

Д. О. ДОБРОВОЛЬСКИЙ, Е. Б. КРОВОТА,  
И. С. ПАРИНА

## КОРПУСНАЯ ЛЕКСИКОГРАФИЯ <sup>1</sup>

(материалы мастер-класса)

Цель данной работы — попытаться ответить на вопрос, насколько текстовые корпуса (в том числе и корпуса параллельных текстов) могут оказаться полезным инструментом как в исследовании немецкого языка, так и в решении задач контрастивной лингвистики, то есть в описании особенностей функционирования форм немецкого языка в сопоставлении с русским. В центре нашего внимания стоят лексические явления сопоставляемых языков и способы их представления в двуязычных словарях. Материал параллельных корпусов провоцирует вопросы лексикографического представления<sup>2</sup> межъязыковых лексических коррелятов, особенно если подобные соответствия не являются эквивалентами в точном смысле.

Значительное внимание в данной работе уделяется также описанию состава релевантных корпусов и способам работы с ними.

### 1. Корпус немецко-русских и русско-немецких параллельных текстов в Национальном корпусе русского языка

Создание корпуса параллельных текстов в качестве одного из модулей Национального корпуса русского языка (НКРЯ) — актуальная лингвистическая задача, над решением которой вот уже больше десяти лет работает целый коллектив исследователей. Такой корпус необходим в первую очередь потому, что надежные, подкрепленные эмпирическими данными ответы на большинство вопросов сопоставительной лингвистики, в том числе лексикологии и лексикографии, могут быть получены только при наличии значительных по объему и общедоступных параллельных корпусов. Ср. подробнее [Добровольский, Кретов, Шаров 2005; Добровольский 2009].

---

<sup>1</sup> Работа выполнена при поддержке РГНФ (грант 12-04-12041), РФФИ (грант 13-06-00403) и Программы фундаментальных исследований Президиума РАН «Корпусная лингвистика». Авторы благодарят Д. В. Сичинаву, ознакомившегося с рукописью статьи, за полезные замечания и советы.

<sup>2</sup> Вопросы использования корпусных данных в — прежде всего, одноязычной — лексикографии обсуждаются в [Atkins, Fillmore, Johnson 2003]. Ср. также спецвыпуск журнала «Sprachtypologie und Universalienforschung», посвященный использованию параллельных корпусов в языковом сравнении и типологии [Cysouw, Wälchli 2007].

Корпус параллельных текстов НКРЯ состоит в основном из множества художественных произведений с их переводами. В относительно небольшом объеме в корпус включены тексты СМИ. Корпус включает в себя, с одной стороны, оригинальные русские тексты с их переводом на другие языки, а с другой — иноязычные тексты с их переводом на русский. В дальнейшем мы будем говорить исключительно о русско-немецких и немецко-русских частях корпуса.

### 1.1. Зачем нужен параллельный корпус?

Создание и использование параллельных корпусов представляется целесообразным как с практической точки зрения, так и с точки зрения развития корпусной лингвистики — одного из наиболее перспективных лингвистических направлений. Корпус параллельных текстов может быть эффективно использован в различных лингвистических исследованиях (в первую очередь, в области контрастивной лексикологии и двуязычной лексикографии), а также в исследованиях по теории перевода, сравнительному литературоведению, культурологии, автоматической обработке текста и др. В разделе 1 нас будут в первую очередь интересовать собственно лингвистические — и прежде всего лексикографические — аспекты использования русско-немецких и немецко-русских параллельных корпусов.

Обращаясь к корпусу, создатели двуязычных словарей получают весьма простой и эффективный инструмент сбора материала и эмпирической проверки своих гипотез, касающихся межъязыковой эквивалентности. Ценность этого инструмента определяется тем, что в лингвистике этап сбора материала является наиболее трудоемким и наименее творческим, а подобные корпуса параллельных текстов позволяют значительно сэкономить время и силы для творческого этапа работы.

Современное состояние сравнительной лексикологии и двуязычной лексикографии характеризуется ориентацией на сопоставление более или менее изолированных языковых структур. Отрицательным последствием подобной ориентации является недостаточный учет узуса, то есть тех особенностей синтаксического и сочетаемостного поведения единиц языка, которые нельзя объяснить их системными признаками. Так, в принципе известно, что та или иная структура одного языка не может быть во всех контекстах переведена на другой с помощью своего стандартного эквивалента. В определенных контекстах язык L2 традиционно прибегает к другим способам описания соответствующей ситуации. Известно также, что не существует продуктивных правил, по которым можно было бы вывести подобные отклонения от «стандартной эквивалентности» из неких более общих принципов. Единственный способ описания подобных отклонений — это их тщательная фиксация на аутентичном материале. Только так можно построить исчерпывающие сопоставитель-

ные описания и создать словари, удовлетворяющие современным требованиям.

Корпус параллельных текстов представляет собой наиболее адекватный инструмент для выполнения этих задач. Та или иная языковая структура, интересующая исследователя, может быть найдена во всех представленных в корпусе контекстах с их переводами на соответствующий язык. Таким образом, исследователь получает в свое распоряжение набор аутентичных контекстов, представляющих интересующую его структуру в ее естественном окружении, а также самые разнообразные эквиваленты этой структуры в языке-цели. Поскольку эти эквиваленты также оказываются встроенными в естественные контексты, на основе полученных с помощью параллельного корпуса материалов могут быть сделаны выводы о зависимости выбора эквивалента от типа контекста. Подобные результаты практически всегда расходятся с теми сведениями, которые мы можем почерпнуть из традиционных, созданных в «докорпусную эру» словарей, являясь тем самым нетривиальными.

Важным параметром, по которому языки могут различаться между собой, является степень употребительности определенных выражений. Так, некоторое выражение *A* языка *L1* может стандартным образом переводиться на язык *L2* с помощью выражения *B* — вполне корректного с точки зрения норм этого языка. Таким образом, выражения *A* и *B* оказываются эквивалентными в рамках системы соответствующих языков. Тем не менее их функциональная эквивалентность часто представляется неполной. В частности, это имеет место в случае, когда одно из выражений оказывается в своем языке существенно более употребительным, чем его переводной эквивалент — в своем. Такие случаи хорошо прослеживаются при сопоставлении оригинальных текстов с их переводами на другие языки. О. Б. Йокояма [2012: 152] замечает по этому поводу: «Со становлением корпусной лингвистики в середине 90-х гг. появились (ранее недоступные) материалы и новые подходы к исследованию, например сравнение статистических показателей определенных языковых признаков в исходных текстах с частотностью тех же признаков в переводных текстах».

Исследование некоторого языкового явления на основе корпуса параллельных текстов может быть по ряду параметров противопоставлено исследованию этого явления на основе большого корпуса оригинальных текстов. Отличие оригинальных текстов от переводов заключается и в объеме (миллионы слов оригинальных текстов, производимых ежедневно, против относительно небольшого количества текстов, переводимых с иностранных языков), и в природе авторства (то есть в степени оригинальности и творческой свободы при порождении текста), а также в культурном контексте (переводные тексты обычно погружены в культуру исходного языка).

Все эти факторы обеспечивают различия между исходными и переводными текстами по целому ряду параметров. Между конкретными словами с близкими значениями в разных языках (которые принято считать эквивалентными) нет взаимно однозначных соответствий. В принципе это касается многих семантических классов слов. Каждое конкретное слово обладает своим уникальным набором сочетаемостных ограничений и предпочтений. Эта идея относится к числу традиционных теоретических положений лингвистики. То новое, что вносит обращение к корпусам параллельных текстов в решение подобных задач, — это возможность эмпирической проверки соответствующих гипотез на представительном материале.

### 1.2. Состав немецко-русского и русско-немецкого корпуса в НКРЯ

На данный момент в состав доступного online немецко-русского корпуса (наряду с небольшим массивом публицистических текстов, любезно предоставленных компанией АВВУУ) входят следующие произведения<sup>3</sup>:

- Heinrich Böll (Г. Бёльль):  
Ansichten eines Clowns — Глазами клоуна;
- Gottfried August Bürger (Г. А. Бюргер):  
Die Abenteuer des Freiherrn von Münchhausen —  
Приключения барона Мюнхгаузена;
- Friedrich Dürrenmatt (Ф. Дюрренматт):  
Justiz — Правосудие;  
Der Verdacht — Подозрение;
- Josef von Eichendorff (И. фон Эйхендорф):  
Aus dem Leben eines Taugenichts — Из жизни одного  
бездельника;
- Michael Andreas Hellmuth Ende (М. А. Х. Энде):  
Momo oder Die Geschichte von den Zeit-Dieben — Момо, или  
История об украденном времени;
- Johann Wolfgang Goethe (И. В. Гёте):  
Die Leiden des jungen Werther — Страдания юного Вертера;
- Hermann Hesse (Г. Гессе):  
Siddhartha — Сиддхартха;  
Peter Camenzind — Петер Каменцинд;
- Ernst Theodor Amadeus Hoffmann (Э. Т. А. Гофман):  
Der goldne Topf — Золотой горшок;  
Klein Zaches, genannt Zinnober — Крошка Цахес по прозвищу  
Циннобер;

<sup>3</sup> Здесь и в дальнейшем сохраняется орфография оригинального издания, вошедшего в корпус.

- Nußknacker und Mausekönig — Щелкунчик и Мышиный  
Король;
- Franz Kafka (Ф. Кафка):  
Der Prozess — Процесс;  
Die Verwandlung — Превращение;
- Erich Kästner (Э. Кестнер):  
Pünktchen und Anton — Кнопка и Антон;  
Emil und die Detektive — Эмиль и детективы;
- Heinrich von Kleist (Г. фон Клейст):  
Michael Kohlhaas — Михаэль Кольхаас;
- Thomas Mann (Т. Манн):  
Der Zauberberg — Волшебная гора;  
Buddenbrooks — Будденброки
- Gustav Meyrink (Г. Майринк):  
Der Golem — Голем;
- Novalis (Новалис):  
Heinrich von Ofterdingen — Гейнрих фон Офтердинген;
- Erich Maria Remarque (Э.-М. Ремарк):  
Der schwarze Obelisk — Черный обелиск;
- Arthur Schnitzler (А. Шницлер):  
Casanovas Heimfahrt — Возвращение Казановы;  
Die Frau des Weisen — Жена мудреца;
- Patrick Süskind (П. Зюскинд):  
Das Parfum. Die Geschichte eines Mörders — Парфюмер.

В стадии подготовки находится роман Теодора Фонтане «Эффи Брист» (Theodor Fontane — Effi Briest), а также статья Вальтера Бен-ьямина «О понятии истории» (Walter Benjamin — Über den Begriff der Geschichte).

Работа над русско-немецким корпусом была начата относительно недавно. Пока он существенно уступает по объему немецко-русскому. Сейчас в открытом доступе находятся лишь пять публицистических текстов, повесть А. С. Пушкина «Капитанская дочка» с сайта Gutenberg-DE, предположительно, в переводе Ф. Фриш (Fega Frisch)<sup>4</sup>, и роман Н. А. Островского «Как закалялась сталь». В ближайшее время корпус будет дополнен следующими произведениями русской классической литературы:

- Н. В. Гоголь:  
Шинель — Der Mantel;  
Мертвые души — Die Toten Seelen;  
Вечера на хуторе близ Диканьки — Abende auf dem Vorwerke  
bei Dikanjka und andere Erzählungen;

<sup>4</sup> Этот параллельный текст, а также перечисленные здесь произведения Кестнера, Шницлера, «Превращение» Кафки и «Как закалялась сталь» Островского были любезно предоставлены проекту НКРЯ А. Барентсеном.

