Sprach- und Kommunikationswissenschaft, 291, Berlin: Mouton de Gruyter.