

Компьютерная лингвистика и интеллектуальные технологии:  
по материалам международной конференции «Диалог 2017»

Москва, 31 мая — 3 июня 2017

## **ОПЫТ КОРПУСНОГО МОДЕЛИРОВАНИЯ ФАКТОРОВ МЕТАФОРИЧНОСТИ НА ПРИМЕРЕ РУССКИХ ГЛАГОЛОВ<sup>1</sup>**

**Бадрызлова Ю. Г.** (yuliya.badryzlova@gmail.com)

Национальный Исследовательский Университет «Высшая  
Школа Экономики» (НИУ ВШЭ), Москва, Россия

## **A CORPUS-BASED STUDY OF FACTORS AND MODELS METAPHORICITY; EVIDENCE FROM RUSSIAN VERBS**

**Badryzlova Yu. G.** (yuliya.badryzlova@gmail.com)

National Research University Higher School of Economics (HSE),  
Moscow, Russia

The paper presents a tentative scheme for the analysis and annotation of the factors that contribute to metaphoricity of verbs in real-world data. The need for a comprehensive metaphor annotation scheme was demonstrated in the comparative study of four state-of-the-art automatic metaphor identification systems (Dunn, 2013c), which showed low agreement on a gold standard. The proposed scheme aims to capture a variety of factors involved in metaphoricity, their cumulative effect, and the gradual nature of metaphoricity.

The proposed inventory of metaphoricity factors includes nonbasic meanings (the less concrete, body-related and precise meaning(s) of a word); semantic shifts (additional semantic components of the meaning), newly attested meanings; morphologically motivated metaphoricity; different types of personification (a switch in the semantic class of the

---

<sup>1</sup> В данной научной работе использованы результаты проекта «Точки роста в языковой системе: корпусные исследования и моделирование», выполненного в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» в 2017 году.

argument from animate to inanimate); a switch in the semantic class of the argument between concrete and abstract; the use of a metonymic argument, participation in idiomatic expressions; and direct metaphors — the direct use of a word involved in a cross-domain mapping.

We also attempt to represent the dynamic interactions between these factors using graph-like visualization. Personifications and direct metaphors are described in terms of conceptual sources and targets. Secondary metaphoricity is attested for words that participate in extended metaphors. Metaphorically related words can be signaled by different types of cues: graphically, or by specific grammatical constructions, or by lexical cues.

The evidence is drawn from 2,036 occurrences of 100 Russian verbs sampled from the Russian syntactically annotated corpus SynTagRus.

**Keywords:** Factors of metaphoricity, metaphor corpus, metaphor annotation scheme, Russian verbal metaphor, dynamic metaphoricity

## Введение

Разработка систем автоматической идентификации метафоры является активно развивающейся областью автоматической обработки текста, в которой используется машинное обучение на основе различных технологических подходов и инструментов — как с использованием созданных вручную баз данных и онтологий, так и с применением статистических, дистрибутивно-семантических методов. Исчерпывающий обзор систем, технологий и инструментов представлен в работе (Veale et al., 2016). Однако при ознакомлении с дизайном этих систем нельзя не заметить, что они значительно различаются не только технологически, но и лежащими в их основе теоретическими положениями, а также задачей и объектом идентификации. По сути, объекты, идентифицируемые различными системами, представляют собой явления разного порядка.

## 1. Различные трактовки метафоры в корпусных ресурсах и автоматических системах

Подобное разночтение в понимании объекта метафоры среди создателей систем автоматической идентификации было наглядно продемонстрировано в проведенном Дж. Данном (Dunn, 2013c) эксперименте по сравнению эффективности четырех систем на едином золотом стандарте. Три из принявших участие в сравнении систем (1–3) основаны на методах дистрибутивной семантики, а одна (4) — на онтологических методах. Однако обратим здесь внимание на различие в понимании их разработчиками того, что собственно является метафорой.

- 1) В основе системы, предложенной (Li and Sporleder, 2009; Li and Sporleder, 2010), лежит предпосылка, что метафорическое употребление является семантической аномалией на фоне неметафорического контекста. Система идентифицирует фразы, являющиеся идиоматическими выражениями, например, “*to trip the light fantastic*” (=to dance), а также отдельные слова в переносном значении, например, “*During the Iraq war, he was*

*a sparrow; he didn't condone the bloodshed but wasn't bothered enough to go out and protest.*” При этом из числа идентифицируемых метафор исключаются высокочастотные лексикализованные метафоры, например, ‘*summit*’ в значении «встреча на высшем уровне».

