



ALEKSANDER KIKLEWICZ

 ORCID: <https://orcid.org/0000-0002-6140-6368>

Uniwersytet Warmińsko-Mazurski w Olsztynie

КОРПУСНЫЕ ИССЛЕДОВАНИЯ В СИНТАКСИСЕ: ВОЗМОЖНОСТИ И ОГРАНИЧЕНИЯ¹

CORPUS RESEARCH IN SYNTAX: OPPORTUNITIES AND LIMITATIONS

The aim of the article is to show the possibilities of using internet corpora in syntactic research. Corpus analysis allows for the verification of axiomatically modeled compositions of units, however the command may be material (exact search) or algorithmic (lexical-grammatical search). Corpus tools are primarily useful in determining the frequency of units and their combinations in texts, and the regularity of predetermined features. Based on the frequency, the typological features of different languages can be described in terms of selected parameters. Corpus analysis also makes it possible to check whether a given lexeme valence property is certified or not, and to determine its functionality. The author also shows that the use of the internet corpus in syntactic research has certain limitations. In the case of frequent phenomena, especially in colloquial speech, corpus analysis is effective, but it does not always allow to document phenomena that are less typical and stylistically marked. It is all the more dangerous in the case of less frequent, but regular units in terms of the implementation of systemic oppositions. The corpus should be equally capable of describing the phenomena of speech and the language system, concludes the author. One conclusion is that corpus analysis should be set up with introspection and polling.

ВВЕДЕНИЕ

В последние десятилетия корпусная лингвистика стала одним из наиболее популярных направлений в области языкознания, а обращение к корпусам стало обычной, рядовой практикой лингвистов. Существование больших и постоянно обновляемых интернет-картотек позволяет ускорить исследовательские процедуры,

¹ Автор благодарит своих коллег: проф. Марка Лазинского (Marek Łaziński) из Варшавского университета и Адама Пшепюрковского (Adam Przepiórkowski) из Института переработки информации ПАН, за полезные советы, касающиеся доступа к ресурсам Национального корпуса польского языка.

а также получить более качественную информацию о языковой системе и ее реализации в текстах. В то же время приходится осознавать некоторые ограничения корпусного метода. Корпус представляет собой определенную выборку и не в состоянии охватить все продукты языковой деятельности. Даже если бы мы вообразили максимально объемный корпус, включающий в себя все тексты, написанные на данном языке, в процессе создания корпуса будут появляться новые тексты, а часть элементов, отмеченных в корпусе, станет функционально нерелевантной.

Лингвистика в широком смысле имеет дело с описанием четырех категорий объектов, постулируемых в общей теории систем. В соответствии с концепцией Юрия Урманцева (1978: 10 и др.) существуют четыре параметра, необходимые для описания каждой системы: 1) первичные элементы; 2) их свойства как основания для категоризации; 3) отношения единства, благодаря которым элементы образуют классы; 4) законы композиции, благодаря которым из элементов строятся конструкции. Эмпирическое исследование призвано предоставить исходную информацию о некоторых из этих параметров, а именно — осуществить инвентаризацию языковых элементов и всех отмеченных композиций. Языковой корпус в интернете идеально подходит для решения этих задач, неслучайно он получил широкое употребление в исследованиях синтаксиса. В данной статье обсуждаются некоторые актуальные проблемы синтаксического анализа с использованием корпусных данных. Главное внимание будет посвящено эмпирическим аспектам реализации модели экспликативного синтаксиса, разрабатываемой группой исследователей, относящихся к польской школе семантического синтаксиса. Основная часть иллюстративного материала получена из Национального корпуса русского языка (<https://ruscorpora.ru/>) — ссылки на корпус имеют вид аббревиатуры НКРЯ.

1. КОРПУСНЫЕ ИССЛЕДОВАНИЯ В ОБЛАСТИ СИНТАКСИСА: КРАТКИЙ ОБЗОР ПРОБЛЕМАТИКИ

Роль интернет-корпусов особенно важна в случае аксиоматических описаний, когда на основе ранее принятых постулатов генерируются абстрактные модели, например, типовые конструк-

ции единиц, реальность, а также функциональность которых следует установить посредством наблюдений. Корпус позволяет введение параметров, по которым компьютерная система осуществляет поиск необходимых единиц (токенов). Эти параметры могут быть введены разными способами. Первый из них имеет материальный характер и опирается на языковую интуицию. Речь идет о ситуациях, когда исследователь, благодаря своей языковой компетенции, генерирует некоторые возможные реализации модели — обращение к корпусу в этом случае должно подтвердить существование гипотетических конструкций². К примеру, задача исследователя может касаться обоснования синтаксического статуса глагола *содействовать* как двухместного предиката второго порядка. В четвертом томе *Словаря русского языка* (Евгеньева 1984: 180) в соответствующей статье приводятся два примера:

- (1) Обстоятельство это [...] содействовало его успеху.
- (2) Наступлению наземных войск активно содействовали удары нашей авиации.

