

КВАНТИТАТИВНЫЕ МЕТОДЫ В ДИАХРОНИЧЕСКИХ КОРПУСНЫХ ИССЛЕДОВАНИЯХ: КОНСТРУКЦИИ С ПРЕДИКАТИВАМИ И ДАТИВНЫМ СУБЪЕКТОМ

Бонч-Осмоловская А. А. (abonch@gmail.com)

НИУ ВШЭ, Москва, Россия

Ключевые слова: дативные субъектные конструкции, предикативы, корпусная лингвистика, иерархическая кластеризация

QUANTITATIVE METHODS IN DIACHRONIC LINGUISTIC STUDIES: THE CASE OF RUSSIAN DATIVE SUBJECTS WITH PREDICATIVES

Anastasia Bonch-Osmolovskaya (abonch@gmail.com)

National Research University Higher School of Economics,
Moscow Russia

The paper aims to demonstrate how quantitative corpus methods used in linguistics research may help to range different realizations of the same phenomenon: the use of dative subjects in predicative and adjective constructions. The core idea of the research is to study the distribution of dative subject constructions with predicative and adjective forms that potentially can be used in such constructions, i.e. aptitude of the construction for explication or omitting the dative subject. While usually the predicates are classified on the basis whether they can potentially be used with dative subject, I study the trends for explicit use of dative (or prepositional beneficiary arguments) among the “dative subject predicates”, and show that the frequency rates of real use of dative subjects can be very different with different predicates. Separate analysis of different morphological forms of the same dative subject lexeme (i.e. adjectives in full and short forms, comparative adjectives and predicatives) shows that they may also exhibit different strategies with explicit dative subjects. Finally data from the 18th and the 21st centuries is compared and hierarchical clustering is used to reveal some diachronic trends.

Key words: dative subject constructions, quantitative methods, corpus linguistics, Russian

1. Введение: теоретические подходы к анализу предикативных конструкций с дативным субъектом¹

Конструкции типа *мне грустно* — с предикативами на -о и дативным субъектом обладают рядом противоречивых свойств, которые делают их исследование весьма непростой задачей. Во-первых, проблемой являются категориальные характеристики этих предикатов. Являются ли они лексически самостоятельными, относящимися к отдельной категории состояния (Щерба 1957, Виноградов 1947) и лишь омонимически совпадающими с соответствующими формам прилагательных и наречий? Или же они являются определенного вида деривацией, предикативными наречиями (Шведова 1980)? Или же они представляют собой несогласованную предикативную форму краткого прилагательного в среднем роде (дефолтный род согласования в русском языке) аналогично тому, как согласуются безличные глаголы (Babby 1974). Второй, безусловно связанной с пониманием категориальной природы предикативов на-о, проблемой является вопрос о продуктивности этих конструкций. Если принять парадигматический анализ Леонарда Бэбби и рассматривать предикативы и прилагательные как одну лексему, выступающую в случае предикативов в специальном мофросинтаксическом контексте, то оказывается, что далеко не все краткие формы прилагательных допускают употребление дативного субъекта. Более того, даже те из предикативов, которые могут употребляться безлично, лишь некоторые свободно присоединяют дативный субъект. Имеется достаточно большая «серая» зона предикативов, которые могут быть употреблены с дативным субъектом лишь при дополнительной контекстуальной поддержке, и при этом такие употребления кажутся сомнительными при изолированной интроспекции. Так, в предложении (1) употребление предикатива *знойно* с дативным субъектом является окказионализмом, который, благодаря контексту, легко интерпретируется, ср. изолированное употребление этой же конструкции в примере (2), которое выглядит существенно более проблематично.

(1) *OK Жара описана хорошо — мне знойно стало и губы пересохли*
(Поиск по «Яндекс-блогам»)

(2) *? Мне стало знойно*

При этом важно отметить, что в целом среди конструкций с дативным субъектом, конструкции с предикативами на -о весьма частотны: как было показано в (Бонч-Осмоловская 2003) предикативы на -о занимают третье по частотности место среди всех конструкций с дативным субъектом после экспериментальных глаголов и модальных предикатов. Важно что, в отличие от экспериментальных и модальных глаголов, списки которых исчисляются примерно

¹ В статье использованы результаты, полученные в ходе выполнения проекта (№14-01-0191), в рамках Программы «Научный фонд НИУ ВШЭ» в 2014-2015 гг.

двумя десятками, предикативов, требующих (или допускающих) дативное кодирование субъекта, на порядок больше. В работе (Циммерлинг 2010) сообщается о списке в 150–200 таких предикатов, и вопрос об их продуктивности остается открытым. Иначе говоря, в случае экспериенциальных глаголов мы можем однозначно сказать, что дативное кодирование субъекта есть словарное свойство предиката, поскольку у нас имеется внятный и законченный список дативных предикатов. С предикативами на-о само составление такого списка и определение того, насколько тот или иной предикат имеет валентность на дативный актанта, является непростою исследовательской задачей.