- 2) Система, разработанная (Turney et al., 2011), определяет степень абстрактности/конкретности слова и сравнивает ее с аналогичными показателями контекста. Теоретическую основу составляют положения теории концептуальной метафоры (Lakoff and Johnson, 1980), согласно которым метафора представляет собой концептуальный перенос из конкретной сферы-источника на абстрактную сферу-цель. Таким образом, объект данной системы главным образом представлен лексикализованными метафорами типа “*deep appreciation*” (в отличие от неметафорического употребления “*deep snow*”).
- 3) Основанием системы, представленной в (Shutova, Sun and Korhonen, 2010), также служит теория концептуальной метафоры. Система осуществляет в тексте поиск лексических единиц, представляющих концепты сферы-источника и сферы-цели. Предполагается, что и источник, и цель должны быть выражены на лексическом уровне, причем глагол представляет сферу-источник, а существительное, находящееся к нему в синтаксическом отношении подлежащего или дополнения, представляет сферу-цель. Таким образом, система идентифицирует пары глагол-прямое дополнение (например, “*grasp theory*”, “*suppress memory*”) и субъект-глагол (например, “*tension mounted*”, “*ideology embraces*” и т.п.).
- 4) Система, описанная в работе (Dunn, 2013a), как и предыдущая, нацелена на поиск лексических единиц, репрезентирующих в тексте сферу-источник и сферу-цель, однако наличие между ними синтаксической связи не является обязательным. Поиск осуществляется на основе онтологического подхода (Nirenburg and Raskin, 2004; Niles and Pease, 2001).

В качестве золотого стандарта для сравнения перечисленных систем Дж. Данн использовал Амстердамский корпус метафоры (сокращенное название VUAMC, по месту создания — Vrije Universiteit Amsterdam) (Steen et al., 2010a), который, в свою очередь, основан на собственных теоретических предположениях и методологических принципах.

Англоязычный корпус VUAMC, содержащий около 200 тыс. словоупотреблений, размечен в соответствии с принципами процедуры MIPVU (сокращение от Metaphor Identification Procedure Vrije Universiteit) (Steen et al., 2010b). В основе процедуры лежит определение базового значения (БЗ) слова. БЗ — это такое из зафиксированных в словаре значений слова, которое является наиболее конкретным, телесно-ориентированным и четким (ibid.: 5–6).

Если в рассматриваемом контексте слово употреблено в значении, которое обладает менее ярко выраженной конкретностью, телесноориентированностью и четкостью, чем БЗ, то такое употребление в MIPVU обозначается термином «непрямая метафора» (indirect metaphor), например:

- (1) *Mr Franklin went there at the end of the 1970s, after the collapse of Keyser Ullman...*

Установление БЗ и определение степени контраста между ним и контекстуальным значением представляет собой наиболее уязвимый и подверженный субъективизму момент процедуры MIPVU; однако он неизбежен и снимается привлечением к анализу двух и более аннотаторов и принятием согласованных решений.

Помимо не прямой метафоры, процедура MIPVU и размеченный в соответствии с ней корпус VUAMC выделяют прямую метафору (*direct metaphor*) и потенциальное олицетворение (*possible personification*).

Прямая метафора — это употребление лексической единицы в ее прямом значении для обозначения концепта, принадлежащего к иной концептосфере, чем сама эта лексическая единица, т. е. когда имеет место междоменный перенос:

(2) *He's like a favourite old coat.*

Потенциальное олицетворение определяется как употребление неантропной (не обозначающей человека) лексемы в качестве аргумента лексической единицы, которая в своем базовом значении требует антропного аргумента:

(3) *Paula's stomach turned a somersault.*

Кроме того, в корпусе MIPVU размечены случаи, когда одновременно реализуются два из вышеперечисленных типов метафоры (*Double*), и пограничные случаи между метафорой и неметафорой (*When In Doubt Leave It In, WIDLII*).

В ходе тестирования по стандарту VUAMC перечисленные выше системы продемонстрировали различные показатели эффективности (см. Dunn, 2013c: 4–9). Однако наиболее интересный для нас результат состоит в невысоком уровне их согласованности (*Fleiss' Kappa*), варьирующемся в диапазоне от 0,259 до 0,293. Это означает, что в большинстве случаев системы принимают правильные и ошибочные решения в различных, не совпадающих друг с другом случаях, что вполне ожидаемо, учитывая представленные выше различия во взглядах на к определению предмета и теоретические предпосылки. На основании этого результата автор эксперимента делает вывод о том, что для эффективной идентификации метафоры на реальных данных необходим синтез подходов и технологий.

