

A Phrasematic Key to the Russian Verb-Nominal Constructions: The Workshop of the Computational Phraseography

Fedorushkov Yury Glebovich¹

Assistant professor at Department of Russian Language of Institute of Russian and
Ukrainian Philology, Adam Mickiewicz University,
Poznan, Poland.

(date of receiving: March, 2018; date of acceptance: September, 2018)

Abstract

The article focuses on the problem of the phrasematic status of verb-nominal phrases considered as "potential phrasemes", such as: *давить на психику*, *ходить на работу*, *забежать на минутку*, *забраться на крышу*, *зайти на страницу*. The analyzed list includes more than one thousand of such structures. A special approach to their valency autonomy in the context is proposed. We take into account the context as supraphrasematic *feature* of potential phrasemes. An individual approach to each potential phrasemes in a particular context has a special role to play. This allows creating a special experience that can be converted to computerized linguistic analysis. We propose to consider the possibility of automatization the detection of the phrasematic status of potential phrasemes by using *machine learning* methods (the artificial intelligence). The described methods allow to automatic retrieval of the potential phrasemes with synchronous recognition of their phrasematic status in particular text and big text corpora of Russian language.

Keywords: Phraseme, Collocation, Idiom, Grammeme, Morphosyntax, 3-Gram, Machine Learning, NLP, Text Corpora.

1. E-mail: jerfed@amu.edu.pl

Фразематический ключ к вербономинальным конструкциям русского языка: мастерская компьютерной фразеогрфии

Федорушков Юрий Глебович¹

Доцент кафедры русского языка института русской и украинской филологии
УАМ, Университет им. А. Мицкевича в г. Познань, Познань, Польша.
дата получения: март 2018 г.; дата принятия: сентябрь 2018 г.)

Аннотация

В статье заострена проблема фразематического статуса вербономинальных словосочетаний, рассматриваемых как "потенциальные фраземы" (ПФ), – таких как *давить на психику, ходить на работу, забежать на минутку, забраться на крышу, зайти на страницу*. В анализируемом списке свыше 1 тысячи таких конструкций. Предлагается особый подход к их валентной автономности в контексте, при этом учитывается сам контекст как супрафразематическая особенность ПФ. Особую роль имеет индивидуальный подход к каждой ПФ в конкретном контексте. Это позволяет создать особый опыт, который можно конвертировать для компьютеризации лингвистического анализа. К рассмотрению предлагается возможность автоматизации определения фразематического статуса ПФ с помощью искусственного интеллекта (методы *machine learning*). Описываемые методы позволяют автоматически эксцерпировать ПФ с параллельным определением их фразематического статуса как в отдельных текстах русского языка, так и в корпусных массивах.

Ключевые слова: Фразема, Коллокация, Идиома, Граммема, Морфосинтаксис, 3-Грамм, Машинное Обучение, Обработка Естественного Языка, Электронный Корпус.

1. E-mail: jerfed@amu.edu.pl

Введение

Данная статья является фрагментом более широкого исследования в области компьютерной лексикографии, связанной прежде всего с её прикладной областью – фразеографической мастерской. Мастерская, о которой идёт речь, касается изучения вербономинальных фразем русского языка с помощью *Системы автоматизации обнаружения и аутентификации многословных единиц во фразематическом ключе (САФК)*. Используются методические решения из сферы NLP (*Natural Language Processing*), а также инструмент *Word2vec*, основанный на векторном представлении слов в машинном обучении. Данный класс методов относится к области искусственного интеллекта. Целью статьи является презентация фразематического ключа, который использует в итоге САФК в процессе итерации языкового опыта, то есть своего рода самообучения, а также при принятии решения относительно конкретной конструкции. Эффект работы с САФК – автоматически найденные, сегментированные вербономинальные конструкции (о заданном морфосинтаксическом типе) и автоматическая индикация их фразематического статуса в виде определённого кода. Другими словами, система САФК в состоянии определить, являются ли такие выражения как *давить на психику, ходить на работу, бежать на работу, забраться на крышу, биться на randevу, выйти на площадку, ступить на землю, выйти на сцену, выходить на палубу, лечь на спину* и др. фраземами, а если так, то к какому, технически говоря, "типу" фразем они относятся: полукolloкациям, коллокациям, идиомам – или всё же это так называемые *свободные словосочетания* (см. ниже)? С формально грамматической точки зрения – это морфосинтаксические 3-граммы. На них мы и сосредоточим своё внимание в статье.