- Портрет — Das Porträt;  
 Записки сумасшедшего — Aufzeichnungen eines Irren;  
 Нос — Die Nase;  
 Невский проспект — Der Newskij-Prospekt;  
 Вий — Der Wij;  
 — Ф. М. Достоевский:  
 Идиот — Der Idiot;  
 Преступление и наказание — Verbrechen und Strafe;  
 Братья Карамазовы — Die Brüder Karamasow;  
 Бедные люди — Arme Leute;  
 — М. Ю. Лермонтов:  
 Герой нашего времени — Ein Held unserer Zeit;  
 — А. С. Пушкин:  
 Повести покойного Ивана Петровича Белкина —  
 Die Erzählungen des verstorbenen Iwan Petrowitsch Belkin;  
 Пиковая Дама — Pique Dame;  
 Дубровский — Dubrowskij;  
 Капитанская дочка — Die Hauptmannstochter (в переводе  
 Фреда Отто — Fred Ottow);  
 — Л. Н. Толстой:  
 Война и мир — Krieg und Frieden;  
 Анна Каренина — Anna Karenina;  
 Крейцерова соната — Die Kreuzersonate;  
 Семейное счастье — Glück der Ehe;  
 Хозяин и работник — Herr und Knecht;  
 — И. С. Тургенев:  
 Отцы и дети — Väter und Söhne;  
 Фауст — Faust;  
 — А. П. Чехов:  
 Дуэль — Ein Zweikampf;  
 — Н. С. Лесков:  
 Леди Макбет Мценского уезда — Die Lady Makbeth  
 des Mzensker Landkreises;  
 Запечатленный ангел — Der versiegelte Engel  
 — И. А. Гончаров:  
 Обломов — Oblomow;  
 — Максим Горький:  
 Трое — Drei Menschen.

В следующем разделе мы на конкретных примерах попробуем показать возможности использования корпуса для описания лексических единиц немецкого языка в сопоставлении с русским.

### 1.3. Параллельные тексты НКРЯ в сопоставительном описании лексики: немецкие показатели пространственного дейксиса: элементы *her* и *hin* и их русские соответствия

В данном разделе будет показано, как работа с параллельным корпусом может помочь в выявлении русских соответствий определенного немецкого слова, обладающего нетривиальными особенностями. Когда у некоторого слова в специфических употреблениях нет стандартных аналогов в других языках, может создаться впечатление, что подобные смыслы непереводимы в принципе. Корпус показывает, что способы передачи таких смыслов все-таки существуют, и позволяет эти способы описать. В качестве материала мы берем такие лексические единицы немецкого языка, про которые принято считать, что они не имеют русских аналогов, и пытаемся с помощью контекстов из параллельного корпуса выявить возможные способы передачи соответствующих смыслов.

В немецком языке есть огромное количество слов с пространственно-дейктическими элементами *her* (≈ ‘сюда’) и *hin* (≈ ‘туда’), выражающими идею направления движения<sup>5</sup>. Слова *hin* и *her* употребляются также и самостоятельно в различных комбинациях с глаголами и адвербиальными конструкциями для выражения пространственных отношений между участниками ситуации. В русском языке нет таких системно организованных (то есть почти грамматических) средств для выражения соответствующих дейктических смыслов. Так, и глагол *hereinführen*, и глагол *hineinführen* переводятся на русский язык с помощью глаголов *вводить/ввозить*, так что идея ‘вводить/ввозить сюда, то есть по направлению к говорящему’ vs. ‘вводить/ввозить туда, то есть по направлению от говорящего’<sup>6</sup> в русском языке часто вообще не выражается.

В работе [Добровольский, Падучева 2008] было показано, что проблема состоит не только в отсутствии у этих немецких слов стандартных русских аналогов, но и в том, что комментарии к немецким *her* и *hin*, апеллирующие к говорящему, искажают реальную картину. Так, человек, желающий войти в некоторое помещение, (в стандартной ситуации) спросит, постучав: *Darf ich herein?* (то есть *Можно мне войти?*), сокращенная форма от *Darf ich hereinkommen?* ‘Можно мне сюда войти?’), а не *Darf ich hinein?* (букв. ≈ «Можно мне туда войти?»). Понятно, что предполагаемое перемещение в пространстве должно совершаться не по направлению к говорящему, а от его исходного местонахождения по направлению к адресату. Дело в том, что в та-

<sup>5</sup> С типологической точки зрения *her* и *hin* — это показатели глагольной ориентации; ср. [Плунгян 2011].

<sup>6</sup> Точнее ‘по направлению к говорящему или к лицу, точку зрения которого принимает говорящий’ vs. ‘по направлению от говорящего или от лица, точку зрения которого принимает говорящий’. Ср. [Добровольский, Падучева 2008].

ких случаях говорящий принимает точку зрения адресата, как бы ставя себя на его место.

Еще более сложно ведут себя такие дейктические элементы в гипотаксисе и нарративе. Ср. функционирование вопросительных местоименных наречий, в состав которых входят *her* или *hin*. Значение данных дейктических элементов (и, соответственно, их перевод на русский язык) меняется в зависимости от того, употребляется ли это наречие в прямом вопросе или в качестве относительного наречия в гипотаксисе. В гипотаксисе субъектом дейксиса является не говорящий, а подлежащее главного предложения. В [Добровольский, Падучева 2008] приводится такой пример: *Alle schauten entsetzt auf den Felsen, woherunter der Bergsteiger gestürzt war* — Все в ужасе посмотрели на скалу, откуда <с которой> сорвался альпинист. Выбор наречия *woherunter* (в отличие от *wohinunter*) мотивирован в этом случае местонахождением людей, смотрящих на скалу (*alle*). Контекст ясно показывает, что они стоят внизу и смотрят на скалу снизу вверх. Иными словами, *hin-* и *her-*слова показывают, где находится наблюдатель, а не говорящий. *Woherunter* — это как бы «где-сюда-низ», а *wohinunter* — «где-туда-низ». Соответственно, если наблюдатель наверху, он скажет «где-туда-низ», то есть *wohinunter*, по-русски что-то вроде *куда* (*вниз*), а если он внизу — *woherunter* — «где-сюда-низ», то есть *откуда* (*вниз*).

В качестве примера интерпретации дейктических слов в нарративном режиме мы использовали контексты (1) и (2) из НКРЯ.

- (1) Prosper Alpanus senkte sich *herab* zu dem Jüngling, an dessen Seite er Platz nahm, während die Libelle aufflog in ein Gebüsch und in den Gesang einstimmte, der durch den ganzen Wald tönte. (Е. Т. А. Hoffmann. Klein Zaches genannt Zinnober)

Проспер Альпанус спустился к юноше и сел подле него, стрекоза упорхнула в кусты, вторя пению, наполнявшему весь лес.

Здесь в принципе возможно и *hinab*. Но в этом случае субъектом дейксиса был бы Проспер Альпанус. А в перспективе повествования, выбранной автором, субъект дейксиса — юноша. Как видно из перевода, русский язык не располагает стандартными средствами, позволяющими выразить эти различия: ‘спустился сюда’ vs. ‘спустился туда’. Более точный перевод был бы *Проспер Альпанус спустился сюда к юноше*.

- (2) Er schaute *hinauf* und erblickte staunend Prosper Alpanus, der auf einem wunderbaren Insekt, das einer in den herrlichsten Farben prunkenden Libelle nicht unähnlich, daherschwebte. (Е. Т. А. Hoffmann. Klein Zaches genannt Zinnober)

Он поднял глаза и с изумлением увидел Проспера Альпануса, летевшего к нему на каком-то диковинном насекомом, нелишенном сходства с великолепной, сверкающей всеми красками стрекозой.



*Hinauf* в предложении (2), которое в тексте предшествует предложению (1), развертывает событие в противоположной перспективе. Субъект восприятия здесь — тот же юноша. Но в одном случае движение направлено к нему, потому что Проспер Альпанус спустился к нему вниз, а в другом — от него. В последнем случае решающим фактором для выбора дейктического элемента является направление взгляда юноши снизу вверх. Итак, субъект дейксиса в нарративном режиме — это один из персонажей, глазами которого описываются события в конкретном контексте. Подобные сдвиги в осмыслении пространственных отношений еще более усложняют поиск адекватных русских соответствий.

Как параллельный корпус может использоваться для исследования особенностей значения и употребления дейктических слов с элементами *hin* и *her*, а также для выявления их контекстно-зависимых русских соответствий? Попробуем ответить на этот вопрос на материале адвербиальных выражений *herein* и *hinein*, которые употребляются и как самостоятельные наречия, и в составе сложных глаголов.

Обратимся к примерам из НКРЯ. В первую очередь, примеры показывают, что, помимо самых простых и однозначных выражений с элементами *tуда* для *hin* и *сюда* для *her*, существует целый ряд способов передать направление движения относительно точки отсчета — говорящего или наблюдателя, которые используются далеко не во всех контекстах, а только там, где это коммуникативно необходимо. Для *herein* это, например, такие предложные группы, как *на нас* (3). А для *hinein* это могут быть, среди прочего, глаголы, более или менее четко указывающие на направление движения от наблюдателя (4). В примере (5) глагол *войти* дейктически нейтрален: войти можно как куда-то, где находится наблюдатель (именно эта ситуация имеет место в данном контексте), так и в какое-то помещение прочь от наблюдателя.

- (3) Sie weht *herein* mit dem Geruch der Erde, der Blumen und dem Fun-  
keln der Sterne. (E. M. Remarque. Der schwarze Obelisk)  
Она дышит *на нас* запахом земли, цветов и мерцанием звезд.
- (4) Die gigantische Maschine *hob sich* mit Dr.h.c. Isaak Kohler in den  
Nachthimmel *hinein*, heulend, brüllend, ab nach Australien, als ich aus  
meinem Volkswagen sprang, den Revolver entschert. (F. Dürrenmatt.  
Justiz)  
Когда я выскочил из своего «фольксвагена», сняв револьвер с  
предохранителя, гигантская машина, уносящая доктора *honoris  
causa* Исаака Колера, с ревом и воем *взмыла* в ночное небо и взяла  
курс на Австралию.

- (5) Dann ging die Tür auf, und Antons Mutter *kam herein*. (E. Kästner. Pünktchen und Anton)  
Вдруг дверь открылась и в детскую *вошла* мать Антона.

Примеры (6—10) показывают, как выбор дейктического элемента в нарративе позволяет установить, кто из персонажей является в данный момент субъектом перспективы повествования, а также в каком месте располагается в каждом конкретном эпизоде точка наблюдения. В русских переводах это чаще всего видно только из контекста.

- (6) <...> da gab ihm der Vater von hinten einen jetzt wahrhaftig erlösenden starken Stoß, und er *flog*, heftig blutend, weit in sein Zimmer *hinein*. (F. Kafka. Die Verwandlung)  
И тогда отец с силой дал ему сзади поистине спасительного теперь пинка, и Грегор, обливаясь кровью, *влетел* в свою комнату.
- (7) „Was denn?“ fragte Pünktchen harmlos. Gottfried Klepperbein vertraf ihr drohend den Weg. „Das weißt du schon ganz gut, stell dich nicht so dumm, mein Herzblatt!“ Pünktchen wollte gern ins Haus, aber er *ließ* sie nicht *hinein*. (E. Kästner. Pünktchen und Anton)  
— Это что же? — простодушно осведомилась Кнопка. Готфрид Клеппербейн грозно загородил ей дорогу. — Сама знаешь, не строй из себя дурочку, дорогуша! Кнопка хотела войти в дом, но он не *пускал* ее.
- (8) Aber es *fiel* so viel Russ *herein*, dass das Mädchen die Luke gleich wieder zuziehen und mit ihrem Taschentuch die Hände K.s vom Ruß reinigen mußte, denn K. war zu müde, um das selbst zu besorgen. (F. Kafka. Der Prozess)  
Но *посыпалось* столько сажки, что девушке пришлось тут же закрыть люк и смахнуть сажу с рук К. своим носовым платком, потому что сам он слишком ослабел.
- (9) Wir wohnten in der Nähe des Ehrenfelder Bahnhofs, blickten vom Fenster unseres Zimmers aus auf die rote Backsteinmauer des Bahndamms, Braunkohlenzüge *fuhren* voll in die Stadt *herein*, leer aus ihr *hinaus*, ein trostreicher Anblick, ein herzbewegendes Geräusch, ich mußte immer an die ausgeglichene Vermögenslage zu Hause denken. (H. Böll. Ansichten eines Clowns)  
Пансион находился неподалеку от Эренфельдской товарной станции; из окон нашей комнаты была видна красная кирпичная стена и товарные составы, в город *шли* платформы с бурым углем, из города *отправлялся* порожняк — отрадная картина, приятный перестук колес: как тут не подумать, что имущественное положение семьи Шниров вполне надежно.

