

Grammatical variation: methods for quantitative corpus analysis (the German preposition *wegen*)

E.B. Krotova

Institute of Linguistics, RAS
Moscow, Russia
elena_krotova@inbox.ru

E.N. Tsvetaeva

Moscow State Linguistic University
Moscow, Russia
elenatsvetaeva@yandex.ru

A.V. Sharandin

Russian State University
for the Humanities
Moscow, Russia
sharandin@gmail.com

D.O. Dobrovol'skij

Russian Language Institute, RAS, Institute
of Linguistics, RAS, Moscow, Russia
Stockholm University, Stockholm, Sweden
dobrovol'skij@gmail.com

Abstract

The corpus study of grammatical variation is not considered a complex task anymore. However, the corpus annotations used to solve the task are based on specific theoretical assumptions which do not necessarily cover all possible variants found in empirical material. On the example of the German preposition *wegen* and its variation in case government this article shows why available tools in corpus managers can be insufficient. An experiment is presented in which data obtained from annotated corpora are automatically labeled using a developed algorithm. An analysis of the corpus material showed that the preposition *wegen* primarily governs the genitive (about 60% of all uses). The dative, which is recognized by modern linguistics as a conventional variant, is relatively rare (5%). Additionally, the analysis also revealed that for 35% of nouns following *wegen* case cannot be determined, a phenomenon which is not mentioned in the corresponding dictionary entries of all lexicographic sources known to us. Thus it is not possible to identify the proportion of such underdetermined instances of case government using only the existing annotation of the analyzed corpora, since it did not include tags for nouns with no clear case assignment.

Keywords: Variation, grammatical norm, corpus analysis, German

DOI: 10.28995/2075-7182-2022-21-349-357

Вариативность грамматической нормы: возможности корпусного и количественного анализа (на материале немецкого предлога *wegen*)

Е.Б. Кротова

ИЯз РАН
Москва, Россия
elena_krotova@inbox.ru

Е.Н. Цветаева

МГЛУ
Москва, Россия
elenatsvetaeva@yandex.ru

А.В. Шарандин

РГГУ
Москва, Россия
sharandin@gmail.com

Д.О. Добровольский

ИРЯ РАН, ИЯз РАН, Москва, Россия
Стокгольмский университет,
Стокгольм, Швеция
dobrovol'skij@gmail.com

Аннотация

Изучение вариативности грамматической нормы с использованием корпусных данных в современных условиях не является сложной задачей, однако разметка корпуса опирается на определенные лингвистические теории и не всегда в должной мере учитывает эмпирический материал.

В статье на примере немецкого предлога *wegen* 'из-за' и вариативности его управления показано, почему имеющиеся средства в корпусных менеджерах оказываются недостаточными; представлен эксперимент, в котором данные, полученные из корпусов, автоматически размечаются с помощью разработанного алгоритма. Анализ корпусного материала показал, что генитив остается базовым управлением предлога *wegen* (около 60% от всех употреблений), а датив, признаваемый современной лингвистикой узаконенным вариантом управления, встречается относительно редко (5%). Анализ выявил, что в 35% случаев у существительного после *wegen* отсутствует указание на падеж, что не упоминается в соответствующих словарных статьях всех известных лексико-графических источников. Выявить долю подобных случаев, пользуясь стандартной разметкой рассмотренных корпусов, не представлялось возможным, так как разметка не предполагала категории беспадежного управления.

Ключевые слова: вариативность, грамматическая норма, корпусный анализ, немецкий язык

1 Введение

Вариативность грамматической нормы свойственна практически всем языкам, один из существующих вариантов может признаваться кодифицированной нормой, а другой – (еще) нет. При этом следует иметь в виду, что норма – понятие динамическое, хотя и более консервативное, чем узус. Смена узуса с течением времени приводит к изменению нормы или к ее модификации в том смысле, что нормативными признаются несколько вариантов. Т.е. понятия универсальной нормы не существует (Hennig and Müller, 2009).