Как видим, синтаксические коллокации глагола — абстрактные существительные *обстоятельство, успех, наступление, удары*. Статус предиката второго порядка (за которым стоит выражаемое данным словом понятие) требует, чтобы его пропозициональные аргументы (указывающие на ситуации, события, состояния, процессы и отношения) были выражены наиболее полным, дискретным, изосемическим способом, т.е. в форме придаточного предложения. Если мы постулируем, что в содержании глагола *содействовать* закодирована пропозициональная структура $P(p, q)$, необходимо привести отмеченные формы синтаксических актантов в первой и второй позиции. Поиск-вый параметр может быть определен с учетом знания о том, как оформляются в русском языке придаточные изъяснительные предложения. Рассмотрим это на примере использования опции «Поиск точных форм» в НКРЯ:

- (3) поисковый параметр: [содействовал тому, что]
Александр содействовал тому, что войска обесславили себя вторжением в Испанию (НКРЯ).

² Михаил Котин (Kotin 2018: 63) пишет, что корпусный анализ должен сопровождаться самоанализом как критерием языковой компетенции.

- (4) поисковый параметр: [содействовало то, что]
Этому содействовало то, что Платов [...] был тесно связан с донским казачеством (НКРЯ).

Материальный параметр в случае синтаксического анализа имеет свои ограничения: предложение, кроме известного исследователю глагольного предиката и синтаксических операторов, содержит определенный лексический, трудно прогнозируемый состав. В этом случае точный поиск невозможен и следует прибегнуть к другой форме — лексико-грамматическому поиску, используя проведенную в текстах разметку. Эта опция необходима для поиска предложений, в которых оба пропозициональных аргумента при предикате *содействовать* выражены в форме придаточных предложений. Запрос имеет вид:

[что] [содействовать V,indic] [что]

Компьютерная программа открывает единственный имеющийся в корпусе пример — из прозы Льва Толстого:

- (5) Тому, что Багратион выбран был героем в Москве, содействовало и то, что он не имел связей в Москве, и был чужой. (НКРЯ)

Интерпретация этого факта — это уже дело исследователя. В зависимости от своих теоретических предпосылок он считает возможным трактовать этот факт как подтверждение сочетаемости глагола с сентенциальным актантом или продолжит поиск в других источниках. Например, если в поисковой системе Google вписать поисковый параметр "то что * содействовало тому что", можно получить доступ к единственному предложению:

- (6) То, что Шелеста сняли с постов, содействовало тому, что на рубеже 1980-х — 1990-х население Центральной Украины не имело уже таких эффективных лидеров и руководителей как Шелест³.

Несмотря на то, что конструкция с сентенциальными актантами представлена единичными примерами использования, они вполне соответствуют как литературной норме, так и системе современного русского языка.

³ <https://pandia.ru/text/77/463/20680-3.php> (16.12.2021).

Как видим, интуиция или интроспекция является не только альтернативным источником лингвистической информации по отношению к текстам, в которых языковые единицы и их конструкции считаются отмеченными, — в определенной степени она также влияет на поисковые процедуры при работе с интернет-корпусом (подробнее об этом см.: Meurers 2005). Другими словами, использование корпуса предполагает наличие некоторых предварительных языковых знаний субъекта, что особенно важно в ситуациях, когда корпус располагает ограниченными возможностями автоматизированного поиска. Так, Арто Мустайоки (Mustajoki 2006: 53 и др.) рассматривает возможности корпусного исследования русских безличных предложений типа

Лодку унесло ветром.

В связи с тем, что используемый исследователем корпус Integrum⁴ не содержит грамматической разметки материала, приходится использовать дополнительные инструменты, чтобы сузить поле поиска. Мустайоки принимает во внимание факт, что в безличных предложениях в позиции каузатора выступают определенные глагольные лексемы: *унесло, убило, зарезало, ударило* и др. Указав эти лексемы в поисковом запросе, исследователь смог ограничить поиск необходимых конструкций. В тех же случаях, когда общее количество конструкций с некоторыми единицами было слишком большим, выбирались случайные примеры. Таким образом был получен подкорпус, состоящий из 2304 предложений, которые были экспортированы в программу Excel и подверглись дальнейшему лингво-статистическому анализу.