Такой синтезирующий подход был реализован в системе *MetaNet* (Dodge, Hong and Stickles, 2015), однако информацией о тестировании этой системы на каком-либо золотом стандарте мы на настоящее время не располагаем.

Для данной статьи наибольший интерес представляют теоретические выводы, к которым приходит Дж. Данн по результатам сравнения систем (Dunn, 2013c):

1. Метафора представляет собой градиентное явление. Одни метафорические высказывания обладают более высокой степенью метафоричности, тогда как другие менее метафоричны. Проблема протестированных систем состоит в том, что они пытаются проводить строгое бинарное разделение между явлениями, имеющими градиентную природу.
2. Метафоричность высказывания может происходить из нескольких различных источников.

В последующем исследовании (Dunn, 2013b) автор предлагает оценивать степень метафоричности высказывания в зависимости от его насыщенности метафорическим материалом.

Эти выводы, а также наблюдения, полученные нами в ходе предшествующих исследований (разметка русскоязычного корпуса по аналогии с VUAMC (Badryzlova, Kerimov and Shekhtman, 2013; Badryzlova et al., 2013) и идентификация лексикализованной метафоры методами машинного обучения с применением дистрибутивной семантики (Badryzlova, 2015; Panicheva and Badryzlova, 2017)) — убеждают нас в востребованности более детализированного корпусного описания факторов и источников метафоричности, включающих в себя собственную семантику лексических единиц, их синтагматические взаимодействия на семантическом и синтаксическом уровне, изменения собственной семантики и сочетаемости, происходящих в результате этих взаимодействий, а также метафорические переходы между концептосферами. Подобное описание, опирающееся на существующий лексикографический и корпусный инструментарий, призвано способствовать более полному моделированию метафорических явлений для целей автоматической идентификации метафоры.

## 2. Методика исследования

Представленная в Разделе 3 классификация факторов метафоричности и методика их описания носят предварительный характер. На настоящем этапе нами ставилась задача обеспечить как можно более подробную детализацию разметки, чтобы максимально охватить круг явлений, способствующих возникновению метафоричности. В дальнейшем эта методика может быть упрощена и адаптирована для целей автоматической аннотации. Представленные количественные результаты носят предварительный характер и для подтверждения достоверности подлежат анализу с участием дополнительных аннотаторов и определением меры согласованности, как принято в практике разметки метафорических ресурсов.

Анализ производился на материале глаголов в синтаксически размеченном корпусе СинТагРус (Boguslavsky et al., 2009), для чего были отобраны 100 глаголов с частотностью 19–25 вхождений и общим числом 2036 вхождений.

В качестве основного лексикографического источника для определения значений лексических единиц использовалась электронная версия Словаря русского языка под ред. А. П. Евгеньевой (Yevgenyeva, 1981–1984). В случае отсутствия в нем того или иного значения применялись онлайн-версия Толкового словаря С. И. Ожегова и Н. Ю. Шведовой (Ozhegov and Shvedova) и онлайн-словарь Русский Викисловарь (Russian Wiktionary).

Разметка производилась с помощью инструмента лингвистического аннотирования brat (Stenetorp et al., 2012).

## 3. Факторы метафоричности

Поскольку анализ производился на материале, представленном глаголами с определенными характеристиками (невысокая частотность, ограниченная полисемия), то предлагаемое описание не претендует на исчерпывающую полноту. Возможно, что включение лексических единиц с другими

характеристиками (частота, количество значений, часть речи) потребует расширения или дополнения предлагаемого описания.

При разработке системы факторов мы опирались на исследования механизмов языкового расширения и типов производных значений, представленных в работах Е. В. Падучевой (Paducheva, 2004) и Г. И. Кустовой (Kustova, 2004).

### 3.1. Небазовые значения

Базовые значения (БЗ) определяются в соответствии с критериями, сформулированными для процедуры MIPVU.

Все значения, не попавшие в группу БЗ, относятся к группе небазовых значений (НБЗ), которые представляют собой один из факторов метафоричности.

Каждое глагольное употребление в корпусе получает помету [BM] (basic meaning) либо [NBM] (non-basic meaning), например:

(4) «Гореть»:

- 'Поддаваться действию огня, уничтожаться огнем.':

[BM]

На юге Москвы горит жилой дом, сообщили РИА "Новости" ...  
в столичном управлении государственной противопожарной службы.

- 'Давать свет, пламя.':

[BM]

Если горит зеленый свет, это значит, что предыдущий поезд  
уехал очень далеко и можно смело отправляться в путь ...

- 'Топиться (о печи).':

[BM]

Печка горела бойко, светлые пятна и темные тени прыгали по стене ...