Если говорить о немецкой грамматической норме, то здесь существуют два взгляда на норму: 1) дескриптивная норма, ориентированная на реальное употребление языка, в том числе на институциональных уровнях, и 2) прескриптивная норма, предписанная регулирующими институтами.

Важно, что в Германии отсутствует некая инстанция нормы на государственном уровне¹. Исключение составляет только „Rat der deutschen Rechtschreibung“, занимающийся исключительно вопросами правописания. В сомнительных случаях носитель языка обращается к авторитетным словарям или справочникам, например, Duden online², DWDS³, Дуден-9 и Вариг-5 (Duden 9, 2021; Wahrig, 2003).

По этой причине в сфере грамматической нормы и ее вариантов на сегодняшний день наблюдается ряд трудно решаемых проблем. Так, свойственная многим языкам – и немецкий здесь не исключение – вариативность падежного управления предлогов затрудняет процесс однозначного определения языковой нормы. Отдельные падежи (как например, родительный/генитив и дательный/датив в немецком), управляемые большим количеством предлогов и конкурирующие друг с другом на протяжении всего исторического развития, представляют собой особую трудность. Определение единой нормы как таковой для них вряд ли вообще возможно. Однако вполне реально определить допустимые варианты нормы с учетом количественных данных, сложившихся под воздействием систематических исторических процессов и определяющих не только и не столько диалектальное, сколько литературное варьирование падежного управления.

На сегодняшнем этапе развития информационных технологий решение обсуждаемой проблемы становится возможным благодаря, с одной стороны, наличию больших корпусов, с другой – благодаря наличию инструментов для обработки естественного языка.

Продемонстрировать это можно на примере немецкого предлога *wegen* 'из-за, по причине', нормативно допускающего вариативность управления.

2 Постановка задачи

Среди всех германских языков падежная система немецкого языка представляет собой промежуточный вариант между архаичной исландской и ярко выраженной аналитической ан-

¹Ср.: „Jeder in einer Grammatik kodifizierte Sprachgebrauch kann zur Norm erhoben werden, auch wenn die Autoren reine Deskription reklamieren.“ (Zifonun et al., 1997, p. 6) – Каждая зафиксированная любой из существующих грамматик особенность употребления языка может получить статус нормативной, даже если авторы формулировали свою задачу как чисто описательную.

²<https://www.duden.de>

³<https://www.dwds.de>

глейской, включая в себя, таким образом, две подсистемы – флективную и аналитическую (Dal, 1960). Этот факт не может не влиять на динамические процессы в морфологической системе языка. Редукция падежной системы в немецком – типичная для большинства германских языков – привела, в частности, к изменению семантики существующих падежей, что в свою очередь способствовало вытеснению их морфологических признаков предложными конструкциями.

Предлоги с генитивным управлением, будучи производными, непервообразными, стоят в предложной системе немецкого языка особняком: они возникли – по историческим меркам – относительно недавно, образовавшись преимущественно от основ существительных в ранненововерхненемецкий период (1350-1650), что и объясняет их управление: первичная функция падежа при них представляла собой адноминальный атрибутивный генитив. В течение продолжительного времени падежное управление при новых предлогах не было стабильным: генитив конкурировал с дативом, что не является случайным. В результате редукции падежной системы в ней остаются и развивают свою семантику падежи с наиболее абстрактным значением, их абстрактно-грамматический характер, а также многозначность – результат позднейшего обобщения на основе слияния форм и значений. Для немецкого языка максимально синкретическим оказался дательный падеж (Zhirmunskiy, 2020). Однако уже с начала XX в. генитив вновь занимает господствующее положение, что касается всей группы этих предлогов. Так, *wegen* в литературном языке употребляется в это время, как правило, с генитивом (Admoni, 1986, p. 119).

Глобальные процессы в падежной системе обусловили не только явление падежного синкретизма, но и способствовали развитию тенденции к проявлению аналитических признаков на грамматическом уровне и фразеологизации немаркированных форм в определенных видах дискурса (Zhirmunskiy, 2020; Schmid, 2017; Nübling, 2006). Весьма существенной становится роль предлогов, их количество и значения увеличиваются, а функции расширяются. С этим, в частности, связано и явление беспадежного употребления существительного после предлога, которое не тождественно явлению падежного синкретизма. Сам факт предложного управления с немаркированным падежом – явление не новое и характерное не только для исследуемых конструкций.