В исследованиях последних лет было предложено несколько подходов к анализу семантических и синтаксических свойств предикативов на-о, которые так или иначе могут быть связаны с дативной субъектной конструкцией. Следует отметить, что в большинстве работ предикативы рассматривались как формы, парадигматически связанные с краткими и полными прилагательными (ср. подход Бэбби к определению категориальной природы предикативов, приведенный выше). В (Бонч-Осмоловская 2003, 2007) предикативы делятся на классы в зависимости от того, что происходит со вторым не дативным актантами. Такого актанта может вовсе не быть (*мне жарко*), или же он может быть выражен инфинитивом (*мне трудно дышать*), зависимой предикацией (*мне интересно, что ты ответишь*) или согласованной с кратким прилагательным номинативной именной группой (*мне понятен твой ответ*). Было показано, что имеются своего рода кластера предикативов, допускающих один или несколько вариантов оформления второго актанта. Так, в отдельную группу выделяются предикативы, встречающиеся либо без актанта вовсе, либо с инфинитивом. В другую группу попали предикативы, присоединяющие именную группу или предложение. Для предикатива *ясно* доминирующей стратегией является управление зависимым предложением. В работе была высказана гипотеза, что расширение списка предикативов происходит за счет наиболее продуктивного класса, который характеризуется отсутствием другого не дативного актанта.

В (Циммерлинг 2010) центральной идеей, классифицирующей предикативы на -о, является соотнесение дативной предикативной конструкции, которая в работе называется предикатом ситуативного признака LexP с адъективной формой — предикатом, соотнесенным с конкретным референтом. Постулируется три класса основ:

- амбивалентные основы, которые «образуются как предикаты, соотнесенные с конкретным референтом, так и предикаты ситуативного признака» *грустный — мне грустно*
- актантно — поляризованные основы, которые допускают только адъективные формы, а предикаты ситуативного признака (дативные субъектные конструкции с предикативом) имеют существенный сдвиг в семантике (ср. *жирный — мне жирно* в значении «слишком много»)
- ситуативно-поляризованные основы, которые не имеют на синхронном уровне соответствующей адъективной формой, или же кроме сдвига в значении имеется еще и сдвиг ударения (ср. *слабый — мне слабó*)

Циммерлинг вводит формат описания свойств предикативов типа LexP в языке и сравнивает данные русского языка с данными литовского, исландского и чешского, в которых тоже имеются подобные конструкции. Формат включает в себя словообразовательную модель (источник, от которого образуется форма), семантические характеристики дативного актанта, такие, например, как +одушевленность или — генеричность, объем класса LexP и продуктивность класса. Таким образом, фактически среди качественных прилагательных, т.е. таких, которые в принципе могут образовывать краткую и адвербиальные формы, выделяется подкласс амбивалентных основ, которые прежде всего и образуют предикативные конструкции с дативным субъектом.

Следующий существенный шаг в изучении свойств основ качественных прилагательных относительно их возможности употребления в дативных предикативных конструкциях был сделан в (Say 2014). В основе подхода также лежит парадигматическое понимание предикатива как непосредственно связанного с формами прилагательного с той же основой. В работе проводится корпусное исследование частотности употребления дативных субъектов с предикативами, согласованными краткими формами и полными формами прилагательных двенадцати лексем, про которые известно, что они могут использоваться в дативных предикативных конструкциях. В результате формулируются две импликации:

1. Если дативный субъект употребляется вместе с полной формой прилагательного, то он употребляется и с краткой формой. Обратное неверно.
2. Если краткая форма прилагательного может встречаться с дативным субъектом, то и соответствующий предикатив (несогласованная форма) тоже встречается с дативным субъектом. Обратное неверно.

Таким образом, в (Say 2014) формулируется иерархия дативных конструкций:

Предикативы > Краткие формы > Полные формы

Основы качественных прилагательных классифицируются на основании того, какую позицию они занимают в иерархии и, следовательно, образуют 4 класса:

- 1 класс не встречается с дативным субъектом (ср. например, **солнечный мне день*, **мне солнечен день*, **мне солнечно*)
- 2 класс употребляется с дативным субъектом только в конструкциях с предикативами (**грустная мне песня*, **мне грустна песня*, но ^{OK}*мне грустно*)
- Основы 3 класса встречаются с предикативами и краткими формами, но не с полными формами прилагательных (**тесный мне пиджак*, ^{OK}*мне тесен пиджак*, ^{OK}*мне тесно*)
- 4 класс допускает дативный субъект в любой из форм (^{OK}*приятный мне человек*, ^{OK}*мне он приятен*, ^{OK}*мне приятно*)

Далее, анализируя семантические отношения между дативным аргументом и аргументной структурой соответствующих прилагательных, автор

показывает, что разные позиции на иерархии соотносятся с разными типами отношений. Датив может быть вообще никак не связан с аргументами прилагательного (тип 1; напр. *холодно*), или может быть соотнесен с внутренним аргументом прилагательного (тип 2; напр. *грустно*), или же непосредственно связан с дативным аргументом прилагательного (тип 3; напр. *приятно*).