- (10) Aber was weiß er schon? Beweisen kann er nichts. Trotzdem kann er mit mir spielen, wie er will. Der Arzt *kommt herein*. Es ist nicht der Direktor; es ist der behandelnde Arzt. „Essen Sie mit uns?“ fragt Bodendiek. (E.M. Remarque. Der schwarze Obelisk)  
 Но какая цена этому знанию? Доказать он ничего не может. И все-таки он играет со мною, как ему вздумается. *Входит* врач. Это не директор — это лечащий врач. — Вы ужинаете с нами? — осведомляется Бодендик.

Так, в примере (6) субъект дейксиса — Грегор, в (7) — Кнопка. Причем важно, что не Кнопка, а другой персонаж является в данном случае субъектом действия и, соответственно, подлежащим в предложении. Аналогично и в (10): субъект действия и подлежащее — врач. Но не он является субъектом дейксиса. Иначе было бы не *kommt herein*, а *kommt hinein*. Выбор дейктического элемента ясно показывает, что наблюдатель находится внутри — в той комнате, в которую входит врач. А значит, субъект наблюдения и дейксиса (как и в примере (9)) — диегетический повествователь, то есть персонаж, ведущий повествование от 1-го лица. В примере (9) комбинация дейктических адвербиалов *herein* и *hinaus* однозначно демонстрирует, что наблюдатель — он же диегетический повествователь — находится в городе, а не за его пределами, что коррелирует с указанием места расположения пансиона, в котором живут герои.

Ориентация немецкого пространственного дейксиса в нарративе на наблюдателя подтверждается выбором в русских переводах таких глаголов, как *появиться, показаться, явиться, доноситься*. Известно, что смысл ‘наблюдатель’ входит в значение этих глаголов, а указание на фигуру наблюдателя является неотъемлемой частью их толкования.

- (11) ⟨...⟩ und es kam immer ein Punkt, wo er mich wie ein heulender Hund ansah, und meistens *kam* gerade dann Sabine *herein*, schaute ihn verbittert an, weil sie wieder schwanger war. (H. Böll. Ansichten eines Clowns)  
 ⟨...⟩ а потом вдруг наступала минута, когда он смотрел на меня глазами затравленного зверя; тут в комнате обычно *появлялась* Сабина и смотрела на Карла с видом жертвы: она опять ходила беременная.
- (12) Wieder *kam* der junge Mann mit dem Schnurrbart *herein*, senkte einmal das Kinn auf die Brust und setzte sich mit sorgenvoll-verschlossener Miene. (Th. Mann. Der Zauberberg)  
 Снова *явился* молодой человек с усами, опустил подбородок на грудь и уселся с озабоченным и замкнутым видом.

- (13) Nach Osten und Süden standen die Fenster des Turmgemachs offen, aus Garten und Feld drangen linde, süße Gerüche aller Art, aus der Landschaft unbestimmte Geräusche zu ihm *herein*, wie die kommende Frühe sie aus der Weite und Nähe zu bringen pflegt. (A. Schnitzler. Casanovas Heimfahrt)  
Окна комнаты, выходившие на восток и на юг, были раскрыты. Из сада и с полей *доносились* нежные и сладкие ароматы, повсюду, вдалеке и вблизи, слышались неясные шорохи, которые обычно приносит с собой наступающее утро.
- (14) Hie und da *spähten* ein paar Schutzleute mit grünlich schillernden Federbüschchen *herein* und gingen in langsamem, schwerem Schritt wieder weiter. (G. Meyrink. Der Golem)  
Время от времени *показывались* городовые с зеленоватыми и блестящими перьями на шляпах и медленными тяжелыми шагами шли дальше.

В основе семантики слова *hinein* (в его пространственном значении<sup>7</sup>), наряду с указанием на направление прочь от говорящего или от лица, точку зрения которого принимает говорящий, лежит идея перемещения из более открытого в более замкнутое пространство. Интересны контексты типа (15) и (16), в которых речь идет о ментальной (духовной или интеллектуальной) сфере. Употребление *hinein* в таких контекстах показывает, что в модели мира, зафиксированной в немецком языке, эта сфера мыслится как замкнутое пространство, причем отдельное от субъекта. Иными словами, даже если это внутренний мир человека, он описывается в языке как пространство, существующее отдельно от этого человека.

- (15) In dies Priestertum, in diesen Hochmut, in diese Geistigkeit *hinein* hatte sein Ich sich verkrochen, dort saß es fest und wuchs, während er es mit Fasten und Buße zu töten meinte. (H. Hesse. Siddhartha)  
В это-то священничество, в это-то высокомерие и духовность и заползло его Я, там оно засело крепко и росло, в то время как он воображал, что умерщвляет его постом и покаянием.
- (16) Mit beginnendem Schlaf versenkte er sich tiefer und tiefer in sich *hinein* und hielt triumphalen Einzug in seiner inneren Festung, auf der er sich ein geruchliches Siegesfest erträumte, eine gigantische Orgie mit Weihrauchqualm und Myrrhendampf, zu Ehren seiner selbst. (P. Süskind. Das Parfum)  
Засыпая, он все глубже погружался в самого себя и совершал триумфальный въезд в свою внутреннюю крепость, где он

<sup>7</sup> У этого слова есть еще и временное значение; ср. словарную статью из НБНС, приведенную ниже.

праздновал в мечтах некую победу обоняния, гигантскую оргию с густым дымом ладана и парами мирры в свою честь.

Рассмотренные здесь особенности немецких лексических единиц с дейктическим значением должны учитываться в их лексикографическом представлении. В первую очередь это касается необходимости переориентации субъекта дейкиса с говорящего (как это традиционно делалось в словарях) на наблюдателя и сходных с ним участников ситуации, которым говорящий как бы делегирует свои дейктические полномочия. В «Новом большом немецко-русском словаре» мы попытались учесть эти особенности следующим образом:

**herein** *adv det* сюда; внутрь (*при обозначении перемещения из более открытого в более замкнутое пространство по направлению к говорящему или лицу, точку зрения которого принимает говорящий*); ~! **входи(те)!**; (nug) immer ~! **входи(те) быстрее [смелей]!**; hier ~! **вход здесь!** ◇ ~, **wenns kein Schneider ist!** *шутл. разг.* ≈ **входи(те), коль не шутите!**

**herein** = *отд. первый частотный элемент гл., указывает на движение внутрь по направлению к говорящему или лицу, точку зрения которого принимает говорящий (тж. перен.)*: hereinkommen **входить**; hereinlassen **впустить**; hereinfliegen **попасть впросак**

**hinein** *adv det* **1.** в, внутрь; вглубь; вплоть до (*при обозначении перемещения из более открытого в более замкнутое пространство по направлению от говорящего или лица, точку зрения которого принимает говорящий*); ins Zimmer ~ **в комнату**; ins Leere ~ **в пустоту**; nun aber ~ ins Bett (mit euch)! **а ну, в кровать!** **2.** *при обозначении времени до, вплоть до*; bis in die Gegenwart ~ **вплоть до наших дней**; bis (tief) in die Nacht ~ **до поздней ночи** ◇ **ins Blaue** ~ *разг.* **наобум, без четкой цели**; **bis ins Herz [Mark, Innerste]** ~ **высок. до глубины души**

**hinein** = *отд. первый частотный элемент гл., указывает на движение внутрь по направлению от говорящего или лица, точку зрения которого принимает говорящий (тж. перен.)*: hineingehen **входить**; hineinblicken **заглядывать (напр. в помещение)**; sich hineindenken **вдумываться**; hineinlegen **привносить (личное отношение в какое-л. дело, напр. чувство в исполнение музыкального произведения)**

### 1.4. Составление поисковых запросов в параллельных корпусах НКРЯ

Для того, чтобы открыть параллельные корпуса НКРЯ, необходимо зайти на сайт Национального корпуса русского языка по адресу <<http://ruscorpora.ru/index.html>> и нажать ссылку «Поиск в корпусе»:

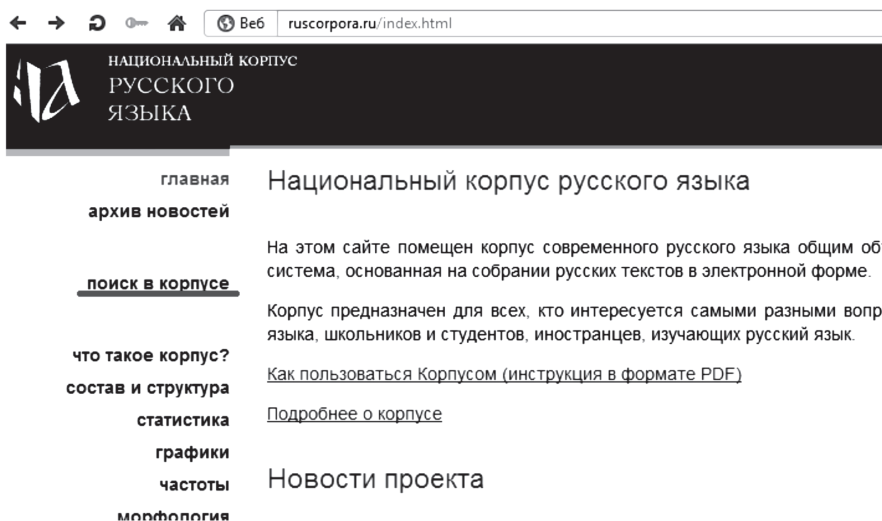


Рис. 1

В появившемся списке корпусов необходимо выбрать «Параллельный» и указать интересующий язык, например немецкий:

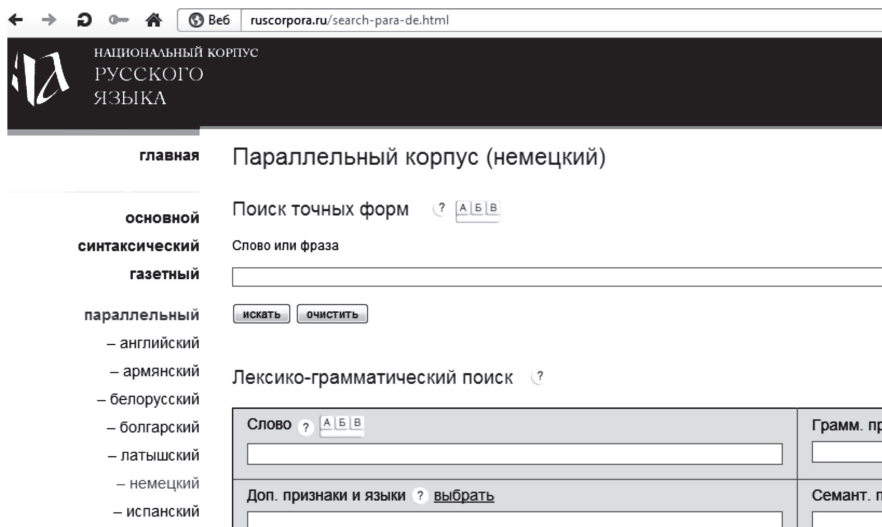


Рис. 2

Далее следует задать подкорпус. Данный шаг можно опустить, если для исследователя не принципиально, будет ли осуществляться поиск по оригинальным текстам или переводным. В ином случае необходимо в правом верхнем углу экрана пройти по ссылке «задать подкорпус». В появившемся списке можно выбрать ряд параметров текста, таких как название текста, автор, год его рождения, год создания, переводчик. Там же можно выбрать язык оригинала, например немецкий, и язык перевода — русский. Кроме того, можно указать сферу функционирования текста, выбрав между художественной, обиходно-бытовой, официально-деловой, производственно-технической, учебно-научной, церковно-богословской сферами, публицистикой, рекламой и электронной коммуникацией:

<b>главная</b>	<b>Параллельный корпус: выбор подкорпуса</b>	
<b>основной</b>	Вы можете задать подмножество корпуса, по которому в дальнейшем будет вести	
<b>синтаксический</b>	Корпус: <input checked="" type="radio"/> двуязычный <input type="radio"/> мультязычный	
<b>газетный</b>	Название	* <input type="text"/>
<b>параллельный</b>	Автор текста	<input type="text"/>
<b>обучающий</b>	Год рождения: от <input type="text"/>	до <input type="text"/> <input type="checkbox"/> Точное вхождение
<b>диалектный</b>	Год создания: от <input type="text"/>	до <input type="text"/> <input type="checkbox"/> Точное вхождение
<b>позтический</b>	Переводчик <input type="text"/>	
<b>устный</b>	Язык оригинала: <input type="text" value="любой"/>	
<b>акцентологический</b>	Язык перевода: <input type="text" value="любой"/>	
<b>мультимедийный</b>	Дата перевода: от <input type="text"/>	
<b>исторический</b>	до <input type="text"/>	
<b>использование корпуса</b>	Сфера функционирования: <a href="#">выбрать</a>	

Рис. 3

Выбор подтверждается кнопкой «Далее», после чего пользователь видит, какие тексты удовлетворяют его параметрам:

Результаты поиска в параллельном корпусе

[перейти на страницу поиска](#)

Объем всего корпуса: 1 506 документов, 4 142 533 предложения, 54 028 815 слов.