Основные теоретические принципы и идеи из области фразематики и семантики и почерпнуты из работ В. Хлебды, И. Мельчука, О. Федосова (см.

Литература). Данная статья по сути обзорная и состоит из идейных пролегомен для области междисциплинарного – на стыке лингвистики и информатики – методологического формализма. Описанные в статье разработки – на экспериментальной стадии и требуют углубленной модификации. Речь пойдёт о теоретико-прикладной фокусировке понятия *фразематический ключ*. Результаты исследования могут внести новый методологический опыт в современную фразеографию. Именно в связи с их интердисциплинарным характером подход к классификации фразем во многом упрощён, на что особенно стоит обратить внимание Читателя. Стоит отметить, что в эксперименте были использованы данные трёх корпусов: из архива газеты *Новая газета* за 2013 год (3421285 слов-вхождений, понимаемых как повторяющиеся сегменты знаков алфавита от пробела до пробела) были автоматически эксцерпированы 3-граммные вербономинальные конструкции; массив *CommonCrawl* – компрессированный архив текстов российского домена *.ru* за 2012 год (около 105 Гб; адрес архива данного массива: <http://statmt.org/ngrams/pages/raw-data.htm> (доступ: 03.01.2018), при чём точное количество слов (с повторениями) неизвестно: теоретически для массива за 2012 и 2013 год общее количество слов – около 500 миллиардов), а также *КРЛПМП – Корпус российской локальной прессы милленийного периода* (около 500 млн слов; подробное описание корпуса в публикации Федорушков 2009) – были использованы для машинного обучения САФК.¹

Основная часть

Прежде чем приступить к анализу, следует отметить, что в данной статье использована польская трактовка некоторых терминов. О разнице в трактовке

1. Значительную техническую поддержку в создании САФК оказали доктор Филипп Гралиньский и инженер Давид Юркевич факультета математики и информатики университета им. Адама Мицкевича в г. Познань, Польша.

терминов *денотат*, *десигнат* в польской и российской научной традиции – см. Маркевич 2007: 121. В польском языкознании денотат – это "множество всех (как прошлых, так и будущих) объектов внеязыковой действительности", напр. "денотат названия *философ* – это множество таких личностей, как Сократ, Платон, Лейбниц, Локк, Татаркевич" (Полански 1999: 111; "десигнат названия *стол* – это каждый экземпляр стола" (там же. 115) [перев. с польск. – Ю.Ф.]).

Итак, рассмотрим примеры: *давить на психику*, *ходить на работу*, *забежать на минутку*, *забраться на крышу*, *зайти на страницу*. Без контекста трудно определить, имеют ли данные выражения фразеологический статус. По мнению одних, это традиционно понимаемые коллокации, по мнению других, это *свободные словосочетания*, не связанные узлами фразеологизации. В конкретной ситуации при вербализации конкретных десигнатов мы создаем словосочетания, то есть конструируем кластеры слов свободно-условно. Рассмотрим другие примеры, но уже с контекстами:

i) *Это белая стена.*

белая стена – это *de facto* сиюминутное выражение, "словосочетание-однодневка", которое испарится, то есть не будет производиться, если, скажем, стена будет перекрашена в другой цвет: атрибут десигната будет иметь совершенно другую лексическую наполненность. Это будет уже, скажем, *красная стена*. Подобных примеров можно создать бесконечное количество. Ср. высказывание: "(...) громадный, теоретически неограниченный класс **свободных словосочетаний**" (Иорданская, Мельчук 2007. 223–224).
Например:

ii) *Это зелёное сукно.*

iii) *Он шагает по тротуару.*

iv) *Я открыл книгу на [число] странице.*

Другое дело, если мы скажем:

v) У него *белый билет*.

vi) Он *вышел на улицу*.

vii) Я *зашёл на страницу*.

белый билет (пример v) – это идиома, зарегистрированная в *Большом словаре русском поговорок* (Мокиенко, Никитина 2007), основана уже на модели метафоры, то есть данное выражение вербализирует не просто материально существующий десигнат названия (документ или удостоверение белого цвета), а метафора-формула, связанная с самим фактом, свидетельствующим об освобождении от воинской службы: ср. *белобилетник* – не просто тот, кто владеет *белым билетом*, а человек, освобождённый от службы в армии.