Изолированно каждый из таких типов может быть соотнесен с одним из теоретических подходов: тип 1 лучше всего описывается конструкционной моделью, именно поэтому этот тип является наиболее продуктивным и допускает образование контекстуально-обусловленных предложений с дативным субъектом, описывающих непосредственную реакцию экспериенцера (ср. пример(1)). Идея состоит в том, что конструкция PRED+NP_{ДАТ} имплицитно означает значение состояния экспериенцера, и сама возможность интерпретации лексических свойств предикатива как экспериенциального состояния обуславливается не столько его собственно семантикой, но контекстом описываемой ситуации в целом.

Для анализа предикативов второго типа наиболее близкой является деривационная интерпретация «дативных имперсональных конструкций» (см. Guiraud-Weber 1984). Фактически тот же самый деривационный процесс имеет место в деагентивных деривациях «я не сплю > мне не спится»

Наконец, третий тип наиболее четко укладывается парадигматическую концепцию Леонарда Бэбби: в этом случае предикативная конструкция с дативным субъектом есть ни что иное, как безличное употребление краткого прилагательного с сохранением его актантной структуры, в частности дативного аргумента.

Примирение трех подходов заключается в более абстрактном понимании конструкционной схемы, которая, по выражению автора, будучи «слепа к аргументной структуре соответствующего прилагательного», использует уже имеющийся датив для выражения экспериенцера, или же понижает внутренний аргумент до дативного слота, или же превносит его как дополнительный конструкционный элемент.

Еще одним важным замечанием, сделанным в выводах рассматриваемой работы, является методологическое наблюдение о зависимости теоретических моделей от привлекаемых данных. Как отмечает автор, расширение объема данных позволяет перейти от конкуренции моделей в сторону их контаминации.

Настоящая работа во многом развивает метод работы с данными, предложенный в (Say 2014) и основанный на обобщении корпусных наблюдений о поведении дативного субъектного аргумента в контексте соотнесенных форм предикатива (безличного употребления), краткой формы и полной формы прилагательного. При этом, в отличие от работы (Say 2014), в работе будут рассматриваться не все предикативы, а лишь такие, которые потенциально могут допускать конструкции с дативным субъектом во всех парадигматически связанных формах: обеих адъективных и собственно предикативной (описанные как тип 4 в (Say 2014)). Основная задача исследования видится в том, чтобы уйти от бинарного описания возможности/невозможности употребления дативного субъекта к более актуальной на сегодняшний день вероятностной оценке

«склонности» или «тяготения» (Stefanowitsch, Gries 2009) конструкции к эксплицитному дативному маркированию субъектного аргумента. Такой подход, во-первых, позволяет построить иерархию выраженной «дативной субъектности» внутри группы предикатов, которые в принципе допускают такие конструкции, а, во-вторых, дает возможность для диахронических исследований постепенных смещений конструкций в сторону одного или другого полюса.

Таким образом, фокус работы состоит в методологическом подходе получения новых данных о дативных предикативных конструкциях. Работа построена следующим образом: в части 2 будут выделены значимые параметры, на основании которых отбираются исследуемые данные, также будут описаны методы сбора и анализа данных, в части 3 будут рассмотрены результаты анализа данных с помощью метода иерархической кластеризации и предложена интерпретация полученных классов. Там же будут подведены предварительные итоги полученным наблюдениям и сформулированы задачи дальнейших этапов исследования.

2. Методы сбора и анализа данных

Основной концептуальной задачей исследования являлся уход от бинарного подхода к анализу предикативных конструкций (может или не может использоваться с дативом) и переход к анализу трендов — статистически значимых изменений, выражающихся в частотности употребления дативного субъекта. Неудовлетворительность бинарного подхода связана в первую очередь с самим характером данных, а именно наличием контекстуально-обусловленных конструкций, которые, как уже было сказано выше, не могут считаться приемлемыми при изолированном употреблении, однако встречаются в языке как спорадические новообразования.

Как показано в (Бонч-Осмоловская 2003), доля таких окказиональных конструкций в общей выборке предикативов с дативным субъектом весьма существенна. Кроме того, представляется чрезвычайно важным включить в поле рассмотрения диахронические данные, позволяющие проследить значимые изменения.