Сегодня благодаря доступности больших корпусов этот способ управления стал более заметен и не может игнорироваться грамматическим описанием, несмотря на то что традиционные словари и грамматики не упоминают это явление. Именно благодаря анализу больших объемов данных стала возможной дифференциация случаев морфологического синкретизма и беспадежного употребления существительных в составе предложной группы.

Изучение вариативности управления немецкого предлога *wegen* не представляется нам сегодня неразрешимой задачей: корпусные исследования вариативности немецких предлогов и, конкретно, предлога *wegen* уже проводились, и результаты представлены, например, на сайте проекта Grammis⁴ Института немецкого языка в г. Мангейме (далее – IDS), где обобщена информация, представленная в корпусе Deutsches Referenzkorpus⁵ (далее – DeReKo). В соответствующей словарной статье указано на возможность употребления предлога *wegen* как с генитивом, так и с дативом, а также без маркировки падежа (*ohne Kasusmarkierung*), однако данных о распределении частотности употреблений с разными падежами не приводится.

В исследовании предлогов, допускающих варьирование (Di Meola, 2009), проведенном также на основе DeReKo, были проанализированы вручную до 1000 контекстов на один предлог. Полученные данные автор разделил на следующие категории: 1) однозначное указание на генитив (для предлога *wegen* это 45,3% от общего числа проанализированных контекстов), 2) однозначное указание на датив (3,3% для *wegen*), 3) употребление в форме,

⁴<https://grammis.ids-mannheim.de>

⁵<https://cosmas2.ids-mannheim.de/cosmas2-web/>

которая может интерпретироваться как генитив (Genitiv-kompatibel) (49,3% для *wegen*), и 4) употребления в форме, которая может интерпретироваться как датив (сама форма исключает генитив, но маркеры датива отсутствуют) (2,1% для *wegen*). Этот анализ не выявляет, однако, доли беспадежного употребления и не позволяет однозначно определить, какие случаи попадают в категории (3) и (4). К тому же, анализ проводился вручную, а потому имеется ограничение на число контекстов.

Существуют также инструменты, с помощью которых можно выявить набор существительных, употребляемых с рассматриваемым предлогом. Сбор такой информации на основе текстов, полученных из корпуса, возможен с помощью Lexpan⁶, разрабатываемого также в IDS. Более того, имеется база данных по предложным группам Prepron⁷, в которой представлен и предлог *wegen* с наиболее частотными управляемыми им существительными.

Тем не менее, вышеприведенные результаты исследований и использованные инструменты не позволяют получить готовый ответ на вопрос о том, каково распределение доли генитивного, дативного и беспадежного управления. Для решения этого вопроса необходимо обращение к крупному корпусу, снабженному морфологической разметкой. На первый взгляд кажется, что задача может быть сведена к нескольким запросам в выбранных для исследования корпусах, в которых исследователь выявляет количество генитивного и дативного управления при рассматриваемом предлоге и наличие иных вариантов. На практике, однако, задача оказывается сложнее.