- 'Быть охваченным каким-л. сильным чувством, со страстью отдаваться чему-л. (какому-л. чувству, делу).':

... Директором и Генеральным конструктором назначают Валентина Глушко,

[NBM]

горящего желанием поставить победную точку в давнем споре с Королевым ...

- 'Блестеть под влиянием какого-л. чувства, мысли (о глазах).':

[NBM]

... (ее) глаза из-под платка горели страстью очереди.

- 'Быть под угрозой срыва, провала из-за опоздания, упущения сроков.':

Обком партии поручал Сафронову под личную ответственность

[GRAPH\_CUE]

[NBM]

наиболее важные, часто "горящие" объекты ...

([GRAPH\_CUE] — graphic cue: неалфавитный знак, указывающий на наличие метафоричности, в данном случае кавычки.)

В некоторых случаях бывает проблематично однозначно классифицировать значение как БЗ или НБЗ — например, когда значение, уже являясь небазовым, продолжает быть слабо связанным со значением, обладающим характеристиками предметности или телесности. Такие значения помечаются тэгом [DUAL]:

(5) «Дрожать»: ‘Испытывать опасение, тревогу; опасаться.’:

Но, как это бывает с людьми, добившимися положения исключительно благодаря

свирепому комплексу неполноценности и потому DUAL CONSTR дрожащими за свой авторитет, он никогда не признавал своих промахов.

([CONSTR] — construction transformation: изменение конструкции, см. п. 3.7.)

Лексемы, все значения которых являются только базовыми или только небазовыми, не обладают собственной внутренней метафоричностью, вызываемой противопоставлением БЗ и НБЗ. Несмотря на это, в некоторых случаях такие лексемы могут становиться метафоричными под влиянием факторов, описанных ниже.

Все значения лексем, не обладающих ни одним БЗ, получают специальную помету [ALL\_NBM] (all non-basic):

(6) «Уважать»:

- ‘Испытывать чувство уважения, почтения по отношению к кому-, чему-л.’:

Шаманизм был истреблен в советское время, хотя раньше на острове

ALL\_NBM  
жили около 40 шаманов, уважаемых во всем Забайкалье.

- ‘Принимать что-л. во внимание, считаться с чем-л., относиться с уважением.’:

По-разному можно относиться к судам, но я в любом случае

ALL\_NBM  
уважаю их решение как законопослушный гражданин.

- ‘Любить, ценить за что-л.’:

ALL\_NBM  
"Финские унитазаы я вообще-то не очень уважаю", - плавно жуя, сказал сантехник.

## 3.2. Семантические сдвиги

Семантические сдвиги ([SEM\_SHIFT], semantic shift) значения констатируются в тех случаях, когда помимо основной семантики лексемы, выраженной словарным определением, в рассматриваемом контексте у значения появляются дополнительные семантические компоненты, не входящие в определение, например:

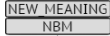
(7)

SEM\_SHIFT  
BM  
В годы юности Ильи Ильича девушки бегали в домотканых сарафанах - и нравились.

(Основная семантика: «Бегать»: ‘Быстро, поспешно идти куда-л., откуда-л.’; дополнительный компонент: «Носить»: ‘Надевать на себя что-л., украшать себя чем-л. по принятой моде.’)

### 3.3. Инновации

При образовании нового значения ([NEW\_MEANING]), в отличие от семантического сдвига, собственная семантика исходного значения практически полностью вытесняется новым семантическим наполнением:

- (8)  ... на сайты, заподозренные в связях с экстремистами, сбрасываются послания на арабском, подписанные "Абдуллой Ахмедом".

«Сбрасывать» (син. «переслать» — ‘Отослать, отправить.’ (дополнительный компонент — ‘электронное устройство’) и «записать» — ‘Комп. Сохранить данные на устройство хранения информации.’)

Особый случай представляют значения, принадлежащие к терминологическим метаязыкам. Зачастую образование термина ([TERM]) происходит путем метафоризации по основанию, которое является непрозрачным для среднего носителя языка, не являющегося специалистом в предметной области:

- (9) Тридцать лет назад в ультразвуковой диагностике начали развиваться

 так называемые методы согласованной фильтрации ...

### 3.4. Метафоричность, обусловленная внутренней формой

Метафоричность, обусловленная внутренней формой слова, наблюдается у лексем, не обладающих собственным БЗ или обладающих слабо выраженным БЗ, когда их морфологические компоненты этимологически связаны с другой лексемой, демонстрирующей в современном языке ярко выраженные характеристики БЗ. Например, глагол «исчерпывать», обладающий слабо выраженным БЗ ‘Полностью использовать.’, этимологически связан с «черпать», который обладает ярко выраженным БЗ ‘Захватывая чем-л., набирать, доставать (обычно снизу, из глубины) что-л. жидкое, сыпучее.’