Найдено 32 документа общим объемом 154 087 предложений, 2 426 750 слов.

[Сохранить подкорпус и перейти к странице поиска](#)

Страницы: 1

1. [Mitsubishi Europe. Speicherprogrammierbare Steuerungen Bedienungsanleitung \[ABBYY LingvoPRO\] \(2012\)](#) [омонимия]
2. [Michael Ashelm. Wer kauft den Sport? Im Reich der Milliardäre \[ABBYY LingvoPRO\] \(2010\)](#) [омонимия не снята]
3. [berlin.de. Die unterschätzte Gefahr: Frauenkrankheit Herzinfarkt \[ABBYY LingvoPRO\] \(2005-2010\)](#) [омонимия не снята]
4. [Diana Engel. Kinder leiden mit: Rat und Hilfe bei häuslicher Gewalt \[ABBYY LingvoPRO\] \(2009\)](#) [омонимия не снята]
5. [nrw.de. Hygiene im Alltag. Influenza im Keim ersticken \[ABBYY LingvoPRO\] \(2009\)](#) [омонимия не снята]
6. [Sondervorlage der Moskauer deutschen Zeitung: Landwirtschaft \[ABBYY LingvoPRO\] \(2009\)](#) [омонимия не снята]
7. [Bundeszentrale für gesundheitliche Aufklärung \(BZgA\). Merkblatt für junge Eltern \[ABBYY LingvoPRO\] \(2004\)](#) [омонимия не снята]
8. [Friedrich Dürrenmatt. Justiz \(1985\)](#) [омонимия не снята]
9. [Patrick Süskind. Das Parfum: Die Geschichte eines Mörders \(1985\)](#) [омонимия не снята]

*Рис. 4*

Если результаты удовлетворяют пользователя, ему необходимо выбрать «Сохранить подкорпус и перейти к странице поиска». Поиск может осуществляться по словоформам, леммам и по определенным грамматическим и семантическим характеристикам слова. Для словоформ используется «Поиск точных форм», где также можно вводить фразы, как показано в примере:

<p><b>главная</b></p> <hr/> <p><b>основной</b></p> <hr/> <p><b>синтаксический</b></p> <hr/> <p><b>газетный</b></p> <hr/> <p><b>параллельный</b></p> <ul style="list-style-type: none"> <li>– английский</li> <li>– армянский</li> <li>– белорусский</li> <li>– болгарский</li> <li>– латышский</li> <li>– немецкий</li> <li>– польский</li> </ul>	<p><b>Параллельный корпус</b></p> <p><b>Поиск точных форм</b> ? <input type="text" value="A B V"/></p> <p>Слово или фраза</p> <p><input type="text" value="den Gästen"/></p> <p><input type="button" value="ИСКАТЬ"/> <input type="button" value="ОЧИСТИТЬ"/></p> <p><b>Лексико-грамматический поиск</b> ?</p> <p>Слово ? <input type="text" value="A B V"/></p> <p><input type="text"/></p> <p>Доп. признаки и языки ? <input type="button" value="выбрать"/></p>
---	--

*Рис. 5*

По такому запросу программа выдаст все тексты подкорпуса, в которых за поисковым объектом **den** следует **Gästen**<sup>8</sup>. Если нажать на

<sup>8</sup> В данной статье поисковые запросы для удобства читателя выделены полужирным шрифтом.



знак вопроса, можно получить справку о том, что подразумевается под «поиском точных форм». Нажав на знак «АВВ», можно вызвать виртуальную клавиатуру. По результатам данного запроса было найдено 7 контекстов<sup>9</sup>:

Найдено 3 документа, 7 вхождений.

Поискать в других корпусах: [основном](#), [акцентологическом](#), [газетном](#), [диалектном](#), [мультимедийном](#), [обучающем](#)

Страницы: 1

1. [Thomas Mann. Der Zauberberg \(1924\)](#) [омонимия не снята] **Все примеры (3)**

de	Übrigens sprang er, bei aller Wichtigkeit, die er der Elbregulierung beimaß, gleich wieder ab von diesen Gästen erzähle, was auch bereitwillig geschah, da Joachim froh war, sich erleichtern und mitteilen zu können.
ru	Впрочем, несмотря на всю важность, которую Ганс Касторп придавал вопросу об Эльбе, о подробней относительно жизни "здесь наверху", а также о пациентах санатория, за что тот впечатлениями. [Томас Манн. Волшебная гора (В. Станевич, 1959)] [омонимия не снята] $\leftarrow \dots \rightarrow$

Рис. 6

Если пользователь желает расширить контекст, ему необходимо нажать на знак  $\leftarrow \dots \rightarrow$ , находящийся после контекста. Указание [омонимия не снята] означает, что найденный контекст может содержать форму, омонимичную искомой. Поиск по текстам со снятой грамматической омонимией возможен только в основном корпусе и отсутствует в параллельных корпусах.

Помимо поиска по словоформам имеется лексико-грамматический поиск, в котором можно задавать определенные грамматические и/или семантические характеристики слова и/или словоформы, а также указывать их последовательность и интервал между ними. В графу для первого слова вводится первый поисковый объект, в данном случае лексема **der**, во вторую графу вводится второй поисковый объект, лексема **Gast**, и указывается максимальное расстояние между ними, от 1 до 2, то есть между лексемами может быть максимум одно слово:

<sup>9</sup> Под термином «контекст» здесь подразумевается отрывок текста, содержащий искомый объект, в том размере, в котором его можно экспортировать из электронного корпуса. Немецким эквивалентом термина «контекст» в этом значении является слово *Beleg*.

## Лексико-грамматический поиск ?

Слово ? А Б В

der

Доп. признаки и языки ? Выбрать

Расстояние: от  до  ?

Слово ? А Б В

Gast

Доп. признаки и языки ? Выбрать

Рис. 7

По такому запросу было найдено 87 контекстов, в числе которых есть и следующий:

- (17) *de Die Gäste, erst wie gebannt, waren aufgesprungen.* [Friedrich Dürrenmatt. Justiz (1985)] [омонимия не снята] ←...→  
 ru Гости, поначалу окаменевшие, вскочили со своих мест. [Фридрих Дюрренматт. Правосудие (С. Фридлянд, 1988)] [омонимия не снята] ←...→

Для лексем можно также выбирать определенные грамматические признаки и дополнительные параметры. Для выбора грамматических признаков необходимо пройти по соответствующей ссылке и в открывшемся меню выбрать интересующие параметры:

## Лексико-грамматический поиск ?

Слово ? <span>А Б В</span>	Грамм. признаки ? <u>Выбрать</u>
<input type="text"/>	NUM <input type="text"/>
Доп. признаки и языки ? <u>Выбрать</u>	Семант. признаки ? <u>Выбрать</u>
<input type="text"/>	<input type="text"/>

Рис. 8


По результатам выбора признаков формируется соответствующий поисковый код. Так, для числительных этот код будет «NUM».

В дополнительных признаках можно выбрать язык, к которому принадлежит искомый объект, его положение относительно знаков

препинания, позицию в начале или конце предложения, а также написание с заглавных букв.

Поиск может осуществляться не только по конкретным лексемам и их грамматическим признакам, но и исключительно по грамматическим признакам. Так, в первом поле для поиска можно указать только грамматические признаки, например числительное, во втором поле поиска — конкретную лексему, например **Gast**. По этому запросу получаем в том числе следующий контекст:

- (18) de Im Guckfenster der Tür erschienen zwei große, schwarze Augen, sahen ein Weilchen die *zwei Gäste* an und verschwanden; die Tür öffnete sich aber nicht. [Franz Kafka. Der Prozess (1914)]  
[омонимия не снята] ←...→  
ru В дверном окошечке показались два больших темных глаза, взглянули на посетителей и снова исчезли, но дверь так и не отворилась. [Франц Кафка. Процесс (Р. Райт-Ковалева, 1965)]  
[омонимия не снята] ←...→

По умолчанию в лексико-грамматическом поиске имеются поля для двух поисковых объектов. Если требуются дополнительные поля, необходимо нажать на стрелку справа от граф с грамматическими и семантическими признаками  (см. рис. 8).

Важно иметь в виду, что поисковые объекты будут располагаться в тексте в том порядке, в котором располагаются их поисковые поля. Так, за поисковым объектом из первого поля следует поисковый объект из второго поля и т. д.

Помимо прочего, в корпусах есть возможность использовать оператор \*, который заменяет неограниченное число символов. Если в лексико-грамматическом поиске задать запрос **Gast\***, можно получить контексты с такими словоформами как *Gastfreundschaft*, *Gastgeber*, *Gaststätten*, *Gasthof* и т. д.

Кроме того, корпуса снабжены возможностью исключать из результатов поиска те или иные словоформы или грамматические характеристики. Для этого в поле поиска перед ними ставится знак минус. Так, если в первое поисковое поле ввести запрос **-Ding**, во второе **an**, а в третье **sich**, то из результатов поиска (всех контекстов с сочетанием *an sich*) будут исключены контексты с сочетанием *Ding an sich*.

Более подробную инструкцию по использованию параллельных корпусов можно найти, пройдя по ссылке «инструкция», которая расположена в окне поиска в верхней правой части экрана.

\* \* \*

Завершая раздел, посвященный немецко-русским и русско-немецким текстам в Национальном корпусе русского языка, отметим,

что обращение к параллельным корпусам помогает выявить нетривиальные семантические и дискурсивные особенности как немецких слов, так и их русских функциональных соответствий.

## 2. Мангеймский корпус немецкого языка

### 2.1. Состав корпуса

Следующий корпус, о котором пойдет речь, с 2004 года носит официальное название Deutsches Referenzkorpus (DeReKo). Он поддерживается Институтом немецкого языка (IDS) в Мангейме, и поэтому известен также как Мангеймский корпус. Мангеймский корпус — это самое крупное в мире собрание немецкоязычных текстов, предназначенных для лингвистических исследований. Его полный объем составляет приблизительно 8,9 миллиардов словоформ (это соответствует 22,2 миллиону книжных страниц). Однако некоторые архивы в составе корпуса защищены авторским правом и доступны только с компьютеров Института немецкого языка. Объем части корпуса, находящейся в свободном доступе, — чуть более 6,5 миллиардов словоформ.