Исследование проводилось на материале Национального корпуса русского языка. Для исследования были выбраны десять частотных основ: *нужный, надобный, полезный, приятный, жаркий, слышный, понятный, плохой, худой и известный*. Важным свойством этих основ является то, что, во-первых, они достаточно распространены и хорошо представлены в корпусе. Во-вторых, за исключением основы *жаркий*, все они теоретически допускают употребление дативного субъекта с обеими адъективными и предикативной формами. В-третьих, они характеризуют разные стороны внутреннего состояния субъекта. Следует отметить, что наравне с дативным маркированием субъекта рассматривались и конструкции, в которых в этой же функции используется предложная группа с *для*, например, *полезный для меня*. Таким образом, исследовалось два вопроса: во-первых, оценивалась доля конструкций с дативным

субъектом для той или иной морфосинтаксической формы лексемы из списка, а во-вторых, были собраны данные о возможной конкуренции предложной и дативной конструкции для каждой такой формы.

Собственно материал исследования был представлен двумя выборками предложений, первая из которых была получена из текстов 18 века, а другая — из текстов 21 века². Общий объем данных составил около 15 тысяч предложений. Такой значительный объем данных был необходим для надежных статистических подсчетов поведения 80 форм (по 4 формы каждой лексемы для каждого века).

Выбранный подход отличается от метода, используемого в Say 2014, в котором используется анализ 100 случайных употреблений для каждого из 12 предикатов. (О недостатках такого подхода см. Goldberg 2011). В настоящем исследовании каждая основа была представлена совокупностью вхождений конкретных форм в выборке текстов 18 или 21 века.

Полученные предложения были размечены вручную по ряду параметров. К каждому предложению была приписана базовая основа, была указана конкретная форма (полная или краткая согласованная адъективная, сравнительная степень³ или же безличная предикативная), далее отмечалось наличие эксплицитно выраженного датива, а также наличие предложной группы с бенефициарным значением (в абсолютном большинстве случаев последние два параметра находились в дополнительной дистрибуции). Наконец, каждое предложение было также маркировано по принадлежности к выборке 18-го или же 21-го века.

Идея состояла в том, чтобы посмотреть на двух разных временных срезах, как соотносятся разные формы одной основы относительно дативного субъектного аргумента и предложного бенефактива. Поэтому основанием для сравнения стали подмножества данных, характеризующиеся пересечением трех общих параметров: общая лексема, один и тот же тип морфологической формы, общий диахронический показатель (век). Такого рода тройки параметров мы в дальнейшем будем называть типами. Для каждого типа (основа+форма+век) были подсчитано общее число вхождений, число вхождений с выраженным дативным субъектом, число вхождений с предложной бенефактивной группой. Абсолютные числа были нормализованы в проценты от общего числа вхождений. Всего мы получили 59 типов с данными, поскольку некоторые типы оказались вообще не представлены в корпусе (например, удобный в 21 веке), некоторые представлены слишком слабо для статистически надежных вычислений. Каждый из 59 типов характеризовался 3 числовыми значениями: количество предложений, в которых отсутствует экспериментальный аргумент, количество предложений, в котором он выражен дативом,

² Для того, чтобы выборки были соразмерными, 21 век был ограничен только текстами, датируемыми 2003 годом.

³ В целом формы со сравнительной степени имеют мало данных, поэтому, несмотря на то, что они будут присутствовать на рисунке 1, отражающем иерархическую кластеризацию всех форм всех лексем обоих веков, эти формы пока не будут учитываться в результатах.

количество предложений, в котором он выражен бенефактивной предложной группой, в основном с помощью предлога *для*. В таблице 1 представлены для примера данные для типов, связанных с лексемой *нужный*. Колонка Total показывает общее количество вхождений этого типа в выборку, колонка None отражает количество предложений, в котором рассматриваемый тип был употреблен без выраженного дативного субъекта или бенефактивной предложной группы с предлогом *для*. Колонки Dat и PP соответственно отражают количество предложений с дативным субъектом или же с предложной бенефактивной группой. Следующие три колонки содержат те же данные, но выраженные в процентном соотношении. Даже из этого небольшого фрагмента таблицы видно и то, насколько по-разному ведут себя в отношении дативного субъектного аргумента разные морфологические формы одной и той же лексем, см., например, употребления предикатива *нужно* и прилагательного *нужный* в 18 веке с выраженным дативным субъектом — в первом случае количество таких предложений составляет 23% от общего количества вхождений типа, во втором случае только 5%. Очевидно также и то, что склонность к выражению дативного субъекта меняется со временем. Так, та же форма полного прилагательного *нужный* в 18 веке предпочитала конструкции с предложной группы для выражения субъектных отношений (*нужный для меня*) — таких конструкций обнаруживаем 15% от вхождений, а дативных только 5%. В 21 веке ситуация меняется противоположным образом, конструкций с выраженным дативным субъектом (*нужный мне*) становится уже 19% от общего количества вхождений типа, а конструкции с предлогом *для* становятся весьма редкими и составляют всего лишь 2%.