Разумеется, поиск в корпусе может иметь сутобо алгоритмический характер, как в случаях, когда нам известны свойства матричной конструкции, а обращение к корпусу должно показать их отмеченность или неотмеченность, а также (в случае отмеченности) частоту употребления. Так, в окне синтаксического поиска НКРЯ⁵ мы можем вписать:

[V,сов] сочин [V,несов]
[V,несов] сочин [V,сов].

⁴ <http://www.integrumworld.com>

⁵ <https://ruscorpora.ru/new/search-syntax.html> (16.08.2021)

Такая команда означает, что мы заинтересованы в нахождении сочинительных конструкций с двумя глаголами, один из которых выражает значение совершенного, а другой — значение несовершенного вида. Программа открывает нам языковые выражения, удовлетворяющие этому требованию, например:

(8) Охрана перепугалась, говорит: подходить нельзя.

Vpf Vpf

(9) Он спасся, пытался плыть, ныряя под воду.

Vpf Vpf

(10) Зачем много зарабатывать, стать культурным человеком?

Vpf Vpf

(11) Краны на кухне даже не сипели, превратились в воспоминание.

Vpf Vpf

По данным корпуса, первый тип является более регулярным: он представлен 35 единицами, тогда как второй — только 21 единицей, что дает основание для некоторых первичных обобщений. Важно то, что конструкции, в которых глаголы согласованы по виду, представлены в значительно большем объеме, при этом несовершенный вид употребляется чаще:

[V,несов] сочин [V,несов] — 536 предложений

[V,сов] сочин [V,сов] — 358 предложений

По мнению специалистов (см. Biber, Fitzmaurice, Reppen 2002; Gries 2009; Mustajoki 2006; Münzberg 2011; Marzo, Heylen, Sutter 2012), корпусные исследования чрезвычайно эффективны при изучении функциональности тех или иных синтаксических явлений, в частности, при определении частоты их употребления. Лукаш Пёнтковский (Piątkowski 2018: 133) показывает, что такое исследование может быть проведено при сопоставлении языков, в частности, немецких и польских вербономинальных конструкций типа:

(12) польск. brać udział ‘принять/принимать участие’

(13) польск. przynosić/przynieść efekt ‘дать/давать эффект’

(14) нем. einen Besuch machen ‘навестить/навещать’

(15) нем. Maßnahmen treffen ‘принять/принимать меры’

Благодаря корпусным данным определяются левые и правые коллокаты каждой рассматриваемой конструкции, а их стати-

стическая обработка и обобщение позволяют получить информацию о типовой сочетаемости вербономинальных фразем.

Таблица 1. Типовая схема сочетаемости немецкого словосочетания
Maßnahmen treffen

подлежащее		обстоятельство	определение		определение/ дополнение
Regierung	trifft	bereits rechtzeitig unverzüglich	(die)erfolgreichen (die) notwendigen geeignete	Maßnahmen	zum Schutz zur Sicherheit zur Verbesserung

Таблица 2. Типовая схема сочетаемости польского словосочетания
podjąć kroki

подлежащее		определение	обстоятельство/ определение	
rząd firma	podjmuje podjęła	kroki	prawne zaradcze	w celu w kierunku przeciwko

Применение корпусных данных при сопоставлении синтаксиса языков можно показать на примере русских и польских конструкций с адъективными (в частности указательными) местоимениями. Предварительные наблюдения свидетельствуют, что в польском языке чаще выступает обратный порядок слов, т.е. конструкции, в которых существительное находится в препозиции по отношению к определению. На это указывают некоторые примеры из параллельного корпуса⁶:

- (16) польск. *Bywalców fakt ten* nie zbulwersował jednak (Andrzej Sapkowski).
[перевод] Завсегдагаев *этот факт*, однако, не взволновал и не мешал им.
- (17) *Problem ten*, przez metodologów uznany pospiesznie za metafizyczny, tlił się na dnie wszystkich nieomal dyskusji i sporów (Stanisław Lem).
[перевод] *Эта проблема*, поспешно признанная метафизической, тлела на дне всех дискуссий и споров.