Работа с Мангеймским корпусом осуществляется с помощью системы COSMAS II (Corpus Search, Management and Analysis System). Портал COSMAS II <<http://www.ids-mannheim.de/cosmas2>> включает в себя несколько элементов:

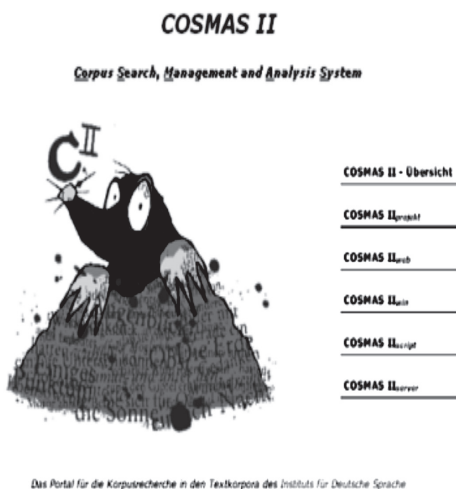


Рис. 9

COSMAS II<sub>web</sub> — приложение для работы с корпусом через портал COSMAS II, то есть прямо в окошке браузера (без предварительной установки каких-либо программ на своем компьютере);

COSMAS II<sub>win</sub> — программа, которую пользователь может скачать и установить на своем компьютере, чтобы впоследствии обращаться к корпусу с ее помощью (а не через браузер). Это приложение предназначено только для операционной системы Windows, и для работы с ним необходимо подключение к сети Интернет, поскольку сами тексты корпуса не сохраняются на компьютере пользователя (допускается лишь экспортирование результатов поиска в корпусе, о котором речь пойдет ниже). Приложения COSMAS II<sub>web</sub> и COSMAS II<sub>win</sub> обладают сходными функциональными возможностями и различаются в основном по графическому оформлению;

COSMAS II<sub>script</sub> — приложение, предназначенное для работы с операционной системой SOLARIS и доступное только для сотрудников Института немецкого языка;

COSMAS II<sub>projekt</sub> и COSMAS II<sub>server</sub> — страницы, содержащие информацию о корпусе, его составе и технических возможностях.

Доступ к корпусу предоставляется бесплатно, но требует регистрации на странице <<http://www.ids-mannheim.de/cosmas2/projekt/registrierung>>. После заполнения регистрационных данных на адрес электронной почты пользователя приходит уведомление с паролем, который затем необходимо указывать при входе в корпус (через приложение COSMAS II<sub>win</sub> или через страницу <<https://cosmas2.ids-mannheim.de/cosmas2-web/menu.registration.login.do>>).

После ввода логина и пароля для осуществления поиска в корпусе при использовании COSMAS II<sub>web</sub> необходимо выбрать в верхней строке меню пункт Recherche. В приложении COSMAS II<sub>win</sub> переход к поиску происходит автоматически.



### Recherche

Das normale Vorgehensmodell bei einer Recherche ähnelt einem Wasserfallmodell. Die einzelnen Schritte sind unten dargestellt. Selbstverständlich können sie im Modell zurückspringen. Änderungen in voran gestellten Aktionen wirken sich aber auf nachstehende aus. So löscht die Auswahl eines neuen Archivs beispielsweise die frühere Auswahl eines Korpus, sodass sie ein neues Korpus wählen müssen.

1. Archiv wählen

↳ 2. Korpus wählen

↳ 3. Anfrage formulieren

*Рис. 10*

Работа с корпусом начинается с выбора того массива текстов, в котором будет осуществляться поиск. Корпус включает в себя различные типы письменных текстов. Институт немецкого языка в Мангейме занимается также разработкой корпуса устной речи (Archiv für

Gesprochenes Deutsch), однако в состав DeReKo он не входит и доступен на отдельной странице — <<http://agd.ids-mannheim.de/index.shtml>>.

Большинство текстов, входящих в корпус DeReKo, представляют период с середины XX века до настоящего времени. В отличие, например, от Британского Национального корпуса, DeReKo не сбалансирован по типам включенных в него текстов и времени их создания, а ориентирован на охват максимально возможного объема материала. Основную часть корпуса составляют тексты газет и журналов Германии (Berliner Morgenpost, Hamburger Morgenpost, Mannheimer Morgen, Nürnberger Nachrichten и др.), Швейцарии (например, St. Galler Tagblatt, Zürcher Tagesanzeiger) и Австрии (Kleine Zeitung, Neue Kronen-Zeitung). В нем также представлены сказки братьев Гримм, сочинения Гете, статьи из интернет-энциклопедии Википедия и беллетристика XX и XXI веков.

Тексты корпуса объединены в архивы, которые делятся на подкорпусы. Так, перед началом поиска пользователю предлагается выбрать основной архив (W — Archiv der geschriebenen Sprache), включающий в себя в свободном доступе чуть более 4 миллиардов словоформ, архив исторических подкорпусов HIST, картотеку Общества немецкого языка или один из архивов, снабженных морфосинтаксической разметкой.

<b>Aktuelles Archiv</b> -	<input type="text"/>	<b>Aktuelles Korpus</b> -	<input type="text"/>
<b>Suchanfrage</b>	<input type="text"/>		
<b>Archivauswahl</b>	<b>Ergebnisse:</b>		

Öffnen Sie bitte eines der folgenden Archive:

W - Archiv der geschriebenen Sprache
Alte Version von "W - Archiv der geschriebenen Sprache" (verfügbar bis 20.12.2013)
W-ÜBRIG - Archiv der aussortierten geschriebenen Korpora
TAGGED-C - Archiv morphosyntakt. annotierter Korpora (CONNEXOR)
TAGGED-T - Archiv morphosyntakt. annotierter Korpora (TreeTagger)
TAGGED-M - Archiv morphosyntakt. annotierter Korpora (MECOLB; ehem. TAGGED)

Рис. 11

В пределах каждого архива необходимо выбрать подкорпус (см. рис. 12): либо все тексты, входящие в тот или иной архив, либо тексты из одного источника (например, подкорпус одной газеты или произведений одного автора).

Aktuelles Archiv	W - Archiv der geschriebenen Sprache	Aktuelles Korpus	-
Suchanfrage	-		
Korpusauswahl	Release: Deutsches Referenzkorpus (DeReKo-2013-II)		Ergebnisse:

Laden Sie bitte eines der folgenden virtuellen Korpora:

Vordefinierte Korpora
W-öffentlich - alle öffentlichen Korpora des Archivs W (mit Neuaquisitionen)
N-öffentlich - alle öffentlichen Neuaquisitionen
W-ohneN-öffentlich - alle öffentlichen Korpora des Archivs W (ohne Neuaquisitionen)
PP - Plenarprotokolle
bio - Biografische Literatur (öffentlich)
bmp - Berliner Morgenpost, Oktober 1997, Mai - Dezember 1998, Januar - Dezember 1999
lsp - Lexikalische Strukturen, Januar 2007 - Juli 2010

Рис. 12

Выбрав подкорпус, пользователь получает возможность задать в появившемся окне поисковый запрос. В корпусе DeReKo используется специальный синтаксис поисковых запросов, позволяющий, например, искать слово или словосочетание во всех его грамматических формах. Подробнее он будет рассмотрен в следующем разделе.

## 2.2. Синтаксис поисковых запросов

Объектом поиска в DeReKo может выступать слово в лексической и/или грамматической форме, знак препинания, служебный знак или число.

Так, при запросе **, dass ich** будет получен следующий результат<sup>10</sup>:

- (19) Eine der großen Herausforderungen dabei für mich ist, **dass ich** bleibe — Kamera, Regie und das Team wechseln, ich nicht. (Braunschweiger Zeitung, 03.01.2013)

Результат по запросу **1979**:

- (20) Dabei habe er **1979** den jungen Beck davon überzeugt, in die Politik zu gehen. (Mannheimer Morgen, 05.02.2013)

<sup>10</sup> В данной статье поисковые запросы для удобства читателя выделены полужирным шрифтом.

Поиск в DeReKo может осуществляться по определенным поисковым шаблонам с помощью операторов. К группе операторов словоформ (Wortformoperatoren) относятся три подгруппы: операторы, заменяющие некоторое количество символов (Platzhalteroperatoren); оператор леммы (Grundformoperator) и оператор игнорирования (Ignorierungsoperator).

К первой подгруппе операторов, заменяющих некоторое количество символов, относятся оператор **\***, оператор **?** и оператор **+**, которые могут комбинироваться как со словом, так и с его сегментом. Оператор **\*** заменяет неограниченное число символов, оператор **?** заменяет ровно один символ, а оператор **+** заменяет либо 0, либо один символ. Так, по запросу **Mond\*** в DeReKo число найденных словоформ превышает 6500; по запросу **Mond?** результаты намного скромнее: 32 словоформы, среди которых *Monda*, *Mondi*, *Mondo*, *Monds*; по запросу **Mond+** получено 38 словоформ, к которым добавились еще *Mond* и *Mönd* (редкая словоформа, употребленная в швейцарской газете); по запросу **Mond??** — 105 словоформ, среди которых *mondän*, *Mondes* и т. д.

Так выглядит диалоговое окно с поисковым запросом<sup>11</sup>:

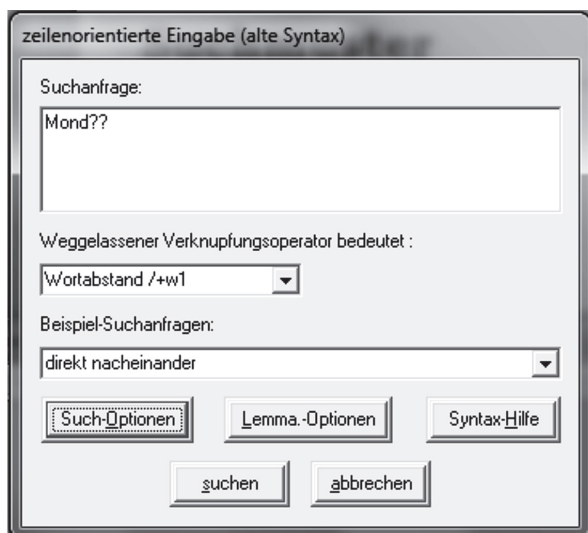


Рис. 13

<sup>11</sup> В разделе 2.2 представлены снимки экрана программы COSMAS Pwin, а не COSMAS Pweb, как в разделе 2.1, по соображениям удобства читателя. Синтаксис поисковых запросов и набор разбираемых функций в данных программах совпадают. Отличия в наборе функций между данными программами имеются, но они не релевантны для данной статьи.



Если при поиске необходимо учесть написание слова со строчной или прописной буквы, то в опциях поиска (Such-Optionen на рис. 13) необходимо выбрать опцию «Учесть написание со строчной или прописной буквы», как показано на рис. 14:

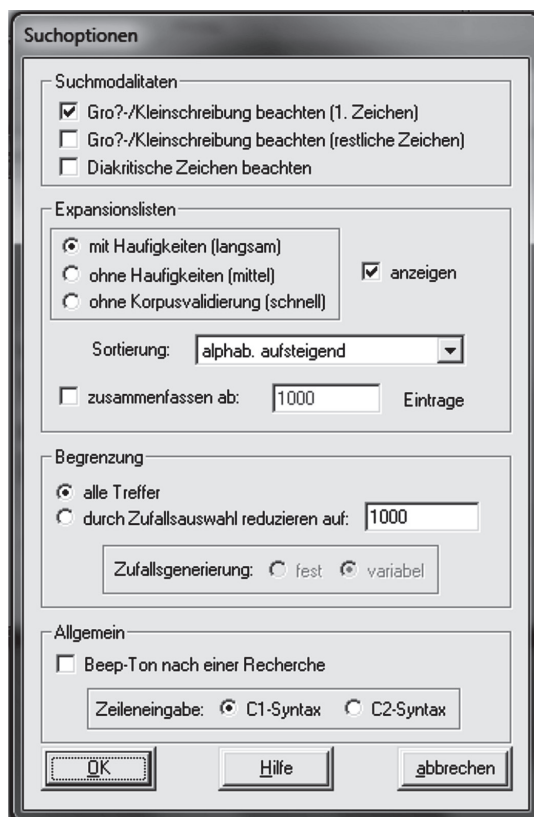


Рис. 14

Помимо операторов, заменяющих некоторое количество символов, к группе операторов словоформ (Wortformoperatoren) относятся оператор леммы (Grundformoperator) и оператор игнорирования (Ignorierungsoperator).