Материал для исследования управления предлога *wegen* основывался на данных нескольких корпусов, имеющихся в Sketch Engine⁸, DeReKo и DWDS⁹ (в основном анализировался немецкий язык в Германии без учета региональных вариантов). В последних двух корпусных менеджерах отсутствуют возможности поиска по падежам, только по частеречным тэгам. В Sketch Engine доступен поиск и по частеречным тэгам, и по падежам. Однако для поставленной исследовательской задачи возможностей поиска оказалось недостаточно, и разметка также местами спорная. К примеру, если в Sketch Engine в корпусе German Web 2018 запрашивать случаи генитивного управления *wegen* (запрос: [lemma="wegen"] [1,3 [case="Gen"]]), то в результатах поиска среди первых десяти контекстов будут: 1) случаи употребления с генитивом (*wegen seiner guten Englischkenntnisse* 'из-за его хорошего знания английского языка'), 2) случаи, в которых дифференциация генитив vs. датив не представляется возможной, так как у существительного женского рода в единственном числе формы генитива и датива совпадают (*wegen der riesigen Nachfrage* 'из-за огромного спроса'), 3) случаи с дативом (*wegen Mängeln der Leistung* 'из-за плохой успеваемости'), 4) случаи без указания на падеж (далее – беспадежное употребление), т. е. с отсутствующим артиклем и с отсутствующим окончанием у существительного либо же с окончанием существительного без показателей генитива или датива (*wegen Sex und Drogen* 'из-за секса и наркотиков'). В случае (4) разметка *Drogen* как формы генитива спорная, так как формальные характеристики генитива отсутствуют. В случае (3) *der Leistung* стоит в форме, идентичной и генитиву, и дативу, т. е. разметка также спорная, кроме того существительное *Leistung* является зависимым не предлога *wegen*, а существительного *Mängeln* (что соответствует дистанции, указанной в запросе, – до трех токенов между предлогом и словом в генитиве). Схожие результаты можно получить при аналогичном поиске для датива.

Можно было бы предположить, что надо смотреть на первое существительное после предлога. Однако и такой подход ведет к ошибкам. После предлога может идти расширенный причастный оборот (*wegen vom Regen aufgeweichter Straßen* 'из-за мокрых от дождя дорог'). Предлог *wegen* здесь управляет генитивом *aufgeweichter Straßen*, хотя первое существительное после него стоит в дативе, но управляется не предлогом *wegen*, а предлогом *von*: *vom Regen*. Более того, предлог *wegen* может встречаться также в постпозиции (*der tiefgrünen,*

⁶<http://uwv.ids-mannheim.de/lexpan>

⁷http://uwv.ids-mannheim.de/prepron/prepron_online.html

⁸<https://www.sketchengine.eu>

⁹<https://www.dwds.de/d/korpora/spk>

satten Farbe wegen ‘из-за насыщенного зеленого цвета’). Случаи употребления *wegen* в постпозиции относительно редки, и поиск существительного в генитиве или дативе, после которого идет *wegen*, дал бы слишком много нерелевантных контекстов. Чтобы учесть такие случаи, необходим поиск по синтаксическим поддеревьям, недоступный в Sketch Engine.

Таким образом, основными препятствиями для работы с корпусом в Sketch Engine стали невозможность поиска по синтаксическим поддеревьям, где вершиной являлся бы предлог *wegen*, и отсутствие в используемой системе тэгов обозначений для случаев, где нельзя точно сказать, идет ли речь о генитиве или дативе (например, обозначение типа Gen/Dat), и для случаев беспадежного употребления (так мы называем случаи, в которых нет никакого указания на падеж: артикля или окончания у существительного). Поэтому было решено использовать имеющиеся корпуса исключительно как источник материала и обрабатывать полученные контексты с помощью скриптов на языке программирования Python и с помощью библиотеки spaCy¹⁰.

3 Материал и метод

Были проанализированы данные следующих корпусов:

1. Корпус немецких газет (размер корпуса – 7 млрд. токенов), DeReKo;
2. Корпус устной речи *Gesprochene Sprache* (размер – 3 млн. токенов), DWDS;
3. Корпус интернет-текстов *German Web 2018* (размер – 5,3 млрд. токенов), Sketch Engine.

Первый корпус – подкорпус из архива корпусов *W – Archiv der geschriebenen Sprache*, DeReKo, составлен на основе имеющихся корпусов письменной речи и полностью состоит из публицистики. Для чистоты эксперимента были исключены австрийские и швейцарские источники и оставлены только немецкие. В виртуальный корпус вошли немецкие газеты с 1990 г. по 2019 г. Размер корпуса при этом составил 7 млрд. токенов.

Также были проанализированы данные корпуса DWDS устной речи (*Gesprochene Sprache*). Корпус охватывает года с 1900 г. по 2001 г. В него входят транскрипты официальных речей, протоколы парламентских заседаний и интервью. Размер корпуса небольшой (менее 3 млн. токенов).