вышел на улицу (пример vi) – в зависимости от контекста это вербономинальное выражение (ВНВ) может иметь зачастую *околоколлокационные* (саму коллокацию мы понимаем согласно дефиниции: "слабоидиоматичные фразеологизмы преимущественно со структурой словосочетания, в которых семантически главный компонент (база) употреблен в своем прямом значении, а сочетаемость со вспомогательным компонентом (коллокатором) может быть задана в терминах семантического класса, но выбор конкретного слова предопределён узусом" – Баранов, Добровольский 2014. 73) или *полуколлокационные*, то есть слабоидеоматические статусы:

1) 'вышел на проезжую часть (отличную от тротуара)';

2) 'вышел на тротуар';

3) 'вышел во внутренний двор';

4) 'вышел в открытое пространство вне стен здания';

5) 'вышел на демонстрацию/ митинг';

6) 'вышел погулять'

7) в зависимости от более широкого контекста может иметь референциально отличное, конденсированное значение: 'вышел подышать свежим воздухом'

и ряд референциально переносных и коннотативно-контекстных значений.

В vii примере ВНВ *зашёл на страницу* абсолютно лишено дословности. Здесь имеется в виду одиночное действие – 'посещение интернет-сайта'.

В случае с примерами v–vii можно с большей уверенностью сказать, что используются готовые формулы – *фраземы*. С примерами i–iv дело обстоит сложнее: тут уже необходим широкий контекст, позволяющий снять неоднозначность фразематического статуса.

В нашей мастерской фраземы мы разделяем на коллокации и идиомы. В современной фразематике актуальной остаётся проблема чёткой калибровки фразематического статуса МЕ, то есть многословных единиц (пол. *wielowrazowców*), традиционно называемых словосочетаниями. Не случайно это именно с о ч е т а н и я . Слова сочетаются по правилам морфосинтаксиса (граммемной сочетаемости, норм валентности) и семантики (рациональная сочетаемость семем – ср. также с формулировкой *онтологическая несовместимость* – Апресян 1995. 161). Но, конечно, этих *stricte* лингвистически ориентированных факторов для фразем мало. Статус "являться фраземой" или "не являться фраземой" подразумевает под собой ориентиры, касающиеся воспроизводимости и повторяемости, то есть частотности в репродукции и/или продукции – так детально обсуждаемых Войцехом Хлебдой в его *Элементах фразематики...* (см. Хлебда 2003. 189–208). Воспроизводимость – особый критерий при выделении или эксцерпции в тексте/корпусе МЕ (на тему автоматизации эксцерпции МЕ – см. Федорушков

2010: 59–85; Федорушков, Шутковский 2010: 1564–1569). Именно репродукты имеют статус фраземы. Такие МЕ не создаются заново, они повторяют особый нейроопыт, то есть репродуцируют МЕ в готовой форме, формуле на уровне морфологии, синтаксиса, семантики, распределения актантных ролей. Модель идиоматичности, особой кодовой символичности основывается зачастую на глубокой и сложной модели переноса, то есть модели метафоры, метонимии, символичности, кодовой закреплённости, наследующей семантические микроотношения (на уровне семы, нескольких сем в семеме, семеме) – такие, как гипо/гиперонимия, меронимия, голонимия, синонимия/близкочность, антонимия, а также фуззинимия (по вопросу дефиниции – см. Интернет-страница Лексипедии – Lexipedia) и др. сематические и лексико-тематические отношения.

Какой же именно критерий является "центральным" при определении статуса МЕ как фраземы при адаптации языковых данных в программной среде? По нашему мнению, это контекст. Во-первых, контекст является супраопределителем значения десигната/денотата – ср.:

i) К ним пришла новая сотрудница. Она оказалась *чёрной овцой* в их офисном стаде. Этакой своенравной бабёнкой, которая, как оказалось, любила всеми помывать.

ii) На лугу паслись *чёрные овцы*.