Разумеется, в каждом языке имеются конструкции с прямым и обратным порядком слов, однако в соответствии с нашей гипотезой в польском языке инверсия представляет собой более регулярное явление. Обратимся к НКРЯ и воспользуемся лек-

⁶ <https://processing.ruscorpora.ru/>

сико-грамматическим поиском. В данном случае следует исключить из рассмотрения употребление указательного местоимения *это* в форме среднего рода, ввиду его регулярного использования в позиции связки, т.е. в рамках другого (пролептического) отношения. Вот результат поиска:

[APRO,nom,(m|f)] [S,nom] – 320.497 (84%)
[S,nom] [APRO,nom,(m|f)] – 61.921 (16%)

В Национальном корпусе польского языка (в версии, разработанной Институтом переработки информации ПАН) поиск осуществляется с помощью команд, которые выглядят несколько по-другому:

[pos=subst & case=nom] [orth=»ten|ta|te|ci»]
[orth=»ten|ta|te|ci»] [pos=subst & case=nom]

В сбалансированном корпусе (объемом 300 млн сегментов) имеется 161.323 конструкции с прямым порядком слов (что составляет 65%) и 87.092 конструкции с обратным порядком слов (35%). Как видим, и в польском материале конструкции с прямым порядком доминируют, однако здесь инверсия охватывает 1/3 всех сочетаний, тогда как в русском языке употребительность этого явления в два раза меньше.

2. КОРПУСНЫЙ МЕТОД В МОДЕЛИ ЭКСПЛИКАТИВНОГО СИНТАКСИСА

Концепция экспликативного синтаксиса была разработана в публикациях Станислава Кароляка (Karolak 1984; 2002), а также дополнена в работах его последователей (Kikiewicz, Korytkowska 2010; Kikiewicz, Korytkowska, Mazurkiewicz-Sułkowska et al. 2019). Предметом исследований являются определяемые в терминах грамматических классов синтаксические структуры, рассматриваемые как репрезентации исходных семантических пропозиций, представляющих конфигурации вершинного (матричного) предиката и его компонентов, обусловленных заключенным в предикате лексическим понятием. Одной из форм представления синтаксических репрезентаций являются, например, экспликативные схемы следующего вида:

- (18) $P(x, q) \rightarrow V_p - N_x - S_q$
Людмила Николаевна знала, что один лишь эгоизм движет
поступками Виктора (НКРЯ).

Важнейшая особенность экспликативного синтаксиса, отличающая его от других моделей (например, московской семантической школы), состоит в том, что синтаксические формы упорядочиваются в соответствии с тем, насколько полно они отражают базовую пропозициональную структуру. Вслед за Галиной Золотовой это свойство определяется как изосемичность. В случае реализации пропозициональных аргументов, то есть таких, которые указывают на ситуации, события, состояния, процессы и отношения, в русском языке имеется целый ряд альтернативных форм выражения, более или менее соответствующих характеру пропозиционального аргумента. Например, глагол *знать* употребляется как двухместный предикат с пропозициональным аргументом во второй позиции. Этот аргумент может быть выражен несколькими способами. В НКРЯ возможные формы сочетаемости глагола с его правым коллокатом можно получить, вписав в окне лексико-грамматического поиска команду [знать V,indic]. Приведенный ниже список содержит формы выражения пропозиционального аргумента, в том числе полные, неполные (являющиеся следствием имплицитации) и сверхполные (являющиеся следствием расщепления):

$q \rightarrow S_q^7$	Эраст знает сам, что происходит в душе его.
$q \rightarrow S_q^q S_q$	О том, что произошло вчера, он знает, что это была провокация националистов.
$q \rightarrow NV_{aq} S_q$	О данном событии он знает, что его причины — в эскалации конфликта.
$q \rightarrow V_q NV_q$	О том, что произошло на площади, Федор знает все подробности.
$q \rightarrow NV_q$	Левон Оганезов знает множество историй.
$q \rightarrow NV_{aq} \emptyset_q$	Петя знает правильное решение задачи.
$q \rightarrow NV_{aq} NV_{aq} \emptyset_q$	О данном событии Иван знает все подробности.
$q \rightarrow NV_{aq} NV_{pq} \emptyset_{vq}$	Святитель Василий знал медицину как языческое искусство.
$q \rightarrow N_{aq} \emptyset_q$	Батюшка знает меня.
$q \rightarrow N_{aq} N_{pq} \emptyset_{vq}$	Любители поэзии знают Шульпякова как одного из последних переводчиков.
$q \rightarrow \emptyset_q$	[С Иваном не спорьте] — Иван знает.

Краткие синтаксические формы (например, именная группа в позиции пропозиционального аргумента) более распростра-

⁷ Определения символов см. в работе: Kikiewicz, Korytkowska, Mazurkiewicz-Sułkowska et al. 2019: 13 и др.

нены в речи, особенно в разговорном стиле, тогда как полные, изосемические формы имеют книжный, а иногда и достаточно искусственный характер. Именно в случае таких форм необходима верификация их отмеченности, т.е. подтверждение того, что данный тип экспликации аргумента предусмотрен валентностью глагола и реализуется в речи. Например, если мы обратимся к русскому глаголу *вздумать*, в первую очередь вспоминаются конструкции с инфинитивом. Одна из них — название народной сказки:

(19) Как волк *вздумал* хлеб печь.