В качестве третьего корпуса выступает корпус *German Web 2018*, представленный в Sketch Engine. Он достаточно крупный (5,3 млрд. токенов) и состоит из текстов, собранных в интернете с разных сайтов.

Здесь следует сделать небольшую оговорку относительно состава последних двух корпусов. Хотя корпус DWDS *Gesprochene Sprache* называется корпусом разговорной речи, это несколько условное обозначение. В нем в основном представлены официальные речи, то есть не спонтанная речь, а подготовленные и, скорее всего, прошедшие редактуру тексты. В то же время корпус *Sketch Engine German Web 2018*, являясь корпусом письменной речи, может отражать и некоторые явления разговорной речи, так как в него входят, в том числе, форумы и блоги, хотя назвать его корпусом разговорной речи тоже нельзя. Поэтому выводы, полученные в результате данного корпусного исследования, будут актуальны прежде всего для письменной речи. Возможно, на корпусе спонтанной устной речи данные будут отличаться.

В выше указанных корпусах объявлялся поиск на случаи употребления словоформы *wegen*. Никаких дополнительных ограничений в запросе не было. Из DeReKo и Sketch Engine было скачано из каждого корпуса до 10 тыс. контекстов (максимально допустимое количество для скачивания), из корпуса DWDS – 465 контекстов.

Все контексты делились на четыре категории:

1. генитив: в эту группу попадали все те случаи, где есть четкий показатель генитива (артикль, окончание прилагательного или окончание у существительного);
2. датив: случаи с четким показателем датива (артикль, окончание прилагательного или окончание у существительного);

¹⁰<https://spacy.io>

3. беспадежное употребление (далее – null): нет артикля, у существительного нет окончания, которое бы указывало на генитив или датив, прилагательное отсутствует;
4. генитив или датив (далее – Gen/Dat): существительные женского рода в единственном числе, у которых эти две формы совпадают.

Последняя группа в анализе игнорируется, так как не имеет значения для данного исследования, а первые три (генитив, датив и беспадежное употребление) составляют 100%. В группу с беспадежным употреблением входят как употребления в единственном (*wegen Körperverletzung* ‘из-за телесных повреждений’), так и во множественном числе (*wegen Kreislaufschwierigkeiten* ‘из-за проблем с кровообращением’). Они объединены в одну группу именно по формальным признакам, то есть по отсутствию указания на падеж. Надо отметить, что в размеченных данных есть возможность отдельно рассматривать случаи употребления в ед. ч. и во мн. ч.

Чтобы получить статистику по управлению предлога *wegen*, которой можно было бы доверять, понадобилось сделать скрипт, который анализирует полученный из корпуса материал¹¹. Этот скрипт написан на языке программирования Python с использованием библиотеки *sraCu*, в которой есть языковые модели для немецкого языка. С помощью нее можно в том числе синтаксически проанализировать текст и получить морфологическую информацию по элементам текста.

В процессе обработки корпусного материала собирается морфологическая информация о каждом элементе синтаксического поддерева, вершиной которого является предлог *wegen*, и определяется, какое существительное зависит непосредственно от него. Для нахождения этого существительного можно пользоваться синтаксическим разбором *sraCu*, но в ряде случаев он ошибочен. Особенно частотны ошибки при постпозиции предлога и при перечислениях, поэтому для определения зависимого существительного мы пользовались в дополнение к разбору *sraCu* собственным алгоритмом.

Особое внимание уделяется случаям с перечислениями, так как в них могут встречаться разные падежи. Например, во фразе *wegen gemeinschaftlicher schwerer räuberischer Erpressung, Freiheitsberaubung, Verstößen* ‘из-за коллективного вымогательства с отягчающими обстоятельствами, лишения свободы, нарушений’ присутствует Gen/Dat (*wegen gemeinschaftlicher schwerer räuberischer Erpressung*), который игнорируется в подсчетах, беспадежное употребление (*Freiheitsberaubung*) и датив (*Verstößen*). Разные падежи или беспадежное употребление внутри одного перечисления подсчитываются отдельно¹². *sraCu* также автоматически определяет падеж, но он не всегда верный, поэтому указанный *sraCu* падеж мы не учитывали.