Пожалуй, этот популярный пример не требует дополнительных рассуждений. В примере ix мы имеем дело с так называемой дословностью, то есть снятием метафорической модели и идиоматического статуса фраземы-идиомы. Во-вторых, расшифровать фразематический статус помогает не только А. расширенный контекст (пример viii, состоящий из трёх предложений), но и Б. – *мультиконтекст* – собрание контекстов, где используется МЕ только с определённым статусом (например, статусом

фраземы или не-фраземы), зависящим от критерия воспроизводимости. С определённой точки зрения, репродуктивность фраземы и является её воспроизводимостью. А и Б. – это важнейшие черты (англ. *features*) языкового сырья – лингвистического опыта, передаваемого в САФК.

Предлагаемые в начале статьи МЕ *давить на психику, ходить на работу, забежать на минутку, забраться на крышу, зайти на страницу* могут иметь разный статус в зависимости от контекста, а воспроизводимость (англ. *reproducibility*, пол. *odtwarzalność*) таких статусов для конкретной ВПВ создаёт фразематическую насыщенность той или иной ВНВ. Так, ВНВ в примере vi может быть полудословной МЕ, фраземой-коллокацией, фраземой-идиомой. Каждый из этих статусов подразумевает под собой присутствие целой серии контекстов (для каждого из статусов).

Данный подход к фразематическому статусу МЕ отображается в своеобразной фразематической шкале. Олег Федосов отмечает "(...) широко понятая сочетаемость означает, что на умозрительной шкале возможностей: **«свободные словосочетания – устойчивые неидиоматические словосочетания (коллокации) – идиоматические словосочетания (фразеологизмы в широком смысле)»** [выд. – Ю.Ф.] (...) далеко не все теоретически возможные сочетания лексем (слов, словоформ), которые объединяет основное качество – некоторая степень устойчивости" (Федосов 2014). С нашей точки зрения, в техническом плане такое классифицирование не противоречит классической, виноградовской классификации ФЕ: *фразеологические сращения, фразеологические единства и фразеологические сочетания* – сочетания с соответственно уменьшающейся метафоричностью. Особенность фразеологических единиц, прежде всего идиом, такова, что семантика конкретной идиомы не является суммой значений отдельно взятых её компонентов (см. Виноградов. 1977: 140–161).

Принимая во внимание приёмы фразематического анализа и классификации МЕ во фразематическом ключе, необходимо сказать, что для нас всё же важна поляризация статуса исследуемых ВНВ: это фраземы либо не-фраземы. Исследуемые ВНВ с "неприкреплённым" статусом мы называем ПФ – *потенциальные фраземы*. Поляризация фразематической шкалы в упрощённой форме выглядит следующим образом:

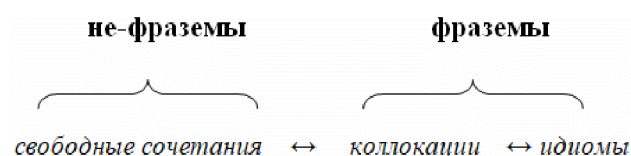


Схема 1. Поляризация во фразематической шкале.

Данный подход связан с самой сферой действий: экспериментом в области компьютерной лексикографии и информатики. Приёмы такого поверхностного фразематического анализа, классификации во фразематической шкале касаются, главным образом, адаптации лингвистических данных в информатическом поле действий. Иными словами, мы имеем дело с переносом лексикографических, то есть лингвистических действий, на поле информатики и программирования.

Однако контекст не всегда помогает определить фразематический статус ПФ. Необходима верификация повторяемости конкретной ПФ в репрезентативно больших текстовых массивах, а также в лексикографических источниках. ПФ может иметь морфосинтаксические варианты, так как границы фраземы иногда расплывчаты и могут включать факультативные члены (аргументы). Например, аргумент *приличные* в примере *x* в идиоме *в люди выйти*.

i) Надо в театр сходить, хочется *в люди выйти*.

ii) Можно и в приличные люди выйти (Интернет-источник: <http://alau.kz/raznye-noski-karima-masimova-stali-novym-trendom-v-seti/> (доступ: 22.01.2018).

В связи с этим надо уточнить: в своём исследовательском материале мы не учитываем вариантов – таких, как в примере xi.