$N_x \quad V_P \quad V_{Inf}$

Языковая интуиция, однако, может быть недостаточно объективной при решении вопроса о том, возможна ли в этом случае сочетаемость глагола с сентенциальным актантами. Более надежным способом является обращение к корпусу, в котором можно найти единственный пример требуемой конструкции:

$q \rightarrow S_q$ А жадный купец *вздумал*, что в этой палке заключена чудесная сила.

$N_x \quad V_P \quad S_q$

Этот пример дает повод задуматься о том, насколько обращение к корпусу может помочь в решении проблемы. Приведенное выше предложение, с одной стороны, является лингвистическим фактом, но, с другой стороны, факт этот единичный и к тому же заимствован из художественного текста начала XX века. Чтобы убедиться в том, что такого рода сочетаемость характерна для современного русского языка, следует обратиться к другим источникам. Так, в поисковой системе Google можно найти следующие примеры:

(20) Ты же знаешь, как быстро вокруг нас рухнет всё общество, если Хан *вздумает*, что ему можно безнаказанно меня подкалывать?⁸

(21) Может, *вздумает*, что именно он замена Хави и останется?⁹

(22) И ежели кто *вздумает*, что планы его расходятся с Высшими Целями Гильдии Серых Магов, [...] он может считать себя в состоянии войны с Гильдией¹⁰.

⁸ <https://context.reverso.net/перевод/русский-иврит/вздумает> (16.12.2021).

⁹ <http://192.168.8.1/html/quicksetup.html> (16.12.2021).

¹⁰ <https://mikportal.org/forum/topic2058s75.html?#entry81724> (16.12.2021).

- (23) Разве не безумным сочтут слепца, который *вздумает*, что знает различия цветов, не зная цвета²¹
- (24) Пока не *вздумали*, что мы здесь короли²².
- (25) Дочь конюха и свиначки *вздумала*, что она принцесса²³.

Несмотря на то, что НКРЯ содержит 40 млн словоупотреблений, и в нем представлены тексты разных стилей, в то же время, как констатирует Владимир Плунгян (2005), 40% текстов относятся к художественному стилю. Напротив, участие научных и официально-деловых текстов относительно невелико и, следует помнить, что полные и компрессированные синтаксические формы по-разному представлены в текстах разных стилей (Korytkowska, Małdziejewa 2002: 161; Kiklewicz 2016: 106). Михаил Котин (Kotin 2018: 63 и др.) справедливо отмечает, что, работая с корпусами, даже весьма обширными, исследователи часто озабочены адекватной фиксацией описываемых единиц и типов в их реальной текстовой среде. Также ученый пишет, что необходимо учитывать зависимость между частотой использования единицы и профилем текстов, образующих корпус, что непосредственно влияет на интерпретацию материала. Часто, по мнению Котина, возникают серьезные коллизии между результатами корпусного анализа и «текущим состоянием» в области категоризации языковых форм. Эти расхождения, по его мнению, дополнительно усиливаются из-за того, что недостаточно принимаются во внимание такие «экстракорпоральные» факторы, как, например, внешние условия речевой деятельности.

Указанные ограничения проявляются в процессе синтаксического анализа и, в частности, при верификации экспликативных схем. Рассмотрим в качестве примера глагол *интерпретировать*. С семантической точки зрения, он представляет собой трехместный предикат второго порядка: $P(x, p, q)$. В соответствии со значением глагола некоторое лицо интерпретирует некоторый факт, положение дел и т.д. как некоторый другой (более известный) факт или положение дел, т.е. ставит в соответствие одному пониманию положения дел другое его понимание. В НКРЯ можно найти большое число примеров того, что по-

²¹ http://bookwu.net/book_filosofiya_677/23_tema-13 (16.12.2021).

²² <https://stihi.ru/2013/08/22/5244> (16.12.2021).

²³ <https://litnet.com/ru/reader/kak-byvshaya-princessa-byt-nalazhivala-b359791?c=> (16.12.2021).

зиции p и q реализованы в форме абстрактных существительных (или соответствующих именных групп), например:

$NS_q NS_r$ Он интерпретирует его замысел как попытку проникновения в историю вопроса.