Таблица 1. Пример распределения разных конструкций внутри вхождения одного типа

TYPE	Total	None	Dat	PP	None%	Dat%	PP%
НУЖНЫЙ_18_ПРЕД	184	115	42	27	63	23	15
НУЖНЫЙ_18_КР_ПРИЛ	346	151	92	103	44	27	30
НУЖНЫЙ_18_ПОЛН_ПРИЛ	528	422	26	80	80	5	15
НУЖНЫЙ_21_ПРЕД	1443	1124	277	42	78	19	3
НУЖНЫЙ_21_КР_ПРИЛ	1255	405	752	98	32	60	8
НУЖНЫЙ_21_ПОЛН_ПРИЛ	330	258	64	8	78	19	2

Далее для анализа таблицы был применен метод иерархической кластеризации данных. С помощью этого метода наиболее близкие по поведению типы объединяются в группы, далее близкие группы объединяются в более общие группы. В результате все данные можно представить с помощью иерархического дерева (см. рисунок 1).

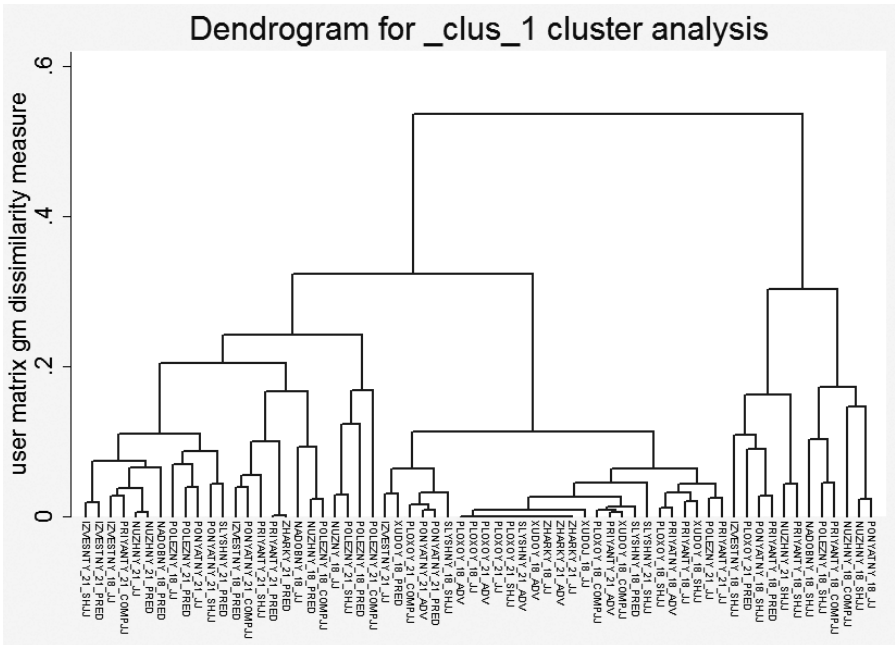


Рисунок 1. Дендрограмма близости типов основ относительно способов выражения экспериенциального элемента. Условные обозначения типов: JJ — полная форма прилагательного, SHJJ — краткая форма прилагательного, COMPJJ — сравнительная степень, PRED — предикатив, не управляющий номинативной ИГ и несогласованный с ней.

3. Результаты и обсуждение

Результаты кластеризации были интерпретированы вплоть до третьего уровня ветвления дендрограммы. Получившиеся в итоге кластеры были сопоставлены с содержательными параметрами классов. Наполнение классов и параметры их выделения отражены в таблице 2. Ниже будут представлены краткие комментарии к анализу дендограммы.

Верхний уровень кластеризации разделяет все типы на те, у которых выраженный дативом экспериенциер или бенефактивная предложная группа встречается в более 40% случаев (экспериенциальный класс), и те типы, для которых частотность таких конструкций ниже (неэкспериенциальный класс). На втором уровне кластеризации в экспериенциальном классе выделяется подгруппа типов, для которых достаточно часто встречается и дативное выражение субъекта, и предложная группа с *для*. Отличительным признаком этой подгруппы является то, что у каждого из типов имеется больше 10% вхождений именно с предложной конструкцией. Такую подгруппу можно условно

называть предложным классом. Эта же подгруппа далее может быть разбита еще на две подгруппы — на собственно предложную, куда входит типы, у которых предложных конструкций больше, чем дативных, и смешанную — такую, в котором количество предложных дативных вхождений для типов представлено примерно поровну. Обратим внимание на то, что обе подгруппы предложного класса содержат в себе только типы, относящиеся к 18 веку. Иными словами, с помощью иерархической кластеризации выделяется, с одной стороны, группа предикатов, чье поведение кардинально изменилось в течение последующих двух веков, а с другой стороны, собственно класс поведения предикатов, который оказывается не представлен в 21 веке, по крайней мере, среди рассмотренных предикатов.