Отдельно анализируются конструкции с постпозицией (*von +*) Gen + *wegen*. Они редко встречаются в корпусе, и у них ограниченное число заполнителей слота. Чаще всего это фраза *von Amts wegen* ‘по долгу службы’.

Сложность представляют также распространенные причастные обороты. Если между предлогом *wegen* и первым существительным после него есть еще один предлог, то существительным, зависящим от *wegen*, скорее всего будет являться последнее существительное в выделенном синтаксическом поддереве (например, *wegen der in seiner Heimat herrschenden Kriegswirren* ‘из-за царящих на его родине военных беспорядков’).

Когда найдено существительное, непосредственно зависящее от предлога *wegen*, определяется искомый падеж. Для этого, помимо информации о роде, числе существительного и его окончании, требуется информация о том, имеется ли у него артикль, и если да, то в какой форме, и стоит ли перед ним прилагательное, если да, то с каким окончанием. После

¹¹ Данные и скрипты представлены по ссылке: https://github.com/ln-kr/german_prepositions

¹² Встретилось менее 1% случаев, когда внутри группы однородных членов встречаются генитив и датив. Гораздо чаще некоторые из однородных существительных стоят в генитиве или дативе, а другие – с немаркированным падежом, ср: *wegen Förderung der Volks- und Berufsbildung sowie der Studentenhilfe und der Förderung des Umweltschutzes* ‘из-за продвижения национального и профессионального образования, а также помощи студентам и поощрения защиты окружающей среды’. Подобные контексты нашлись в корпусах письменной речи DeReKo и German Web.

DeReKo, корпус немецких газет с 1990 г.	%
генитив	60%
датель	4%
null	36%
DWDS, корпус <i>Gesprochene Sprache</i>	%
генитив	66%
датель	6%
null	28%
Sketch Engine, корпус <i>German Web 2018</i>	%
генитив	51%
датель	12%
null	37%

Таблица 1: Данные по употреблению в генитиве, датель и с беспадёжным управлением, %

учета всех этих параметров определяется падеж – генитив, датель, Gen/Dat или отмечается случай беспадёжного употребления.

Как можно видеть, определение падежа, которым управляет предлог, допускающий варьирование, является часто нетривиальной задачей. В выше разобранном случае с *wegen* задачу усложняет возможность нахождения предлога в постпозиции, наличие перечислений, внутри которых могут встречаться существительные в разных падежах или в беспадёжном употреблении, расширенных причастных оборотов и наличие предложных групп, зависящих не от предлога, а от других элементов поддерева.

4 Результаты и дискуссия

Подсчет проанализированных контекстов дал следующие результаты (см. табл. 1).

Таким образом, генитив встречается примерно в 60% случаев, беспадёжное употребление – в 35%, датель – в 5%¹³.

В целом можно сказать, что датель в проанализированных корпусах представлен слабо, минимум 4% в публицистике, максимум 12% в корпусе интернет-текстов. Гораздо более значимым с точки зрения вариативности нормы оказывается беспадёжное употребление. Оно значительно чаще встречается, чем употребление в датель, в трети случаев, о чем не говорится в учебниках и словарях.

Помимо прочего, полученные данные позволяют проанализировать контекстные условия, в которых встречается рассматриваемый предлог. Рассмотрим ниже употребления с датель и беспадёжное употребление.

По данным корпусов из Sketch Engine и DeReKo существительные, встретившиеся в датель больше одного раза, в основном имеют форму мн.ч. (например, *wegen Mängeln* ‘из-за недостатков’, *wegen Verstößen* ‘из-за нарушений’) и означают либо какие-либо недостатки и проблемы (*wegen Problemen* ‘из-за проблем’, *wegen Vorwürfen* ‘из-за упреков’), либо какие-либо нарушения и преступления (*wegen Drogendelikten* ‘из-за преступлений в сфере оборота наркотиков’, *wegen Eigentumsdelikten* ‘из-за имущественных преступлений’). Соответственно, существительные в датель после *wegen* встречаются в основном в текстах юридической направленности.