Итак, при определении фразематического статуса мы опираемся на интуицию и знание нашей языковой компетенции. Конечно, зачастую без контекста ПФ определение её статуса не имеет смысла. Однако очень важно учитывать такое первое, интуитивное определение фразематического статуса для ПФ. Так, например, для ПФ *бегут на подмогу* мы можем дать вступительную оценку 'ПФ имеет статус фраземы' (**F** – см. ниже таблица 1). После этого мы исследуем контекст и иногда меняем решение, учитывая возможную поспешность выводов и огромное количество свободных словосочетаний: 'ПФ не имеет статуса фраземы' (**N-F**). Далее мы ищем ПФ в словарях (**Сл.**), где не находим искомой ПФ. Заглянув в поисковик Яндекс, мы обращаем внимание, что повторяемость данной ПФ – 166 (**YA**), а версии с инфинитивом (*бежать на подмогу*) – уже 1000 раз (**YA-inf**). Это заставляет нас исследовать контексты и прийти к выводу: на данную ПФ стоит обратить внимание, это вероятнее всего именно фразема, то есть мы вновь меняем статус ПФ на **F**, принимая во внимание и её идиоматическую природу (**F-I**). В итоге мы меняем, вернее, принимаем решение 4 раза. При адаптации данные квалификации (**Q**) должны иметь именно кодовые обозначения: 1) **F** → 2) **N-F** → 3) **F** → 4) **F-I**. Ниже представлены эффекты для ПФ. Отметим, что все используемые примеры почерпнуты из архива *Новая газета* за 2013. Веб-сайт: <https://www.novayagazeta.ru/> (доступ 30.11.2017).

Таблица 1. Фрагмент рандомизированного языкового материала, использованного в системе автоматизации обнаружения и аутентификации МЕ во фразематическом ключе (САФК).

	Контекст	Q ₁	Q ₂	Сл.	YA	YA (inf)	Q ₃	Q ₄
i)	«Беркут» продолжает налегать. Люди &&&бегут на подмогу#### к штабу с других концов Майдана. Кто-то кидает в толпу спецназа огнетушитель.	F	N-F	–	Y 166	Y 1000	F	F-I
ii)	Мне приходили идеи во сне, я просыпался и с первым поездом &&&бежал на работу#### , чтобы проверить их на опыте.	F	F	+У	Y 925	Y183	F	F-I
iii)	Хотя после открытия против нее уголовного дела об организации заказного убийства задача усложнилась многократно. Над задачей идет &&&биться на рандеву#### с Януковичем Арсений Яценюк.	N-F	N-F	–	Y 2	Y 2	N-F	N-F

Рассмотрим пример хiii. Система отыскала и верно выделила 3 аутентичных контекста и каждый ПФ отметила как **F-I**. Надо отметить, что в переработанном тексте для технических нужд отсутствуют знаки препинания:

Таблица 1. Фрагмент языкового материала из массива *CommonCrawl*. В случайных контекстах на основе полученного опыта САФК обнаружил, произвёл соответствующие обозначения границ ВНВ (&&&, ###) и фразематически идентифицировал ВНВ.

Контекст из массива <i>CommonCrawl</i>	Рекомендация САФК – автоматически присвоенный квалификатор, то есть фразематический статус на уровне решения Q₄
на киностудию возвращался с работы в двенадцать ночи и утром &&& бежал на работу### мне нравилось идти в легкий летний дождик чтобы по голове	F-I
скрывал что обзавелся люксовой машиной парковал ее за километр и &&& бежал на работу### через всю исаакиевскую площадь	F-I
только чтобы помучить моего друга иосифа ландау пристрелили когда он &&& бежал на работу### у эсэсовцев были специальные квартиры куда они уводили евреев чтобы	F-I

Заключение

Конечно, результаты данного эксперимента с использованием САФК требуют ряда дополнений и дискуссий. В статье не описана полная классификация фразематической шкалы, включающая (кроме прочих) квазиидиомы, а также учитывающая так называемую *валентную автономность* ПФ и особую классификацию "не-ПФ", т.е. ошибок – сочетаний слов, не являющихся 3-граммными ВНВ. Исследовательское поле

(языковое сырьё, т.е. начальный языковой материал) – свыше 1000 ПФ составляют ядро языкового сырья. Но на практике они и являются фразематическим ключом в его практическом применении, т.е. вводе и обработке в САФК. – см. 2–8 колонны (табл. 1) данных, включая контекст и квалификаторы. Следующая порция опыта – переработанные и исправленные варианты автоматической квалификации. В частности появляются области решения Q_5 и Q_6 , которые обогащают опыт САФК и в конечном итоге еще полноценнее минимизируют субъективную оценку исследователя относительно фразематического статуса вербономинальных конструкций.