Другая подгруппа экспериенциального класса характеризуется использованием дативного маркирования для оформления экспериенцера, она может быть в свою очередь поделена на такую, в которой дативные конструкции доминируют и встречаются больше чем в половине всех вхождений типа (такой класс условно назовем супердативным), а также на группу, в которой дативные конструкции тоже составляют значительную часть вхождений типа, но все-таки чуть меньшую, чем в супердативной. Интересно, что в этой группе присутствуют и типы, относящиеся к 18 веку, и типы, относящиеся к 21 веку. Так же, как и предложный классы, дативный классы не очень многочисленны: и предложный класс, и дативный насчитывают по шесть типов.

Неэкспериенциальный класс на следующем уровне ветвления дендограммы подразделяется на группу, в которой выраженный экспериенциальный (или бенефактивный) аргумент все-таки встречается, и группу (так называемый смешанный класс), в которой экспериенциальное или бенефактивное отношение выражается совсем редко (так называемый слабый класс). Смешанный класс на следующем уровне дифференцирует типы по частоте встречаемости предложной группы. Заметим, что самым многочисленным классом среди всех кластеров является смешанный дативный класс — такой, в котором количество конструкций с дативным субъектом или предложной группой составляет менее 20%, но больше 10% от общего количества вхождений, и при этом дативные конструкции доминируют. В этот класс входят типы 18 и 21 века, всего к смешанному дативному классу относится 21 тип. При этом лишь типы *известно* и *нужно* представлены и вхождениями 18-го века, и вхождениями 21 века, иначе говоря, только эти два предиката сохраняют свое поведение в диахронии. Все остальные типы попадают в смешанный дативный тип только в 21 веке, или же как, например, *надобный* и *надобно* вовсе выходят из употребления. Ниже в таблице 2 представлены подробные параметры выделения каждого класса, и приведены все типы, которые попадают в тот или иной класс.

Таблица 2. Дифференциация типов предикатов с помощью иерархической кластеризации: основания для выделения классов и примеры вхождения типов в каждый класс

Класс 1	Экспериментальный (Exp > 40%)	Предложный (PP > 20%)	Экспериментальный пред- ложный класс PP > DAT	18: нужен, нужнее, понятный
Класс 2	Экспериментальный (Exp > 40%)	Предложный (PP > 10%)	Экспериментальный смешанный PP DAT	18: надобен, полезен, приятнее
Класс 3	Экспериментальный (Exp > 40%)	Дативный класс (Dat > 20%)	Супердативный DAT > 55%	18: приятен 21: нужен
Класс 4	Экспериментальный (Exp > 40%)	Дативный класс (Dat > 20%)	Дативный 40% < DAT < 55%	18: известен, понятен, приятен 21: плохо
Класс 5	Неэкспериментальный (Exp < 40%)	Смешанный (20% < Exp < 10%)	Смешанный предложный PP > DAT	18: нужный, полезно 21: полезен, полезнее
Класс 6	Неэкспериментальный (Exp < 40%)	Смешанный (20% < Exp < 10%)	Смешанный дативный DAT > PP	18: известный, известно, надобный, надобно, нужно, полезный, полезнее, 21: жарко, известен, известно, нужный, нужно, полезно, понятнее, понятный, понятен, приятнее, приятно, приятен, слышно, жарко
Класс 7	Неэкспериментальный (Exp < 40%)	Слабый (Exp < 20%)	слабый DAT > 10%	18: слышен, худо 21: известный, хуже
Класс 8	Неэкспериментальный (Exp < 40%)	Слабый (Exp < 20%)	нулевой DAT < 10%	18: жаркий, плохой, плох, приятный, слышный, слышно, худой, худ, хуже, 21: жаркий, плохой, плох, полезный, приятный, слышен

Получив сведения о диахронических изменениях частотности употребления дативной конструкции, мы можем проверить, насколько полученные нами данные соответствуют иерархии, сформулированной в Say 2014,