Оператор леммы & позволяет найти все словоформы искомого слова. Например, по запросу **&Mond** можно получить Mond, Mondes, Monds, Monde.

Оператор игнорирования \$ используется при включенной опции «Учесть написание со строчной или прописной буквы» и позволяет не учитывать написание со строчной или прописной буквы для словоформы, перед которой он стоит. Так, по запросу **Tag der \$offenen Tür** будут найдены *Tag der offenen Tür* и *Tag der Offenen Tür*.

Если поиск осуществляется по неаннотированному корпусу<sup>12</sup>, то невозможно указать искомую часть речи. Так, если ищется глагол *sein* в разных словоформах (поисковый запрос **&sein**), то одновременно будут найдены и все словоформы для соответствующего притяжательного местоимения. Убрать явно неподходящие словоформы из области поиска (например, *seiner*, *seinem*, *seines* и т. п.) можно, поставив минус рядом с соответствующими словоформами в появляющемся после отсылки запроса списке:

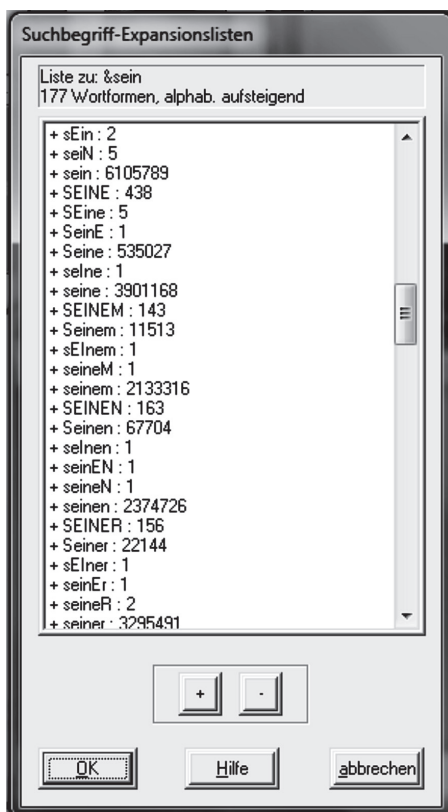


Рис. 15

Помимо этого, оператор леммы позволяет осуществлять поиск не только по всем словоформам, но и по сложным словам и дериватам. Так, при запросе **&laufen** и выбранной опции «сложные слова» (Komposita) можно найти такие слова, как *Abendlaufzeit*, *Eiskunstlaufen*; при опции «дериваты» (Ableitungen) — *abgelaufen*, *auseinanderlaufen* и т. д.

<sup>12</sup> С типологической точки зрения *her* и *hin* — это показатели глагольной ориентации; ср. [Плунгян 2011].

Подключить данные опции можно, нажав на Lemma-Optionen в диалоговом окне составления запроса и выбрав в открывшемся окне *Komposita* для сложных слов и *Sonstige Wortbildungsformen* для дериватов (рис. 16 и 17):

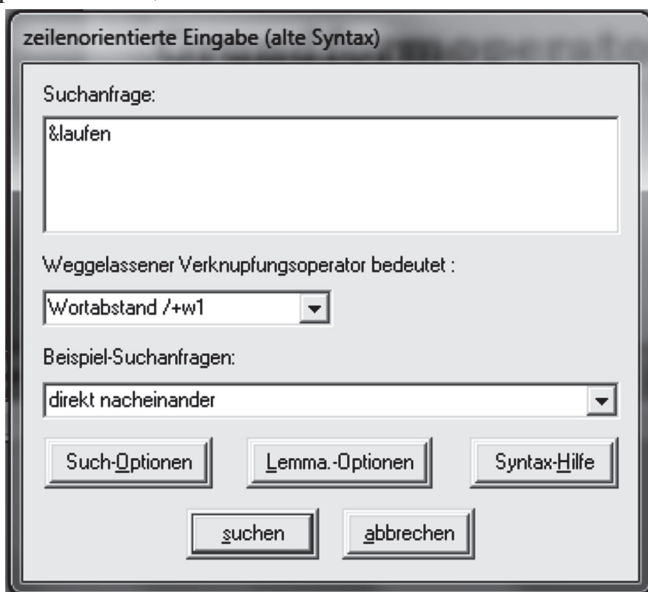


Рис. 16

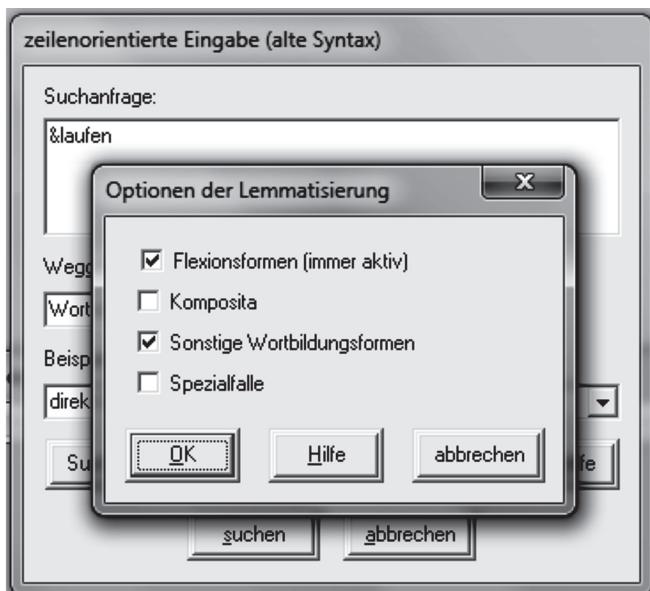


Рис. 17

Таким образом, мы рассмотрели первую группу операторов: операторы словоформ, к которым относятся операторы, заменяющие некоторое количество символов; оператор леммы и оператор игнорирования.

Помимо них, имеются логические операторы UND, ODER, NICHT, операторы дистанции и операторы аннотации.

Оператор UND позволяет искать тексты, в которых одновременно присутствуют два искомого объекта. Так, при запросе **anscheinend und scheinbar** можно получить следующий контекст:

- (21) Dabei geht es weniger um Phobien, von denen es bekanntlich unzählige gibt; von der Arachnophobie (Angst von Spinnen) bis zur Zelophobie (Angst vor Eifersucht). So rätselhaft, so herausfordernd wie die Angst ist *anscheinend* kein menschliches Gefühl. Angst vor Armut plagt den Millionär vielleicht noch mehr als den Habenicht. (Nürnberger Zeitung, 22.02.2013)

Как можно видеть, в данном отрывке присутствует только слово *anscheinend*. Слово *scheinbar* находится в этом же тексте, но не в отрывке, отображаемом в полученном контексте.

Подобным образом действуют операторы ODER и NICHT. При запросе **anscheinend oder scheinbar** ищутся все тексты, в которых присутствует либо *anscheinend*, либо *scheinbar*. При запросе **anscheinend nicht scheinbar** ищутся все тексты, где встречается словоформа *anscheinend*, но нет словоформы *scheinbar*.

Если логические операторы *und*, *oder*, *nicht*, *and*, *or*, *not* (можно использовать обозначения операторов как на немецком языке, так и на английском) становятся частью поискового запроса, то необходимо ставить их в кавычки. Например:

**Kind „und“ Kegel  
groß „oder“ klein  
„nicht“ ein wenig**

Обратим внимание, что если вместо **Kind „und“ Kegel** написать **Kind und Kegel**, то программа воспримет **und** как логический оператор и будет искать тексты, в которых одновременно присутствуют словоформы **Kind** и **Kegel**. Причем искомые объекты могут находиться в разных абзацах одного текста. Следует также обратить внимание на логический оператор *not*, который совпадает по написанию (если отключена опция «Учесть написание со строчной или прописной буквы») с немецким существительным *Not*. Чтобы программа не восприняла немецкое *Not* как логический оператор, необходимо поставить его в кавычки, как это сделано в следующем поисковом запросе:

mit Mühe „und“ „Not“

Третья группа операторов — операторы дистанции. К ним относятся оператор дистанции между словами *w* (Wortabstandsoperator), оператор дистанции между предложениями *s* (Satzabstandsoperator) и оператор дистанции между абзацами *p* (Absatzabstandsoperator). С помощью операторов дистанции указывается, каким должно быть расстояние от одного объекта поиска до другого.

Операторы дистанции вводятся косой чертой. Сразу после оператора (без пробела) указывается максимальное расстояние до искомого объекта. Так, при запросе **Sonne /w2 &scheinen** будут искаться контексты, в которых словоформа *Sonne* и *scheinen* в любой из возможных для последнего словоформ разделены не более чем одним словом, например:

- (22) Während über der katalonischen Mittelmeer-Küste die *Sonne scheint* und die Luft frühlingshaft erwärmt, war die deutsche Handball-Delegation nicht erbaut über die Gefrierschrank-Atmosphäre in der Halle. (Braunschweiger Zeitung, 12.01.2013)
- (23) Endlich *scheint* die *Sonne* wieder, und dann minus 4 Grad — genau das ideale Wetter für unsere Winterwanderung. (Braunschweiger Zeitung, 30.01.2013)

Также можно учитывать, в какой последовательности искомые объекты должны появиться в контексте. Знак + после косой линии указывает, что искомым объект находится после объекта, указанного до косой линии, знак — указывает на обратное. Так, по запросу **Sonne /+w2 &scheinen** мы получим контекст (22). В то время как по запросу **Sonne /-w2 &scheinen** возможен только контекст (23).

Возможным является также одновременно указывать минимальный и максимальный интервалы между искомыми объектами. Для этого после соответствующего оператора (*w*, *s* или *p*) через двоеточие указываются минимальный и максимальный интервалы, например запрос **Sonne /+w2:3 &scheinen** дает следующие результаты:

- (24) Die *Sonne* soll immer *scheinen*. (Braunschweiger Zeitung, 24.01.2013)
- (25) Ich wünsche mir zudem, dass die *Sonne* immer *scheint*. (Braunschweiger Zeitung, 24.01.2013)

Интервал может быть также равен нулю. Подобный интервал необходим при поиске знаков препинания, не разделенных пробелом с первым искомым объектом. Запрос **Sonne /w0** , дает следующий контекст:

- (26) Ab und zu schoben sich dicke Wolken vor die *Sonne*, doch das war den Besuchern egal: Mega-Ansturm beim Hafengeburtstag! (Hamburger Morgenpost, 12.05.2013)

Аналогично работают оператор дистанции между предложениями *s* и оператор дистанции между абзацами *p*.

Так, по запросу **Gegenwart /s1:2 Zukunft** будут найдены контексты, в которых в одном предложении имеется слово *Gegenwart*, а в соседнем предложении (до или после) или через предложение слово *Zukunft*, например:

- (27) Was viele Leute für die *Zukunft* halten, ist bereits *Gegenwart*. Und was sie für die *Gegenwart* halten, ist längst Vergangenheit. (Nürnberger Nachrichten, 16.07.2013)

Если необходимо, чтобы оба слова находились в одном предложении, интервал после оператора должен быть равен нулю. Результат по запросу **Gegenwart /s0 Zukunft**:

- (28) Thomas Bach blickt nicht gern zurück. Seine *Gegenwart* ist die *Zukunft*, und die ist seit langer Zeit systematisch geplant. (Mannheimer Morgen, 10.05.2013)

Поиск может производиться и по параграфам. Так, по запросу **Gegenwart /p1:2 Zukunft** будут получены контексты, в которых в одном параграфе имеется слово *Gegenwart*, а в соседнем или через один параграф есть слово *Zukunft*.