В то же время мы не находим в корпусе ни одного примера того, что пропозициональный аргумент реализуется в форме придаточного предложения. На этом основании можно было бы констатировать, что данный глагол не предусматривает сочетаемость с сентенциальными актантами, хотя такой вывод был бы опрометчивым: характер полученных данных может быть слишком зависим от профиля собранных в корпусе текстов. И действительно, если мы обратимся к другим источникам, например, к поисковым системам интернета, окажется, что конструкции с сентенциальными актантами также существуют. Вот несколько примеров из поисковой системы Rambler.ru:

- $S_q NS_r$ То, что происходит, можно интерпретировать как начало избирательной кампании¹⁴.
- $NS_q S_r$ Совокупность предлагаемых норм можно интерпретировать как то, что для расчета прибыли КИК на основании финансовой отчетности КИК должна располагаться в стране-партнере РФ¹⁵.
- $S_q S_r$ То, что нарисовано на картинке, можно интерпретировать как то, что «тирамису побеждает пончик с желе».
То, что они услышали, он интерпретирует как то, что Европа дает наказ правительству.
То, как ведет себя ЕС, можно интерпретировать как то, что Запад начал некое «тактическое отступление» по украинскому вопросу¹⁶.

О том, что приведенные выше синтаксические формы отражают реальные, закодированные в языковой системе валентные свойства глагола, свидетельствует не только их материальная отмеченность в текстах, но и их совместимость с лингвистической интуицией. В связи с этим можно сослаться на высказывание Казимежа Полянского, который придавал этому источнику лингвистической информации особое значение. В предисловии к дистрибутивно-генеративному словарю польских глаголов он писал:

¹⁴ <https://www.svoboda.org/a/28311562.html> (16.12.2021).

¹⁵ <https://docplayer.com/29940887-Podgotovlena-novaya-versiya-popravok-v-deofshorizacionnoe-zakonodatelstvo.html> (16.12.2021).

¹⁶ Три последних примера из работы: Киклевич 2016: 16.

В соответствии с постулатом порождающей грамматики лингвистика не должна ограничиваться так называемыми текстовыми корпусами — в поисках лингвистических фактов очень часто приходится обращаться к собственной лингвистической интуиции и интуиции других пользователей языка¹⁷ (Polański 1980: 6).

Наиболее эффективным представляется комплексный метод регистрации языкового материала, т.е. с использованием текстовых источников, языковой интуиции, анкетирования, эксперимента, интернет-корпуса и поисковых систем интернета. Это можно показать на примере исследования конструкций русского языка с расщеплением пропозиционального аргумента, т.е. таких, в которых одной аргументной позиции (в семантической структуре) соответствуют два синтаксических актанта (в формально-грамматической структуре), например:

- (26) Он раздражает меня как факт, внезапно нарушивший наш покой (НКРЯ).
 || Меня раздражает то, что он представляет собой факт, внезапно нарушивший наш покой...
- (27) Меня раздражает белый карандаш как невероятная глупость (НКРЯ).
 || Меня раздражает то, что белый карандаш представляет собой невероятную глупость.

Глагол *раздражать* реализует пропозициональную схему $P(p, y)$, которая означает: 'некоторое положение дел вызывает раздражение некоторого лица'. В первом предложении пропозициональный аргумент выражен элементами *он... как факт*, во втором предложении — элементами *белый карандаш... как невероятная глупость*.

Когда возникает вопрос о подобной сочетаемости другого глагола той же лексико-семантической группы — *удивлять*, мы находим в корпусе только несколько примеров из текстов XIX века:

- (28) Но и между произведениями древних, не большая ли часть удивляют нас как памятники высокого духа, а не как образцы совершенного вкуса.
- (29) Как люди, отжившие свой век, они удивляли и забавляли нас своей оригинальностью и разными причудами.

¹⁷ „Zgodnie z postulatem gramatyki generatywnej nie pozwalającym lingwiście ograniczać się do tzw. korpusu tekstowego trzeba w poszukiwaniu faktów językowych bardzo często sięgać do intuicji językowej własnej i innych współużytkowników języka”.

- (30) Как иностранец, он удивляет князя Владимира и его княгиню заморскими подарками.
 (31) Как общественный симптом, — продолжал он, — это меня несколько не удивляет.

Возникает дилемма: дают ли эти несколько примеров текстов 150–200-летней давности основание утверждать, что *удивлять*, как и другие эмотивные глаголы, допускает расщепление пропозиционального аргумента? Лингвистическая интуиция подсказывает нам, что такие конструкции возможны и сегодня, однако, чтобы задокументировать их, нужно обратиться к другим источникам. В поисковой системе Rambler.ru можно найти интересные для нас предложения с глаголом *удивлять*, причем они не являются единичными, ср.