или — иначе — работает ли иерархия не на бинарном уровне категорического противопоставления (может/не может использовать дативный субъект), а на градуальном уровне противопоставления, который мы получаем благодаря делению предикатов на множество классов. Мы ожидаем, что если количество вхождений с выраженным дативным субъектом увеличивается у полной формы, то тогда, это количество будет увеличиваться у краткой формы и у предикатива. Противоречить иерархии будет ситуация, когда количество конструкций с выраженным дативным субъектом у более низкой позиции (предикатива) понижается, а у более высокой позиции (напр. краткое прилагательное) увеличивается. Представим данные таблицы 2 в виде отдельной таблицы, отражающей изменения каждой из форм: курсивом будет отображаться употребление 18 века, а подчеркиванием употребление 21 века. При этом заметим, что иерархия, предложенная в Say 2014, рассматривала исключительно дативные аргументы, а не бенефактивные предложные группы. Таким образом, мы также можем проверить, работает ли иерархия относительно употребления предложных групп с предлогом *для*. В таблице 3 представлена динамика изменений принадлежности к классам типов, относящихся к одной лексеме

Таблица 3. Диахронические смещения по принадлежности к классам морфологических типов лексем. Курсивом выделены типы, относящиеся к 18-му веку, а подчеркиванием к 21-му. В таблице приведены краткие обозначения классов, полные обозначения и обоснования их выделения см. в таблице 2.

экспериментальный				неэкспериментальный			
предложный		дативный		смешанный		слабый	
предлож	смеш	супер-датив	датив	предлож	датив	слабый	нулевой
1	2	3	4	5	6	7	8
<i>понятный</i>			<i>понятен</i>		<u>понятный</u> <u>понятен</u>	<i>понятно</i>	
<i>нужен</i>	<u>нужен</u>			<i>нужный</i>	<u>нужный</u> <u>нужно</u>		
			<i>известен</i>		<u>известный</u> <u>известен</u> <u>известно</u> <u>известно</u>	<i>известный</i>	
	<i>полезен</i>			<u>полезен</u> <u>полезно</u>	<i>полезный</i> <u>полезно</u>		<i>полезный</i>
			<i>приятен</i> <i>приятно</i>		<u>приятен</u> <u>приятно</u>		<i>приятный</i> <u>приятный</u>
					<u>жарко</u>		<i>жаркий</i> <u>жаркий</u>

экспериментальный				неэкспериментальный			
предложный		дативный		смешанный		слабый	
предлож	смеш	супер-датив	датив	предлож	датив	слабый	нулевой
1	2	3	4	5	6	7	8
			<u>плохо</u>				<u>плохой</u> <u>плохой</u> <u>плох</u> <u>плох</u>
					<u>слышно</u>		<u>слышен</u> <u>слышен</u> <u>слышно</u>

Столбики с классами в таблице упорядочены в последовательности уменьшения доли дативных субъектных конструкций. Следуя предсказаниям иерархии, предложенной в Say 2014, мы ожидаем, что любая краткая форма прилагательного будет расположена в таблице не правее полной формы прилагательного той же лексемы, а предикатив той же лексемы будет занимать позицию не правее краткой формы прилагательного. Можно заметить, что в целом это предсказание соблюдается. Однако необходимо сделать несколько оговорок.

Во-первых, особым образом ведет себя предикатив *понятно* в 21 веке: дативные конструкции с этим предикативом встречаются реже, чем с полной и краткой формой прилагательного. У нас слишком мало данных об предикативном безличном употреблении соответствующей лексемы в 18 веке, поэтому в итоговую выборку вхождения *понятно* в 18 веке не вошли. Иными словами, мы не можем посмотреть динамику изменения поведения этого предикатива, но можем только зафиксировать текущее положение вещей, нарушающее иерархию. Возможное объяснение тут состоит в распространении употребления предикатива *понятно*, как вводного слова, не имеющего собственной аргументной структуры:

- (3) *А в «Терре» запросто, без всяких этих штучек, так что она мне как раз подходит. Тебе-то, **понятно**, там не нравится. Тебе бы как раз в эти ваши клубы ходить...* [Анна Берсенева. Полет над разлукой (2003–2005), ruscorpora.ru]
- (4) *Плата за отопление, воду, канализацию и электричество в эту сумму, **понятно**, не входит: все это жильцы получают не от своего товарищества, а от городских служб — им и платят.* [Владимир Абгафоров. Плата за проживание (2003) // «Мир & Дом. City», 2003.04.15, ruscorpora.ru]

Далее, поведение, противоречащее иерархии, обнаруживает предикатив *нужно*, который расположен правее краткого прилагательного *нужен*. Опять же у нас недостаточно данных для понимания того, к какому классу относился бы *нужно* в 18 веке, и мы не можем проследить динамические изменения в поведении этой формы. Заметим, впрочем, что семантически близкий предикатив *надобно*, используемый в 18 веке в том же контексте, а потом вытесненный *нужно*,

относится к тому же самому, что и *нужно*, популярному классу неэкспериментальных дативных типов.