### 3.5. Олицетворение

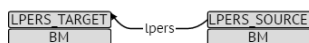
В отличие от корпуса VUAMC, определяющего олицетворение как смену антропного аргумента на неантропный, мы предлагаем рассматривать олицетворение шире — как смену класса аргумента с одушевленного на неодушевленный. При этом мы различаем источник олицетворения [TARGET] — лексему, чей аргумент меняет класс, и цель олицетворения [SOURCE] — лексему, заполняющую позицию аргумента.



В зависимости от того, насколько одушевленность аргумента интегрирована в значение, мы предлагаем различать лексикализованное олицетворение, олицетворение, образованное контекстуальной сменой класса аргумента, и метафорическое олицетворение.

Лексикализованное олицетворение ([LPERS], lexical personification) выражается зафиксированным в словаре значением (БЗ или НБЗ) лексемы, которое, во-первых, в качестве аргумента принимает только неодушевленные объекты, и, во-вторых у этой же лексемы имеется БЗ, требующее исключительно одушевленного аргумента:

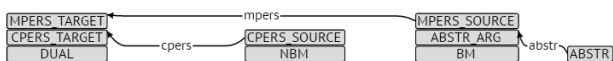
- (10) «Съедать»: ‘Причинить повреждения чему-л. (о редких, ядовитых веществах, ржавчине и т. п.):’:



Но, несмотря на все меры ухода, ржавчина понемногу съедает старые мечи.

Олицетворение, образованное контекстуальной сменой класса аргумента с одушевленного на неодушевленный ([CPERS], contextual personification) происходит, когда значение лексемы, допускающее выбор между этими двумя типами аргумента, в рассматриваемом контексте принимает неодушевленный аргумент:

- (11) «Завоевывать»: ‘Привлечь на свою сторону, расположить к себе.’:



Это произведение во все времена завоевывало сердца, покупало свободу и стоило людям жизни.

([ABSTR\_ARG и ABSTR] — смена семантического класса аргумента с конкретного на абстрактный, см. п. 3.6)

Метафорическое олицетворение ([MPERS], metaphoric personification) наблюдается, когда значение лексемы (БЗ или НБЗ) в норме требует только одушевленного аргумента, но в рассматриваемом контексте принимает неодушевленный аргумент, как «покупать» в примере (11).

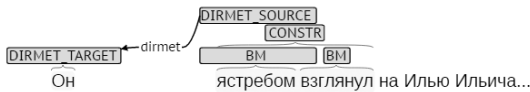
### 3.6. Смена семантического класса аргумента: конкретный → абстрактный

Семантическая сочетаемость некоторых значений определена достаточно широко таким образом, что она допускает как абстрактные, так и конкретные аргументы. Когда в рассматриваемом контексте выбор происходит в пользу абстрактного аргумента [ABSTR], этот выбор может придавать глаголу [ABSTR\_ARG] некоторую степень метафоричности: см. «покупать» в (11).

### 3.7. Трансформация конструкции

Специфические модели грамматической сочетаемости ([CONSTR]) могут являться индикаторами метафоричности. Например, конструкция ДРОЖАТЬ ЗА СУЩ<sub>ВИН</sub> в примере (5) является индикатором метафорического употребления, выраженного НБЗ. А в следующем примере инструментальная конструкция ВЗГЛЯНУТЬ СУЩ<sub>ТВОР</sub> сигнализирует источник прямой метафоры (см. п. 3.10):

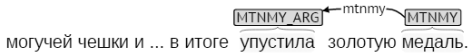
(12)



### 3.8. Метонимический сдвиг в аргументе

Метонимический аргумент, сам являясь результатом семантического переноса, предположительно, может оказывать влияние и на семантику глагола. (Глагол, принимающий метонимичный аргумент — [MTNMY\_ARG], метонимичный аргумент — [MTNMY]):

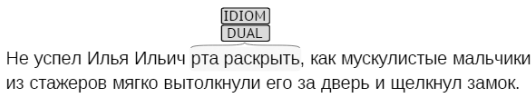
(13) Юлия Чепалова так увлеклась борьбой с полькой, что прозевала рывок



### 3.9. Вхождение в состав идиоматических единиц

Такое фундаментальное качество идиом [IDIOM] как переосмысление лексико-грамматического состава позволяет отнести их к числу факторов, способствующих метафоричности. При этом фразеологические сращения, утратившие мотивировку значения, всегда будут иметь небазовое значение; в то же время, значения фразеологических единств — единиц, в той или иной степени сохраняющих внутреннюю форму, в некоторых случаях могут быть описаны как [DUAL]:

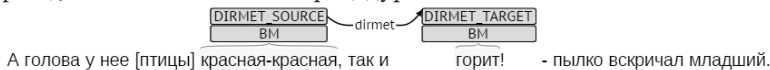
(14)



### 3.10. Прямая метафора

Для обозначения лексических единиц в прямом значении, участвующих в схемах концептуальных переносов из сферы-источника [SOURCE] на сферу-цель [TARGET], мы используем термин «прямая метафора» ([DIRMET], direct metaphor), заимствованный из процедуры MIPVU:

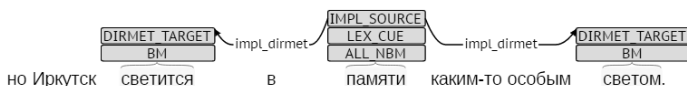
(15)



В этом предложении наблюдается метафорическая схема с возможной формулировкой ЯРКО-КРАСНЫЙ ЦВЕТ ЭТО ПЛАМЯ. При этом и сфера-источник, и сфера-цель выражены эксплицитно, т.е. представлены соответствующими лексическими единицами.

Однако зачастую на поверхностном уровне лингвистической структуры может быть представлена только одна из концептосфер, участвующих в переносе:

(16) Многие, случившееся куда позже, давно позабыто,

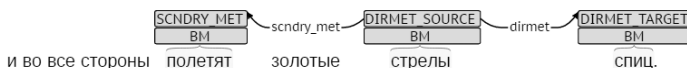


В этом примере активируется схема переноса, которую можно условно сформулировать как ХОРОШИЕ ВОСПОМИНАНИЯ ЭТО СВЕТ, при этом лексическое выражение получает только сфера-источник. Однако в предложении содержится лексический ключ [LEX\_CUE] «память» который уточняет имплицитно выраженную сферу-цель ([IMPL\_SOURCE]), указывая, к какой именно концептосфере происходит обращение.

### 3.11. Сложные метафорические комплексы

Описанные выше факторы метафоричности зачастую действуют не изолированно, а совместно друг с другом, образуя сложные метафорические комплексы. Особую сложность структура таких комплексов приобретает в случаях, когда в них задействована прямая метафора. В таких сложных комплексах некоторые лексические единицы наделяются вторичной метафоричностью ([SCNDRY\_MET], secondary metaphoricity) за счет их включения в качестве сторонних элементов в схему метафорического переноса, где они участвуют в развертывании метафоры. Например, в следующем предложении глагол «полететь» задействован в развертывании метафоры, заданной схемой переноса СПИЦЫ ВЕЛОСИПЕДНОГО КОЛЕСА ЭТО СТРЕЛЫ.

(17) [... я чувствовал щекощущий упор верткой педали под ступней, зуд в ладонях, вспотевших на горячих резиновых ручьяках руля, миг - и деревья послушно замелькают мимо меня, и заскользят под колесом их тени,]



В следующем примере (18) присутствует схема переноса: ВРАЩЕНИЕ ЗЕМЛИ ВОКРУГ СОЛНЦА ЭТО МЕХАНИЗМ, где сфера-цель выражена целой клаузой. В эту схему переноса в качестве вторичного развертывающего элемента [SCNDRY\_MET] входят глаголы «заводить», и «остановиться». «Заводить», в свою очередь, является источником олицетворения типа [MPERS] (метафорическое олицетворение), целью которого является неодушевленный аргумент «природа».

(18)



Очевидно, что метафорические комплексы способны проявлять большое разнообразие структур, вследствие чего описание таких комплексов представляет собой обширную задачу, требующую сбора, анализа и разметки репрезентативного корпусного материала и отдельного развернутого исследования.

## Результаты анализа

Предложенная схема аннотирования направлена на выявление и описание градиентной природы метафорических употреблений. Однако, если в целях сравнения воспользоваться общепринятым бинарным делением на метафорические и неметафорические употребления, то границу между этими двумя классами можно провести следующим образом: полностью неметафоричными можно считать употребления глаголов в базовом значении (БЗ) без каких-либо дополнительных факторов метафоричности. Употребления в небазовом значении (НБЗ), а также любые сочетания БЗ или НБЗ с дополнительными факторами можно считать метафорическими употреблениями, которые при этом различаются по степени насыщенности метафорическим материалом. С точки зрения такого бинарного разделения, 51,44% проанализированных в данном исследовании глагольных употреблений обладают метафоричностью. Это гораздо больше, чем в корпусе VUAMC, где только 18,68% глаголов были отнесены к классу метафорических употреблений (Steen et al., 2010b:194–208).