Наиболее частотные случаи беспадёжного управления – это существительные женского

¹³Личные местоимения после *wegen* употребляются практически всегда в датель в современном узусе, ср. *wegen mir* ‘из-за меня’, *wegen dir* ‘из-за тебя’ и пр. (адвербиальные формы с исходно генитивным элементом типа *meinetwegen* имеют другое значение).

рода, обозначающие либо нарушения и преступления (*wegen Beihilfe* 'из-за пособничества', *wegen Beleidigung* 'из-за оскорбления', *wegen Sachbeschädigung* 'из-за повреждения имущества'), либо травмы (*wegen Verletzung* 'из-за травмы, повреждения'). В основном существительные в беспадежном употреблении встречаются в текстах юридической направленности и в спортивных новостях (например, сообщения о том, что из-за травмы какой-либо игрок не участвует в соревнованиях).

Возможно, в корпусе устной спонтанной речи доля датива среди существительных оказалась бы выше. Об этом свидетельствуют и результаты других исследований (Babenko, 2016; Marossek, 2016).

Основные лексикографические источники современного немецкого языка Duden и DWDS, доступные также в Сети, сообщают, что предлог *wegen* может управлять как генитивом, так и дативом, однако акценты касательно модели управления этого предлога расставлены в этих словарях по-разному. Duden указывает, что оба падежа – генитив и датив – представляют собой абсолютно равноправные употребления, не нарушающие грамматические нормы немецкого языка, причем употребление датива с *wegen* в письменной речи является стандартным разговорным вариантом. Duden, не вдаваясь в подробности, замечает на полях, что после *wegen* зависимое существительное может также употребляться в несклоняемой форме.

DWDS выделяет только два режима предложного управления для предлога *wegen*: основной с генитивом и альтернативный с дативом. Про беспадежное управление *wegen* пользователь этого корпусного словаря не получает вообще никакой информации. Дативные употребления DWDS снабжает дополнительно пометами «разговорный», «южно-немецкий», «австрийский». Представляется, что подобная характеристика употребления *wegen* несколько устарела, поскольку сегодня дативное управление *wegen* более не имеет территориальных ограничений и достаточно широко представлено в практике носителей немецкого языка в немецкоязычных странах.

Как видно из вышеупомянутых описаний, картина языковой нормы в отношении употреблений предлога *wegen* остается несколько неясной, особенно для неносителя немецкого языка. Основной отечественный двуязычный словарь под руководством О.И. Москальской (как и DWDS) сообщает пользователю, что *wegen* управляет генитивом, а датив представлен лишь как разговорный вариант в южных регионах Германии. Корпусные данные это опровергают.

Приведем фрагмент словарной статьи предлога *wegen* из «Немецко-русского словаря актуальной лексики», находящегося в данный момент в печати. Этот словарь, как представляется, учитывает современное словоупотребление и с помощью различных метаязыковых комментариев помогает пользователю разобраться в особенностях употребления предлога *wegen*. Сильной стороной этого словаря является также наличие основанных на корпусных данных иллюстративных примеров, что особенно важно для лексикографического представления служебной лексики.

wégen *ppr* (*G*, в некоторых формах и разг. *D*) *тж.* в *постпозиции* *ради*; *из-за*; *вследствие*; *благодаря*; ~ *éines Mótorschadens könnte die Fahrt nicht förtgesetzt wéren* *из-за* [вследствие] поломки двигателя поезде пришлось прервать; *séines Gejámms* ~ [séinem Gejámmer ~] *máchten wir éine Páuse* *из-за* его нытья мы сделали перерыв; с личными местоимениями *тж.* *D*: ~ *dir* *из-за* тебя; в *препозиции* возможно беспадежное управление – в случае отсутствия при суц. артикля или определяющих слов: ~ *Kránkheit* по болезни; ~ *Betrúg* по причине обмана

5 Заключение

Получение достоверных статистических данных по вариативному управлению предлогов с помощью имеющихся для немецкого языка корпусных менеджеров оказывается затруднительным по ряду причин. В DeReKo и DWDS нет разметки падежей, а в Sketch Engine