Наконец, краткое прилагательное *полезен* показывает в 21 веке существенную относительно других форм этой леммы долю конструкций с предложной группой. При этом, иерархия выражения дативного субъекта нарушается и в 18 веке. Предикатив *полезно* в 18 веке имеет меньшую долю дативных конструкций по сравнению с соответствующим кратким прилагательным, у *полезно* количество таких вхождений равно 16%, а у предикатива *полезен* достигает 32% от общего количества вхождений этого типа. В 21 веке, *полезно* встречается с дативным субъектом в 11% случаев, а *полезен* лишь в 7%, однако в 14% случаев с предложной бенефактивной группой. Таким образом, уменьшение дативных субъектных конструкций отчасти компенсируется распространением бенефактивно-предложного оформления, но только в контексте кратких прилагательных.

Отметим также, что сдвиг большого количества типов в сторону смешанного дативного класса, и тем более существенное расширение использования предикативов, таких как *слышно* или *жарко*, указывает на то, что мы имеем дело с продуктивной лексико-грамматической конструкцией. Следует уточнить, что под конструкцией, в данном случае, имеет смысл понимать не схематическое описание Dat+Pred, но скорее совокупность распределения выраженного и не выраженного датива.

Итак, в работе был опробован метод иерархической кластеризации на диахронических данных. Такой метод ориентирован на изучение языкового употребления и позволяет фиксировать изменения в этом употреблении, которые в дальнейшем ведут к более существенным лексическим и грамматическим сдвигам. Анализ получившихся классов иерархической кластеризации данных по экспериментальному маркированию субъекта (дативного и предложного) показал, с одной стороны, очевидные тенденции к унификации использования дативной конструкции, а с другой стороны, выделил случаи, которые по каким-то причинам не укладываются в направление общего сдвига. Накопление и анализ более значительного количества данных позволит в дальнейшем генерализовать предложенные выводы.

Литература

1. Бонч-Осмоловская А. А. Конструкции с дативным субъектом в русском языке, Диссертация на соискание ученой степени кандидата фил. наук, Москва, 2003.
2. Бонч-Осмоловская А. А. Конструкции с дативным субъектом с предикативами на -о/-е, Киселева, Плунгян, Татевосов (ред), Корпусные исследования по русской грамматике, Москва, 2009.
3. Виноградов, В. В. Русский язык (грамматическое учение о слове). Москва — Ленинград, 1947.

4. Циммерлинг А. В. Именные предикативы и дативные предложения в европейских языках // Компьютерная лингвистика и интеллектуальные технологии, По материалам ежегодной Международной конференции Диалог. — 2010.
5. Шведова Н. Ю. (ред.) Русская грамматика, Москва, 1980.
6. Щерба Л. В. О частях речи в русском языке // Избранные работы по русскому языку. — 1957. — С. 63–84.

References

1. Babby, L. H. 1974. Towards a formal theory of “parts of speech”. In *Slavic transformational syntax*. Brecht, R. & Chvany, C. (Eds). Ann Arbor: Univ. of Michigan. 151–181.
2. Bonch-Osmolovskaya A. A. Dative subject constructions in Russian (*Konstruktsii s dativnym subyektom v russkom yazyke*), Dissertation, Moscow, 2003.
3. Bonch-Osmolovskaya A. A. 2009 Dative subject constructions with predicatives on -o/-e (*Dativnye subyektnye konstruktsii s predikativami na -o/-e*), Kiseleva, Plungyan, Rakhilina, Tatevosov (eds) *Corpus studies in Russian grammar (Korpusnye issledovaniya po russkoy grammatike)*, Moscow, 2009.
4. Goldberg A. E. Corpus evidence of the viability of statistical preemption // *Cognitive Linguistics*. — 2011. — Т. 22. — № 1. — С. 131–153.
5. Say S. On the nature of dative arguments in Russian constructions with “predicatives” // *Current Studies in Slavic Linguistics [Studies in Language Companion Series, 146]*. Amsterdam: John Benjamins. — 2013. — С. 225–245.
6. Shcherba L. V. On parts of speech in Russian, *Selected works on Russian language*, Leningrad, 1957, pp. 63–84.
7. Shvedova N. Yu. (ed) *Russian Grammar*, Moscow, 1980.
8. Stefanowitsch A., Gries S. T. Collocations: Investigating the interaction of words and constructions // *International journal of corpus linguistics*. — 2003. — Т. 8. — № 2. — С. 209–243.
9. Vinogradov V. V. *The Russian Language (grammatical theory about word) (Russkiy Yazyk (grammaticheskoe ucheniye o slove))*, Moscow-Leningrad, 1947.
10. Zimmerling A. V. Nominal predicatives and dative clauses in European languages (*Imennye predikativy i dativnye predlozheniya v evropejskikh yazykakh*) *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog 2010” [Komp’yuternaya Lingvistika i Intellektual’nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii “Dialog 2010”]*, Bekasovo.