Литература

- 1- Апресян, Ю.Д. (1995). *Избранные труды. Том II. Интегральное описание языка и системная лексикография*, Москва, Изд-во «Школа: «Языки русской культуры».
- 2- Баранов А.Н., Добровольский Д.О. (2014). *Основы фразеологии (краткий курс). Учебное пособие*, Москва, Изд-во «Флинта».
- 3- Виноградов В.В. (1977). *Об основных типах фразеологических единиц в русском языке*, [в:] В.В. Виноградов, Избранные труды. Лексикология и лексикография. Москва, Изд-во «Наука», 140–161.
- 4- Иорданаская Л.Н. Мельчук И.А. (2007). *Смысл и сочетаемость в словаре*, Москва, Изд-во «Языки славянских культур».
- 5- Лексипедия (Lexipedia). Лексико-семантический графовый словарь зависимостей. Интернет-источник: <http://www.lexipedia.com/english/fuzzynum> (доступ: 11.11.2013).
- 6- Маркевич, Х. (2007). *Utarczki i perswazje 1947–2006* (Markiewicz H.), Краков, Изд-во «Wydawnictwo Uniwersytetu Jagiellońskiego». Интернет-источник: <https://books.google.pl/books?id=Xg0ati-BKNAC&printsec=frontcover&hl=pl#v=onepage&q&f=false> (доступ: 23.01.2018).
- 7- Мокиенко, В.М., Никитина, Т.Г. (2007). *Большой словарь русских поговорок*, Москва, Изд-во «Олма Медиа Групп». Интернет-источник: https://dic.academic.ru/dic.nsf/proverbs/121_46/Белый (доступ: 03.01.2018).
- 8- *Толковый словарь русского языка*, под ред. Д.Н. Ушакова, Москва: Гос. ин-т "Сов. энцикл."; ОГИЗ; Гос. Изд-во «Иностран. и нац. слов.», 1935–1940, в 4-х т.

- 9- Федорущков Ю. (2010). Методы автоматизации экспертной атрибутивной конструкции русского языка (Fedorushkov Y., Metody automatyzacji ekspercyj konstrukcji atrybutywnych języka rosyjskiego). [в:] *Na tropach reproduktów. W poszukiwaniu wielowrazowych jednostek języka*, под ред. В. Хлебды, Ополе: Изд-во «Wydawnictwo UO», 59–85.
- 10- Федорущков Ю. (2009), Экспертная слов в кавычках из текстов российской прессы (Fedorushkov Y., Ekscerpca wyrazów cudzyslowowych z tekstów prasy rosyjskiej). серия: *Semiosis Lexicographica* ЛП, Варшава, Изд-во «Wydawnictwo TAKT».
- 11- Федорущков Ю., Шутковский Т. (2010). Лексико-грамматическая сочетаемость атрибутивных словосочетаний русского языка в контексте методов компьютерной экспертной, [в:] *II Международная конференция „Русский язык и литература в международном образовательном пространстве: современное состояние и перспективы”, Гранада, 8-10 сентября 2010 г.*, под ред. Р. Гузмана Тирадо, Л. Соколовой, И. Войтяковой, т. II, Гранада: Изд-во «Rubiños – 1860, S.A.», 1565–1569.
- 12- Федосов О.И. (2002). Передвигаются ли границы фразеологии? (О т.н. «лексических фраземах» на материале русского, чешского и венгерского языков), [в:] *Słowo. Tekst. Czas: nowa frazeologia w nowej Europie (Materiały VI Międzynarodowej Konferencji Naukowej, (Szczecin, 6-7 września 2001 r., Greifswald, 8-9 września 2001 r.)*, под ред. М.А. Алексеенко, В.М. Мокиенко, Г. Вальтера, Щецин – Greifswald, Изд-во «Wydawnictwo Naukowe US».
- 13- Федосов О.И. (2004). Коллокации на фоне комбинаторики лексем, [в:] *Frazeologia słowianska i inne płaszczyzny systemu językowego*, под ред. Я. Бартошевской, В.М. Мокиенко, Г. Вальтера, Гданьск, Изд-во «Wydawnictwo Uniwersytetu Gdańskiego», 37–49.
- 14- Хлебда В. (2003). Элементы фразематики: введение во фразеологию адресанта (Chlebda W. *Elementy frazematyki: wprowadzenie do frazeologii nadawcy*), изд. II, Ласк, Изд-во «Oficyna Wydawnicza Leksem».
- 15- *Энциклопедия общего языкознания (Encyklopedia językoznawstwa ogólnego)*. под ред. К. Полянского, (Polański, K.) (1999). Вроцлав, Изд-во «Zakład Narodowy im. Ossolińskich – Wydawnictwo».