- (29) Prof. Georg Nagler, Rektor der DHBW Mannheim: ⟨...⟩ Ich freue mich auf Begegnungen mit der Geschichte und der *Gegenwart* — aber noch mehr auf viele Begegnungen mit den Menschen dieser Region. Prof. Udo Dahmen, Rektor der Popakademie Mannheim: Das Jahr 2013 wird für uns ein besonderes. Wir feiern das zehnjährige Jubiläum der Popakademie Baden-Württemberg. Höhepunkt wird unser Jubiläumskonzert am 19. Juli. In einer Jubiläumspublikation werden wir zeigen, was wir bisher bewegt haben und was wir in *Zukunft* bewegen wollen.

Операторы дистанции *w*, *s* и *p* можно использовать не только для включения искомого объекта в область поиска, но и для его исключения. Для этого косая линия заменяется на знак *%*. Так, если нас интересуют предложения, в которых есть слово *klipp*, но нет слова *klar*, необходимо составить следующий запрос: **klipp %s0 klar**. Если знак *%* заменить на косую линию, будут найдены все предложения, в которых одновременно встречаются *klipp* и *klar*. Результат по запросу **klipp %s0 klar**:

- (30) Was wurde sie einst romantisch besungen, die schöne Müllerin oder gar der Müller selbst, der bei Tag und Nacht stets wach war, weil die Mühle am rauschenden Bach „*klipp klapp*“ machte! (Nürnberger Nachrichten, 17.05.2013)

Если поисковый запрос содержит более одного оператора дистанции, необходимо использовать круглые скобки. Предположим, нас интересует идиома (*etw. A*) *im Auge behalten*. Тогда составляется следующий запрос: **(im /+w1 Auge) /s0 &behalten**, то есть между *im* и *Auge* дистанция равна нулю, словоформа *Auge* идет сразу за *im*. В этом же предложении должен присутствовать глагол *behalten* в любой из своих словоформ.

В поисковом запросе для данной идиомы можно учесть и возможность варьирования глагольного компонента. Тогда запрос будет выглядеть следующим образом: **(im /+w1 Auge) /s0 (&behalten OR &haben)**. В отличие от предыдущего запроса, он содержит указание на то, что в одном предложении с предложной группой *im Auge* должен присутствовать либо глагол *behalten*, либо глагол *haben* в любой из своих словоформ.

Данный поисковый запрос можно несколько упростить. Так, не обязательно указывать, что слова идут друг за другом, с помощью оператора **/+w1**, если настройки по умолчанию не менялись (см. рис. 13). В диалоговом окне, в котором составляется поисковый запрос, по умолчанию искомые объекты следуют друг за другом. Таким образом, поисковый запрос для идиомы (*etw. A*) *im Auge behalten* будет выглядеть следующим образом: **(im Auge) /s0 (&behalten OR &haben)**.

К четвертой группе операторов относится оператор аннотации MORPH, который применяется только в корпусах, снабженных морфосинтаксической разметкой. Для этого необходимо в самом начале работы выбрать архив аннотированных корпусов TAGGED-M, как показано на рис. 18:

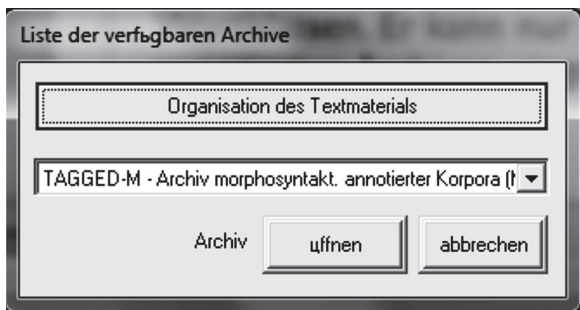


Рис. 18

В таком корпусе искомому объекту можно приписать определенные грамматические характеристики. Для этого сначала указывается

ся искомая словоформа, затем ставится /w0 и в диалоговом окне выбирается оператор MORPH, как показано на рис. 19:

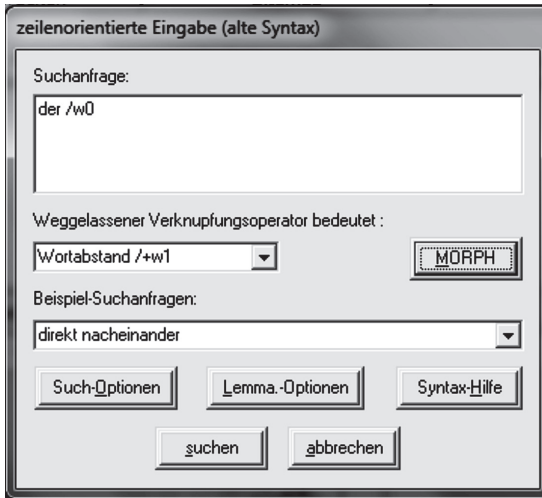


Рис. 19

Далее из списков выбираются грамматические характеристики искомого объекта. Например, можно указать, что искомым объект *der* должен быть относительным местоимением женского рода, стоящим в дательном падеже:

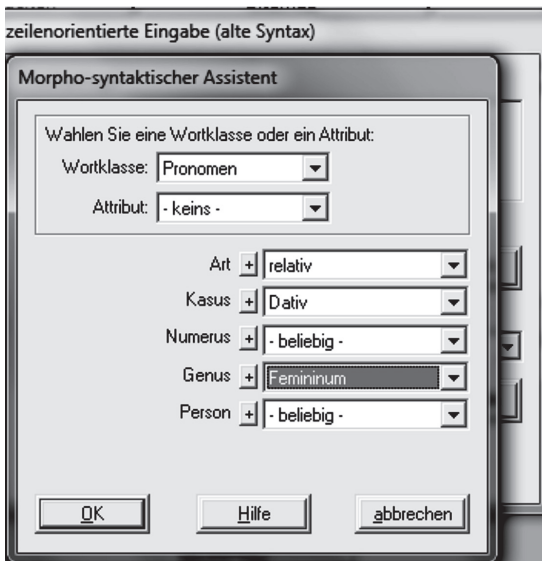


Рис. 20



Программа формирует соответствующий поисковый код, и мы получаем следующий поисковый запрос: **der /w0 MORPH(PRN rel dat)**. Пример из DeReKo по данному запросу:

- (31) Eine solche Koalition, mit *der* einige Vertreter von CDU wie Grünen durchaus liebäugeln, könnte zum Sprengsatz an der alternativen Basis werden. (Mannheimer Morgen, 07.04.1991)

Однако следует отметить, что морфосинтаксическая разметка корпусов DeReKo не свободна от ошибок, поэтому возможен информационный шум.

Поиск может осуществляться также по частям речи. Предположим, что нас интересует предложение, в котором за существительным среднего рода в единственном числе и дательном падеже следует глагол в действительном залоге, претерите или перфекте изъявительного наклонения. После выбора соответствующих характеристик (см. рис. 21, 22) программа формирует следующий запрос: **MORPH(NOU dat sg) /w1 MORPH(VRB npas past indic)**, по которому можно найти такой контекст:

- (32) An jedem *Hof war* es Aufgabe des Herolds, die Identität eines Schildträgers sofort bei seiner Ankunft aus seiner Kenntnis der „Wappen“ auf dem Schild zu verkünden. (Mannheimer Morgen, 07.01.1991)

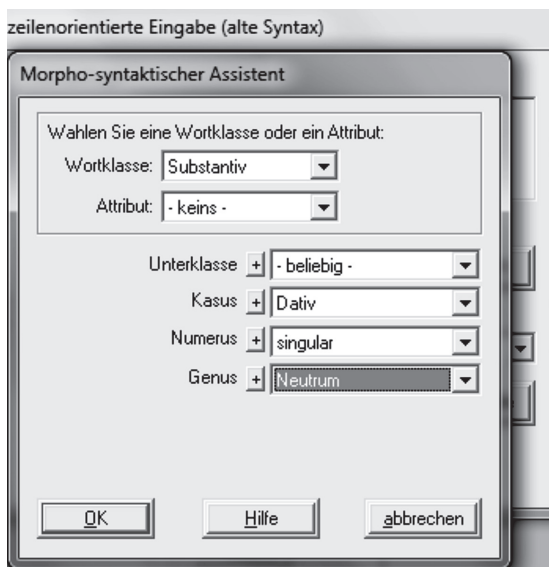




Рис. 22

Почитать подробнее о синтаксисе поисковых запросов можно также на сайте Института немецкого языка в Мангейме; см. [Syntax der zeilenorientierten Suchanfragesprache].

### 2.3. Особенности синтаксиса поисковых запросов DeReKo в сравнении с DWDS

Как было указано выше, корпуса DeReKo не являются сбалансированными, и в большинстве из них отсутствует морфосинтаксическая разметка. В то же время, на сайте Электронного словаря немецкого языка DWDS представлен сбалансированный и аннотированный корпус немецкого языка XX и XXI веков (Kernkorpus des 20./21. Jahrhunderts), который содержит около ста миллионов словоформ и по объему значительно уступает DeReKo. Данный корпус не единственный, имеющийся на сайте DWDS. Помимо него имеется ряд других корпусов, сбалансированных и несбалансированных, полный список которых представлен по ссылке <<http://www.dwds.de/ressourcen/korpora/>>.

Для работы со сбалансированными корпусами необходимо на выше указанной странице выбрать справа от строки поиска в меню «DWDS-Standardsicht» строку «Referenzkorpora»:

Anmelden | Registrierung beim DWDS



Search bar with a magnifying glass icon and a dropdown menu.

- DWDS Referenzkorpora ▾
- DWDS Standardsicht
- DWDS Wörterbuchsicht
- DWDS Referenzkorpora
- DWDS Zeitungskorpora
- DWDS Statistiken
- Retrodigit. Wörterbücher
- DWDS Wortprofil
- WDG-Vergleich

**Ressourcen - K**

- 0. Vorbemerkung
- 1. Referenzkorpora
- 2. Zeitungskorpora
- 3. Spezialkorpora
- 4. Einige Zahlen zu den

**0. Vorbemerkung**

Gefördert von der Deutschen Forschungsgemeinschaft mit dem Projekt Textbasis erstellt. Hierbei entstand das DWDS-Kernkorpus, das erst

*Рис. 23*

При поисковом запросе **zu Gast** программа выдаст все контексты, содержащие данную последовательность поисковых объектов в сбалансированных корпусах DWDS. Наибольшее количество контекстов предоставляет, как правило, сбалансированный и аннотированный корпус немецкого языка XX и XXI веков:

The screenshot shows the DWDS search interface with the query "zu Gast". The search results are displayed in three panels:

- Kernkorpus 20 (eingeschränkte Version):** Shows 10 search results (Treffer: 293, davon anzeigbar: 251). The first result is: "1 ...den dann bei einem Herzog zu Gast, der zum Schein mi...".
- Kernkorpus 21:** Shows 1 search result (Treffer: 1): "1 ...den Husumer Theodor Storm zu Gast hatte und beide na...".
- Regionenkorpus (C4):** A message box stating: "Dieses Panel steht nur angemeldeten Nutzern zur Verfügung. Sie können sich hier anmelden. Sie haben noch keinen Zugang? Hier können Sie sich kostenlos registrieren."
- DDR-Texte:** A message box stating: "Dieses Panel steht nur angemeldeten Nutzern zur Verfügung. Sie können sich hier anmelden."