Можно предположить, что такое различие обусловлено как особенностями предложенной схемы аннотирования, так и характером выборки: она состоит из глаголов с невысокой встречаемостью в корпусе (19–25 вхождений) и небольшим количеством значений (в среднем около 5 значений на глагол), в то время как в корпусе VUAMC анализировались все глагольные вхождения, включая высокочастотные и высокополисемичные.

В Таблице 1 представлена совместная встречаемость некоторых факторов метафоричности с базовыми и небазовыми значениями (в % от всех метафорических употреблений).

**Таблица 1.** Совместная встречаемость факторов метафоричности с базовыми и небазовыми значениями (None — отсутствие дополнительных факторов метафоричности)

	NBM	BM	Total
None	39,42		
ABSTR_ARG	2,05	11,07	13,12
SEM_SHIFT	3,12	7,42	10,54
CPERS_SOURCE	5,59	2,05	7,64
DMET_SOURCE, SCNDRY_MET	1,83	4,19	6,02
CONSTR	5,7	0	5,7
MTNMY_ARG	2,15	3,12	5,27
LPERS_SOURCE	3,66	1,51	5,17
IDIOM	2,58	0	2,58
MPERS_SOURCE	0,76	1,62	2,38
NEW_MEANING	1,72	0,22	1,94
DMET_SOURCE	0,22	0,86	1,08
TERM	0,54	0,11	0,65

Как видно из таблицы, значительное число зафиксированных метафорических употреблений (39,42 %) выражено небазовым значением без каких-либо дополнительных факторов (NBM). Наличие абстрактного аргумента (ABSTR\_ARG) и семантического сдвига аргумента (SEM\_SHIFT) способствуют метафоричности в 13,12 % и 10,54 % случаев соответственно. Трансформация конструкции и метонимический сдвиг в аргументе встречаются в 5,7 % и 5,27 % метафорических употреблений соответственно.

Среди трех типов олицетворения наиболее частотным (7,64 %) является олицетворение, образованное контекстуальной сменой класса аргумента с одушевленного на неодушевленный (CPERS), тогда как лексикализованное олицетворение (LPERS) встречается в 5,17 % случаев. Наименее частотным из всех типов олицетворения является метафорическое олицетворение — 2,38 %. Прямая метафора почти всегда реализуется в составе сложных комплексов (DMET\_S, SCNDRY\_MET), где на нее воздействует вторичная метафоричность окружающих метафорических употреблений — 6,02 % (ср.: прямая метафора вне метафорических комплексов (DMET\_SOURCE) встречается только в 1,08 % случаев).

## Заключение

В данной статье была предложена методика анализа и описания факторов метафоричности глагола, а также их синтагматического взаимодействия. Очевидно, что не все из предлагаемых факторов имеют одинаковый вес в формировании метафоричности; вклад различных типов факторов в метафоричность может быть исследован психолингвистическими методами. Для подтверждения предложенных положений и получения репрезентативных количественных данных в перспективе необходимо привлечение более значительного объема материала (в т. ч. представляющего другие части речи), а также проведение анализа с привлечением большего числа аннотаторов.

## References

1. *Badryzlova Y. G.* (2015), Identification of Russian verbal metaphor by methods of machine learning trained on a corpus annotated for verbal arguments [Identifikatsiya metaforicheskikh upotrebleniy russkikh glagolov metodami mashinnogo obucheniya na osnove korpusa s rametkoy glagolnykh argumentov], MA thesis, National Research University Higher School of Economics, Moscow.
2. *Badryzlova Y. G., Isaeva E. V., Kerimov R. D., Shekhtman N. G.* (2013), Using the linguistic metaphor identification procedure (MIPVU) on a Russian corpus: rules revised and extended [Pravila primeneniya protsedury lingvisticheskoy identifikatsii metaphory (MIPVU) v russkoyazychnom korpuse: lingvokognitivnyy opyt (utochneniya i dopolneniya)], Humanitarian Vector [Gumanitarnyy Vektor], Vol. 4(36), pp. 19–29.