- (32) Он меня удивляет, скорее, как чудо Природы.

$$\begin{matrix} N_p & N_y & & V_p & & & & N_p & \emptyset_p \\ & & & & & & & & \end{matrix}$$
 (33) Меня это удивляет как факт и не удивляет как результат их работы.

$$\begin{matrix} N_y & N_p & & V_p & & N_p & \emptyset_p \\ & & & & & & \end{matrix}$$
 (34) Появление [...] животных уже мало кого удивляет как явление.

$$\begin{matrix} & & & NV_p & & & & N_y & & V_p & & N_p & \emptyset_p \\ & & & & & & & & & & & & \end{matrix}$$
 (35) Твоя энергетика [...] удивляет как феномен!

$$\begin{matrix} & & & NS_p & & & & & & V_p & & \emptyset_y & & & & & N_p & \emptyset_p \\ & & & & & & & & & & & & & & & & & \end{matrix}$$

Таким образом, можно с уверенностью утверждать, что расщепление пропозиционального аргумента (с использованием предложно-падежной конструкции *как* + N_{Nom}) в предложениях с глаголом *удивлять/удивить* в современном русском языке является документированным явлением, и в этом отношении валентность данного глагол соответствует большинству единиц класса *verba sentiendi*.

ЗАКЛЮЧЕНИЕ

Корпусные данные исключительно важны для современных лингвистических исследований, как в области лингвистики языка, так и лингвистики речи. В синтаксических исследованиях обращение к корпусу позволяет верифицировать абстрактные синтаксические модели, документировать синтаксические связи, а также лексические реализации отдельных синтаксических позиций. Используя данный инструмент, мы устанавливаем ча-

стоту употребления тех или иных синтаксических форм, а тем самым и продуктивность синтаксических моделей.

Практика исследований показывает, что у корпусного анализа есть и определенные ограничения. Корпус оказывается недостаточно репрезентативным в случае менее частотных явлений и особенно тех, которые маркированы со стилистической точки зрения. Объективность корпусных данных повышается, если дополнительно используются другие источники материала: интроспекция (наблюдение, рефлексия), анкетирование, эксперимент, качественный и количественный анализ текстов, анализ словарных дефиниций.

Обращение к интроспекции (а также к языковой интуиции) может показаться необоснованным, особенно с учетом факта, что интернет-корпусы содержат миллионы примеров употреблений слов, а среднестатистический пользователь языка знает их только несколько десятков тысяч. Объяснение этого факта кроется в том, что корпусная разметка обычно касается характеристик единиц, и только в случае синтаксических корпусов возможен поиск по их отношениям, но этот поиск, следует признать, еще не совершенен. Языковая же компетенция включает не только единицы и их свойства, но также их системные отношения, включая динамические процессы и возможные состояния системы (в том числе и так называемые потенциализмы).

Из опыта корпусных исследований вытекает общий вывод, что нельзя отождествлять естественный язык ни с каким компендиумом форм его проявления. Так же, как, согласно Протагору, человек — это мера всех вещей, он также является мерой идентичности языковой системы. Понятие языкового факта не ограничивается материальной формой единицы большего или меньшего формата — оно всегда предполагает более или менее социализированное и, в принципе, латентное по способу существования знание носителей языка о контекстах употребления единиц. В связи с этим к идеологии популярного сегодня панэмпиризма (см. Андрухина 1992: 18; Dębowski 2007: 120) следует относиться с осторожностью.

REFERENCES

- Andryukhina, Lyudmila M. *Stil' nauki: kul'turno-istoricheskaya priroda*. Yekaterinburg: Izdatel'stvo Yekaterinburgskogo universiteta, 1992 [Андрухина, Людмила