Bibliography

- 1- Аpresjan, Ju.D. (1995). *Izbrannye trudy. Tom II. Integral'noe opisanie jazyka i sistemnaja leksikografija*, Moskva, Изд-во «Shkola: «Jazyki russoj kultury».
- 2- Baranov A.N., Dobrovolskij D.O. (2014). *Osnovy frazeologii (kratkij kurs). Uchebnoe posobie*, Moskva, Изд-во «Flinta».

- 3- Vinogradov V.V. (1977). *Ob osnovnyh tipah frazeologicheskikh edinic v russkom jazyke*, [v:] V.V. Vinogradov, Izbrannye trudy. Leksikologija i leksikografija. Moskva, Izd-vo «Nauka», 140–161.
- 4- Iordanaskaja L.N. Mel'chuk I.A. (2007). *Smysl i sochetaemost' v slovare*, Moskva, Izd-vo «Jazyki slavjanskih kul'tur».
- 5- Leksipedija (Lexipedia). Leksiko-semanticheskij grafovyy slovar' zavisimostej. Internet-istochnik: <http://www.lexipedia.com/english/fuzzynym> (dostup: 11.11.2013).
- 6- Markevich, H. (2007). *Utarczki i perswazje 1947–2006* (Markiewicz H.), Krakov, Izd-vo «Wydawnictwo Uniwersytetu Jagiellońskiego». Internet-istochnik: <https://books.google.pl/books?id=Xg0ati-BKNAC&printsec=frontcover&hl=pl#v=onepage&q&f=false> (dostup: 23.01.2018).
- 7- Mокienko, V.M., Nikitina, T.G. (2007). *Bol'shoj slovar' russkikh pogovorok*, Moskva, Izd-vo «Olma Media Grupp». Internet-istochnik: <https://dic.academic.ru/dic.nsf/proverbs/12146/Belyj> (dostup: 03.01.2018).
- 8- *Tolkovyj slovar' russkogo jazyka*, pod red. D.N. Ushakova, Moskva: Gos. in-t "Sov. jencikl."; OGIZ; Gos. Izd-vo «Inostr. i nac. slov.», 1935–1940, v 4-h t.
- 9- Fedorushkov Ju. (2010). Metody avtomatizacii jekscerpeii atributivnyh konstrukcij russkogo jazyka (Fedorushkov Y., Metody automatyzacji ekscerpeji konstrukcji atrybutywnych języka rosyjskiego). [v:] *Na tropach reproduktów. W poszukiwaniu wielowyrazowych jednostek języka*, pod. red. V. Hlebdy, Opole: Izd-vo «Wydawnictwo UO», 59–85.
- 10- Fedorushkov Ju. (2009), Jekscerpcija slov v kavychkah iz tekstov rossijskoj prasy (Fedorushkov Y., Ekscerpeja wyrazów cudzyslowowych z tekstów prasy rosyjskiej). serija: *Semiosis Lexicographica* LII, Varshava, Izd-vo «Wydawnictwo TAKT».
- 11- Fedorushkov Ju., Shutkovskij T. (2010). Leksiko-grammaticeskaja sochetaemost' atributivnyh slovosochetanj russkogo jazyka v kontekste metodov komp'juternoj jekscerpeii, [v:] *II Mezhdunarodnaja konferencija „Russkij jazyk i literatura v mezhdunarodnom obrazovatel'nom prostranstve: sovremennoe sostojanie i perspektivy”*, Granada, 8-10 sentjabrja 2010 g., pod red. R. Guzmána Tirado, L. Sokolovoj, I. Vojtjakovoj, t. II, Granada: Izd-vo «Rubiños – 1860, S.A.», 1565–1569.
- 12- Fedosov O.I. (2002). Peredvigajutsja li granicy frazeologii? (O t.n. «leksicheskikh frazemah» na materiale russkogo, cheshskogo i vengerskogo jazykov), [v:] *Slowo. Tekst. Czas: nowa frazeologia w nowej Europie (Materiały VI Międzynarodowej Konferencji Naukowej, (Szczecin, 6-7 września 2001 r., Greifswald, 8-9 września 2001 r.)*, pod red. M.A. Alekseenko, V.M. Mокienko, G. Val'tera, Shhecín – Greifswald, Izd-vo «Wydawnictwo Naukowe US».