эта разметка не всегда корректна. Имеющийся набор тэгов для падежей не предполагает случаев, в которых форма слова может одновременно соответствовать двум падежам (дативу и генитиву) или где вообще нет никакого указания на падеж. Синтаксический анализ полученного из корпусов текстового материала с помощью иных средств (библиотеки *sraСу*) помогает приблизиться к цели, однако сам по себе не является достаточным. По этой причине для получения информации об управлении предлога необходимо дополнить набор *sraСу* собственными алгоритмами, в которых будут учтены все детали (возможность нахождения предлога в постпозиции, особая обработка перечислений и расширенных причастных оборотов) и соблюдена выбранная классификация (генитив, датив, беспадёжное употребление).

На примере предлога *wegen* мы показали возможности современных эмпирических методов исследования, которые позволяют внести ясность в проблематику вариативности грамматической нормы, актуальной для построения корпусной грамматики разных языков.

Новизна полученных результатов состоит в том, что впервые при классификации употреблений предлога *wegen* в особую группу выделяются случаи беспадёжного употребления существительных и именных групп, управляемых этим предлогом. Полученные статистические данные по различным типам управления являются ценным лингвистическим результатом, который используется в лексикографии и преподавании немецкого языка. Теоретический вывод из данного исследования состоит в том, что полученные результаты отражают общие типологические тенденции в грамматике современного немецкого языка.

Acknowledgements

Авторы благодарят анонимных рецензентов за внимательное прочтение работы, конструктивную критику и замечания, учтенные в окончательной версии статьи.

References

- Vladimir G. Admoni. 1986. *Teoreticheskaya grammatika nemetskogo yazyka (na nemetskom yazyke)* [Theoretical grammar of the German language (in German)]. Prosveshcheniye, Moskva, 4 edition.
- Natalia S. Babenko. 2016. Yeshche raz o kategorii padezha: 'obshchiy padezh' v sovremennom nemetskom yazyke [Once again about the category of case: 'common case' in modern German]. *Voprosy psikholingvistiki* [Questions of psycholinguistics], 28:38–43.
- Ingerid Dal. 1960. Entwicklungstendenzen im germanischen Kasussystem. *Untersuchungen zur germanischen und deutschen Sprachgeschichte*, 2:158–170.
- Claudio Di Meola. 2009. Rektionsschwankungen bei Präpositionen – erlaubt, verboten, unbeachtet. *Deutsche Grammatik - Regeln, Normen, Sprachgebrauch*, P 195–221.
- Duden 9. 2021. *Sprachliche Zweifelsfälle: das Wörterbuch für richtiges und gutes Deutsch*. Dudenverlag, 9. überarbeitete und erweiterte Auflage edition.
- Mathilde Hennig and Christoph Müller. 2009. *Wie normal ist die Norm? Sprachliche Normen im Spannungsfeld von Sprachwissenschaft, Sprachöffentlichkeit und Sprachdidaktik*. Kassel University Press, Kassel.
- Diana Marossek. 2016. *Kommst du Bahnhof oder hast du Auto? Warum wir reden, wie wir neuerdings reden?* Hanser Berlin im Carl Hanser Verlag, München.
- Doris Nübling. 2006. *Historische Sprachwissenschaft des Deutschen. Eine Einführung in die Prinzipien des Sprachwandels*. Gunter Narr Verlag, Tübingen.
- Hans Ulrich Schmid. 2017. *Einführung in die deutsche Sprachgeschichte*. J. B. Metzler Verlag, Stuttgart.
- Wahrig. 2003. *Fehlerfreies und gutes Deutsch. Das zuverlässige Nachschlagewerk zur Klärung sprachlicher Zweifelsfälle*. Bertelsmann, Gütersloh / München.
- Viktor M. Zhirmunskiy. 2020. *Istoriya nemetskogo yazyka* [History of the German language]. Stereotypical edition URSS.
- Gisela Zifonun, Ludger Hoffmann, et al. 1997. *Grammatik der deutschen Sprache (GDS)*. De Gruyter, Berlin / New York.