3. *Badryzlova Y. G., Kerimov R. D., Shekhtman N. G.* (2013), The main principles and methods of the Russian-language corpus of conceptual metaphor: using BRAT annotation tool. [Osnovnyye metody i printsipy sozdaniya russkoyazychnogo korpusa kontseptualnoy metaphory: opyt razmetki s primeneniyyem instrumenta BRAT], The Perm University Herald. Russian and Foreign Philology [Vestnik Permskogo Universiteta. Russkaya i Zarubezhnaya Filologiya], Vol. 3(23), pp. 82–92.
4. *Boguslavsky I., Iomdin L., Frolova T., Timoshenko S.* (2009), Development of a Russian Tagged Corpus with Lexical and Functional Annotation, MONDILEX Third Open Workshop, Bratislava, pp. 83–90.
5. *Dodge E., Hong J., Stickles E.* (2015), MetaNet: Deep semantic automatic metaphor analysis, Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Denver, pp. 40–49.
6. *Dunn J.* (2013a), Evaluating the premises and results of four metaphor identification systems, The 2013 International Conference on Intelligent Text Processing and Computational Linguistics, Samos, pp. 471–486.
7. *Dunn J.* (2013b), How linguistic structure influences and helps to predict metaphoric meaning, Cognitive Linguistics, Vol. 24(1), pp. 33–66.
8. *Dunn J.* (2013c), What metaphor identification systems can tell us about metaphor-in-language, The First Workshop on Metaphor in NLP, Atlanta, pp. 1–10.
9. *Kustova G. I.* (2004), Types of Derivative Meanings and Mechanisms of Language Extension [Tipy proizvodnykh znacheniy i mekhanizmy yazykovogo rasshireniya], Languages of the Slavonic Culture [Yazyki Slavyanskoy Kultury], Moscow.
10. *Lakoff G., Johnson M.* (1980), *Metaphors We Live By*, The University of Chicago Press, Chicago—London.
11. *Li L., Sporleder C.* (2009), Classifier combination for contextual idiom detection without labelled data, The 2009 Conference on Empirical Methods in Natural Language Processing, Vol. 1, Singapore, pp. 315–323.
12. *Li L., Sporleder C.* (2010), Using gaussian mixture models to detect figurative language in context, Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, pp. 297–300.
13. *Niles I., Pease A.* (2001), Towards a standard upper ontology, The 2nd international conference on Formal Ontology in Information Systems, Ogunquit, pp. 2–9.
14. *Nirenburg S., Raskin V.* (2004), *Ontological Semantics*, MIT Press, Cambridge, MA.
15. *Ozhegov S. I., Shvedova N. Yu., editors* (1997–1999), *Explanatory Dictionary of the Russian Language* [Tolkovyy Slovar Russkogo Yazyka], electronic resource: <http://slovari.ru/default.aspx?s=0&p=244>
16. *Paducheva E. V.* (2004), Dynamic Models in the Semantics of Vocabulary [Dinamicheskie modeli v semantike leksiki], Languages of the Slavonic Culture [Yazyki Slavyanskoy Kultury], Moscow.

17. *Panicheva P., Badryzlova, Y.* (2017), Distributional semantic features in Russian verbal metaphor identification, *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2017”* [Komp’yuternaya Lingvistika i Intellektual’nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii “Dialog 2017”], Moscow, In press.
18. *Russian Wiktionary* (2017), electronic resource: <https://ru.wiktionary.org/wiki>
19. *Shutova E., Sun L., Korhonen A.* (2010), Metaphor identification using verb and noun clustering, *Proceedings of the 23rd International Conference on Computational Linguistics, Beijing*, pp. 1002–1010.
20. *Steen G. J., Dorst A. G., Herrmann J. B., Kaal A., Krennmayr T.* (2010a), *VU Amsterdam Metaphor Corpus*, available at: <http://ota.ahds.ac.uk/headers/2541.xml>
21. *Steen G. J., Dorst A. G., Herrmann J. B., Kaal A., Krennmayr T., Pasma T.* (2010b), *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*, John Benjamins Publishing Company, Amsterdam—Philadelphia, PA.
22. *Stenetorp P., Pyysalo S., Topić G., Ohta T., Ananiadou S., Tsujii J.* (2012), *Brat: a web-based tool for NLP-assisted text annotation*, *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon*, pp. 102–107.
23. *Turney P. D., Neuman Y., Assaf D., Cohen Y.* (2011), *Literal and metaphorical sense identification through concrete and abstract context*, *The 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh*, pp. 680–690.
24. *Veale T., Shutova E., Klebanov, B. B.* (2016), *Metaphor: a computational perspective*, *Synthesis Lectures on Human Language Technologies, Vol. 9.1*, pp. 1–160.
25. *Yevgenyeva A. P., editor* (1981–1984), *Dictionary of the Russian Language, 2nd ed., Vol. 1–4*, *Russian Language [Russkiy Yazyk]*, Moscow.