- 13- Fedosov O.I. (2004). Kollokacii na fone kombinatoriki leksem, [v:] *Frazeologia slowianska i inne plaszczyny systemu jazykowego*, pod red. Ja. Bartoshevskoj, V.M. Mokienko, G. Val'tera, Gdan'sk, Izd-vo «Wydawnictwo Uniwersytetu Gdanskiego», 37–49.
- 14- Hlebda V. (2003). Jelementy frazematiki: vvedenie vo frazeologiju adresanta (Chlebda W. *Elementy frazematyki: wprowadzenie do frazeologii nadawcy*), izd. II, Lask, Izd-vo «Oficyna Wydawnicza Leksem».
- 15- *Jenciklopedija obshhego jazykoznanija (Encyklopedia językoznawstwa ogólnego)*. pod red. K. Polan'skogo, (Polański, K.) (1999). Wroclav, Izd-vo «Zakład Narodowy im. Ossolińskich – Wydawnictwo».

HOW TO CITE THIS ARTICLE

Fedorushkov Yury Glebovich (2018). A Phrasematic Key to the Russian Verb-Nominal Constructions: The Workshop of the Computational Phraseography. *Issledovatel'skiy zhurnal russkogo yazyka i literatury*. 12(2):pp:209-225

DOI: 10.29252/iarll.12.209

URL: <http://journaliarll.ir/article-1-150-en.html>



کلید فرازیماتیک برای ساختارهای نامشی-کلامی زبان روسی کارگاه فرهنگ‌نویسی رایانه‌ای اصطلاحات

یوری گلباویچ فیداروشکوف^۱

دانشیار مؤسسه زبان و ادبیات‌شناسی روسی و اوکراینی، دانشگاه میتسکوویچ پازنان،

پازنان، لهستان.

(تاریخ دریافت: مارس ۲۰۱۸؛ تاریخ پذیرش: سپتامبر ۲۰۱۸)

در مقاله حاضر مسئله وضعیت فرازیماتیک گروه‌واژه‌های نامشی-کلامی‌ای چون: عصبانی کردن، سر کار رفتن، برای چند لحظه به جایی سر زدن، رفتن به پشت بام، وارد صفحه اینترنتی شدن، که به‌عنوان «اصطلاحات و ترکیبات بالقوه» به‌شمار می‌روند، بررسی می‌شود. در فهرست مورد مطالعه بیش از هزار نمونه از اینگونه ترکیبات و ساختارهای زبانی وجود دارد. رویکرد مقاله حاضر تأکید بر استقلال و عدم وابستگی اینگونه ترکیبات و ساختارها در متن است. در اینجا رویکرد موردی برای هر ترکیب و ساختار فرازیماتیک اهمیت ویژه‌ای می‌یابد. این امر امکان ایجاد تجربه‌ای جدید را میسر می‌کند که با تکیه بر آن بتوان کار رایانه‌ای کردن تحلیل زبان‌شناختی را محقق ساخت. در مقاله موضوع امکان اتوماسیون تعیین وضعیت فرازیماتیک گروه‌واژه‌های نامشی-کلامی با استفاده از هوش مصنوعی (روش‌های machine learning) مطالعه می‌شود. روش‌های مورد مطالعه امکان تمایز خودکار گروه‌واژه‌های نامشی-کلامی از تعریف وضعیت فرازیماتیک هم در متون گوناگون زبان روسی و هم در مجموعه‌های متون موجود در رایانه را به دست می‌دهند.

واژگان کلیدی: واحد کنایه‌ای زبان، جابجایی، اصطلاح، مؤلفه دستوری، مورفوسیتناکسیس، مؤلفه دستوری سه‌گانه، یادگیری ماشینی، پردازش زبان طبیعی، رایانه.