- M. *Styl nauki: kulturowo-istoryczna natura*. Ekaterinburg: Wydawnictwo Ekaterinburskiego uniwersytetu, 1992.
- Biber, Douglas, and Fitzmaurice, Susan, and Reppen, Randi. Eds.. *Using Corpora to Explore Linguistic Variation*. Amsterdam: Benjamins, 2002.
- Dębowski, Józef. "Wolność nauki a zasada bezzłożeniowości. Między «epistemologicznym anarchizmem» a «epistemologicznym restrykcyjizmem»." *Wolność w epoce poszukiwań*. Szulakiewicz, Marek, and Karpus. Zbigniew Eds. Toruń: Wydawnictwo UMK, 2007: 109–122.
- Gries, Stefan G., "What is Corpus Linguistics?" *Language and Linguistics Compass* 2009, no. 3: 1–17.
- Karolak, Stanisław. "Składnia wyrażen' predykatywnych." *Gramatyka współczesnego języka polskiego. Składnia*. Topolińska, Zuzana. Ed. Warszawa: PWN, 1984: 11–212.
- Karolak, Stanisław. *Podstawowe struktury składniowe języka polskiego*. Warszawa: Instytut Sławiści PAN, 2002.
- Kiklewicz, Aleksander, and Korytkowska, Małgorzata, and Mazurkiewicz-Sułkowska, Julia et al. *Zintegrowany opis semantyczno-syntaktyczny czasowników bułgarskich, polskich i rosyjskich (verba cogitandi i verba sentiendi)*. Część I. Warszawa: Instytut Sławiści PAN & Fundacja Sławiści, 2019.
- Kiklewicz, Aleksander, and Korytkowska, Małgorzata. *Podstawowe struktury zdaniowe współczesnych języków słowiańskich: białoruski, bułgarski, polski*. Olsztyn: CBEW, 2010.
- Kiklewicz, Aleksander. "Sintaksicheskiye kharakteristiki russkikh i pol'skikh internet-forumov (na materiale prostykh i slozhnykh predlozheniy s mental'nymi predikatami)." *Interaktion von Internet und Stilistik, Internet und Stil*. Branko Tosović, and Arno Wonisch Eds. Graz: Universität Graz, 2016: 93–110 [Kiklewicz, Aleksander. "Синтаксические характеристики русских и польских интернет-форумов (на материале простых и сложных предложений с ментальными предикатами)." *Interaktion von Internet und Stilistik, Internet und Stil*. Branko Tosović, and Arno Wonisch. Eds. Graz: Universität Graz, 2016: 93–110].
- Korytkowska, Małgorzata, and Małdziejewa, Wiara. *Od zdania złożonego do zdania pojedynczego. Nominalizacja argumentu propozycjonalnego w języku polskim i bułgarskim*. Toruń: Wydawnictwo UMK, 2002.
- Kotin, Michail L. "Möglichkeiten und Grenzen der korpusbasierten Grammatikforschung." *Biuletyn Polskiego Towarzystwa Językoznawczego* 2018, no. LXXIV: 61–70.
- Marzo, Stefania, Heylen, Kris, Sutter Gert De. "Developments in Corpus-based Contrastive Linguistics." *Corpus Studies in Contrastive Linguistics*. Marzo, Stefania, and Heylen, Kris, and Sutter Gert De. Eds. Amsterdam: Benjamins, 2012: 1–7.
- Meurers, Detmar W. "On the use of electronic corpora for theoretical linguistics: Case studies from the syntax of German." *Lingua* 2005, no. 115/11: 1619–1639.
- Münzberg, Franziska. "Korpusrecherche in der Dudenredaktion: Ein Werkstattbericht." *Grammatik und Korpora 2009. Dritte Internationale Konferenz*. Mannheim, 22.–24.9.2009. Konopka, M., and Kubczak, J. and Jacqueline, C. Mair et al. Eds. Tübingen: Narr Verlag, 2011: 181–197.
- Mustajoki, Arto. "The Integrum Database as a Powerful Tool in Research on Contemporary Russian." *Integrum: tochnye metody i gumanitarnye nauki*. Galina Nikiporets-Takigava. Ed. Moskva: Letniy sad, 2006: 50–75.

- Piątkowski, Łukasz. "Zastosowanie analiz korpusowych w badaniach leksyko-gramatycznych. Perspektywa porównawcza niemiecko-polska." *Zeszyty Naukowe Towarzystwa Doktorantów UJ Nauki Humanistyczne* 2018, no. 21 (2): 131–150.
- Plungyan, Vladimir A. "Zachemmy delayem natsional'nyy korpus russkogo yazyka?" *Otechstvennyye zapiski* 2005, no. 2: 296–308 [Плунгян, Владимир А. "Зачем мы делаем национальный корпус русского языка?" *Отечественные записки* 2005, no. 2: 296–308].
- Polanski, Kazimierz. *Słownik syntaktyczno-generatywny czasowników polskich. 1/A-M*. Wrocław etc.: Ossolineum, 1980.
- Urmantsev, Yuriy A. "Nachala obshchey teorii sistem." *Sistemnyy analiz i nauchnoye znaniye*. Dmitriy P. Gorskiy. Ed. Moskva: Nauka, 1978. 7–41 [Урманцев, Юрий А. "Начала общей теории систем." *Системный анализ и научное знание*. Дмитрий П. Горский. Ed. Москва: Наука, 1978. 7–41].
- Yevgen'yeva, Anastasiya P. *Slovar' russkogo yazyka*. T. 4. Moskva: Nauka, 1984 [Евгеньева, Анастасия П. *Словарь русского языка*. Т. 4. Москва: Наука, 1984].