*Рис. 24*

Языки поисковых запросов в DeReKo и DWDS, к сожалению, не совпадают. Основные отличия при поиске по словоформам представлены в следующей таблице:

DeReKo	DWDS
&Arzt	Arzt
Arzt	@Arzt
Arzt*	Arzt*
*arzt	*arzt
&gut Arzt	“guter Arzt”
das &gut &Beispiel	“das gute Beispiel”

Рис. 25

Далее представлены основные различия при поиске с операторами дистанции и логическими операторами:

DeReKo	DWDS
Kanzler /w2 Schröder	“Kanzler #1 Schröder”
Kanzler OR Schröder	Kanzler    Schröder
Kanzler NOT Schröder	Kanzler && !Schröder

Рис. 26

В аннотированных корпусах DWDS также возможен поиск по грамматическим характеристикам слова. Для этого используется оператор **with \$p=**. Так, если необходимо найти существительное Kohl как имя нарицательное в данной конкретной словоформе, составляется поисковый запрос: @Kohl with \$p=NN. Где NN — сокращение, принятое для нарицательных существительных (от normales Nomen). Неудобство по сравнению с поиском в DeReKo состоит в том, что требуемые сокращения необходимо вносить самому. Посмотреть их можно в таблице тэгов STTS [STTS Tag Table]. По вышеуказанному поисковому запросу получаем, к примеру, следующий результат:

- (33) Dass er dies alles freilich nicht sah, nicht mit Augen sah, sondern mit seiner immer schärfer und präziser riechenden Nase erwitterte: die Raupe im *Kohl*, das Geld hinterm Balken (...). (Süskind, Patrick, Das Parfum, Zürich: Diogenes 1985)

Описание языка поисковых запросов с примерами на сайте DWDS представлено в соответствующем кратком обзоре: [Kurzübersicht zur Suche mit der Suchmaschine DDC].

#### 2.4. Результаты поиска в корпусе (на примере DeReKo)

Результаты поиска в корпусе DeReKo отображаются в двух форматах — «ключевое слово в контексте» (KWIC, Key Word in Context) и полнотекстовые примеры (Gesamt-Volltext).

KWIC — это фрагмент контекста длиной в одну строку, в середине которого находится выделенное цветом искомое слово или сочетание слов.

The screenshot shows the DWDS search results page. At the top, there are search criteria: 'Gesamt-KWIC' and 'Gesamt-Volltext'. Below this is a table of search results. The table has columns for 'Auswahl aktiv.', 'Alle aktiv.', and 'Alle deaktiv.'. The results are numbered 1 through 7. Each result shows a date and a snippet of text with the search term highlighted in yellow. The search term is 'Suchanfrage'.

Auswahl aktiv.	Alle aktiv.	Alle deaktiv.
<input type="checkbox"/>	M13/MAL.03218	aufkündigen, sollten die Verbeeinnahmen pro Suchanfrage weniger als 40 Prozent des Google-Du
<input type="checkbox"/>	M13/MAL.03218	stets verfehlt worden. Yahoo nehme jetzt pro Suchanfrage sogar weniger ein als vorher bei den
<input type="checkbox"/>	M13/JUL.06357	ibt ein einfaches Beispiel: „Bei einer Suchanfrage etwa über Google wird nacheinander j
<input type="checkbox"/>	NON13/APR.07356	itätsfeststellung zur Polizeieinspektion. Die Suchanfrage ergab, dass der 15-Jährige in der SI
<input type="checkbox"/>	NUN13/APR.00317	Die Antwort ist die Sportstättenbörse. Eine Suchanfrage hört sich dann so an: „Gesucht: Spor
<input type="checkbox"/>	A08/FEB.01854	armen. Laut Experten verbraucht eine einzige Suchanfrage bei Google so viel Strom wie eine 10
<input type="checkbox"/>	A08/SEP.07645	seiten abrutt. So verursacht bei Google eine Suchanfrage über mobile. google.ch lediglich 15

Рис. 27

Длина KWIC (количество слов до и после искомого) может регулироваться в разделе «Опции». Программа снабжена также функцией сортировки KWIC по алфавиту или по хронологическому принципу. Даже беглый просмотр KWIC позволяет сделать некоторые полезные выводы, например о (не)употребительности устойчивых сочетаний слов, существовании незарегистрированных в словаре вариантов или новых фразем.

Так, просмотр KWIC к лексеме *Affenzahn* позволяет установить, что зарегистрированная в [БНС] идиома *einen Affenzahn drauf haben* = *нестись с бешеной скоростью* встречается достаточно редко — всего в 10 из 478 случаев употребления в корпусе лексемы *Affenzahn*. В то же время в большинстве найденных контекстов искомое слово является компонентом другого устойчивого выражения, не зареги-

стрированного ни в [БНС], ни в [Duden 11 2002] — *mit einem Affenzahn (etw. machen) = с дикой скоростью (делать что-л.)*<sup>13</sup>.

На основе KWIC к сочетанию слов *Bart* и *murmeln* (во всех грамматических формах) можно убедиться в том, что указанная в [Duden 11 2002] фраза *sich in den Bart murmeln (etw. A)* устарела: в корпусе она встречается всего в 9 контекстах, четыре из которых представляют собой фрагменты сказок братьев Гримм.

Наконец, при просмотре KWIC к *an den Mann bringen (etw. A)* выясняется, что помимо этой, зарегистрированной в [Duden 11 2002] формы в 1600 из 7000 контекстов с данной фраземой встречается вариант *an die Frau und an den Mann bringen (etw. A)*.

Кроме того, при просмотре KWIC могут быть сделаны выводы о синтаксическом поведении фразем: возможности их употребления в пассивном залоге, сочетаемости с адverbiallyными или атрибутивными модификаторами.

Следующая опция, *Gesamt-Volltext*, позволяет просмотреть полные предложения с искомой языковой единицей.

The screenshot shows a web interface for KWIC search. On the left is a sidebar with navigation options: Wortformliste, Ergebnisse, Kookkurrenzanalyse, Gesamt-KWIC, Gesamt-Volltext (with sub-options: unsortiert, alphabetisch, chronologisch, zufällige Sort., nach Kook.), and Export. The main area has tabs for 'Auswahl aktiv.', 'Alle aktiv.', and 'Alle dektiv.', along with navigation arrows and 'Seite 1 von 3' and 'KWIC'. Below this is a table of search results:

	Auswahl aktiv.	Alle aktiv.	Alle dektiv.	
				NUN98/NOV.01090 NN, 13.11.1998, S. 17
1	<input type="checkbox"/>			Geplant ist außerdem, schon in naher Zukunft die Technik des sogenannten "Keyword-Advertising" einzusetzen. Durch diese neuartige Methode bekommen die Nutzer der Suchmaschine nur solche Anzeigen zu sehen, die thematisch mit Ihrer Suchanfrage zu tun haben.
				HM07/OKT.03041 MDPO, 30.10.2007, Beilage, S. 2
2	<input type="checkbox"/>			Das Portal führt eine komplett neu gestaltete Suchmaschine ein. Der User kann mit der neuen Funktion seine Anfrage in eigenen Worten formulieren, vergleichbar mit einer Kleinanzeige. Zum Beispiel: "BMW blau mit Sportsitzen". Als Ergebnis erhält er eine Liste aller BMWs, die seiner Suchanfrage entsprechen.
				Titel unbekannt (interner Fehler)
3	<input type="checkbox"/>			Ich bin sehr verwundert darüber, warum der Film eigentlich unter dem Artikel "Das zauberhafte Land" gelistet ist und selbst eine Suchanfrage bei Wikipedia nach "Der Zauberer von Oz" diesen Eintrag nicht anzeigt, sondern ausschließlich zum Kinderbuch führt! Ich persönlich kannte den Film z.B. nur unter dem Namen "Der

Рис. 28

Полнотекстовым этот режим можно назвать лишь условно, поскольку материалы, которые входят в корпус, например газетные статьи, защищены законом об авторском праве и не отображаются целиком. Однако в разделе «Опции» можно установить длину выдаваемых контекстов, достаточную для многих лингвистических выводов, — до нескольких абзацев. Отображаемые контексты снабжаются информацией об источнике и времени его создания и могут быть отсортированы по различным принципам.

<sup>13</sup> Приводимые здесь и далее данные получены в ходе работы над немецко-русским корпусным словарем «Современная немецкая фразеология», составлением которого занимается инициативная группа лингвистов под руководством Д. О. Добровольского. Примеры словарных статей представлены на сайте Института немецкого языка в рамках проекта Usuelle Wortverbindungen в рубрике Deutsch-russische Idiome online <[http://wvonline.ids-mannheim.de/idiome\\_russ/index.htm](http://wvonline.ids-mannheim.de/idiome_russ/index.htm)>.

Система COSMAS II также снабжена функцией экспорта, позволяющей сохранять на свой компьютер данные, полученные в ходе корпусного анализа (как KWIC, так и полнотекстовые примеры) в виде файла в формате rtf.

Подробное изучение контекстов способствует уточнению данных о семантике фразем, например выявлению не зарегистрированных в существующих словарях значений.

Так, для идиомы *um die Ohren hauen* (*jmdm. etw. A*) и ее вариантов в [Duden 11 2002] указано значение ‘упрекать кого-л. в чем-л.’. Однако в корпусе идиома встречается и в другом значении — ‘ошарашивать кого-л. чем-л., потрясать кого-л. чем-л.’:

- (34) Rasch, bevor sich Hollywood der Sache annimmt und bevor uns Wolfgang Petersen, Oliver Stone oder noch Schlimmere ihre Spezialeffekte *um die Ohren hauen* wollen wir noch mal daran erinnern, dass die „Odyssee“ ein Text ist. (Nach: Mannheimer Morgen, 25.01.2005)

Скорее, пока Голливуд не взялся за дело и пока Вольфганг Петерсен, Оливер Стоун или еще кто похуже не *ошарашили* нас своими спецэффектами, хотелось бы напомнить, что «Одиссея» — это все-таки текст.

При составлении двуязычного словаря анализ контекстов также позволяет более точно подобрать соответствия для фразем исходного языка (или убедиться в различиях между традиционными словарными эквивалентами в двух языках). Так, для *sein [das eigene] Licht unter den Scheffel stellen* в [БНС] предлагается в качестве эквивалента *закрыть талант в землю*. Но примеры из корпуса свидетельствуют о том, что значение немецкой идиомы связано не с неумением найти применение своим способностям, а с преуменьшением собственных достижений:

- (35) Vogelschutzgruppe Marienberg ist hocheifrig über die Tatsache, dass das ehemalige Basaltabbaugebiet offiziell unter Schutz gestellt wurde. Vorsitzender Hans Müller wertete diese Entscheidung als gemeinsamen Erfolg, ohne *das eigene Licht unter den Scheffel zu stellen*. (Nach: Rhein-Zeitung, 26.03.1997)

Группа защитников птиц Мариенберга очень обрадована тем фактом, что территория бывшего базальтового карьера официально взята под защиту. Руководитель группы Ханс Мюллер назвал это решение общим успехом, *не преуменьшая и собственных заслуг*.

Следующая функция, доступная в корпусе DeReKo — это *Koorkurrenzanalyse*, то есть анализ совместной встречаемости слов. С помощью программного обеспечения, разработанного сотру

Института немецкого языка, исходя из общего объема - корпуса подчитывается, насколько часто в нем могут встретиться два слова, причем учитывается степень статистической ожидаемости данных сочетаний. Затем показываются все существенные отклонения от этого вероятного результата — слова, которые встречаются в окружении искомого слова существенно чаще, чем должно бы было быть в среднем, или существенно реже. Результат анализа выводится в виде таблицы с указанием логарифмического отношения правдоподобия (LLR).

Suchanfrage		Ergebnisse der Kookkurrenzanalyse						Release: Deutsches Referenzkorpus (DeReKo-2013-I)		Ergebnisse: 545	
Wortformliste		<input type="checkbox"/> Gesamt-KWIC		<input type="checkbox"/> Gesamt-Volltext		<input type="checkbox"/> Export		<input type="checkbox"/> Kookkurrenzanalyse			
Ergebnisse											
Kookkurrenzanalyse											
Erstellungen											
→ Kookkurrenzen											
Gesamt-KWIC											
Gesamt-Volltext											
Export											
#	LLR	hmm1.	Häufig	links	rechts	Kookkurrenzen	syntagmatische Muster				
1	86	12	12	-2	-2	Musikindustrie erforderlich	100% Musikindustrie - Suchanfrage erforderlich In				
2	65	28	4	1	2	Everyhit	50% Charts auf Everyhit				
3	40	31	3	1	2	Theoria	100% mit Ihrer Suchanfrage - Theoria				
4	38	54	23	1	2	Google	78% eine ... Suchanfrage bei in Google				
5	38	74	20	1	1	erforderlich	100% Musikindustrie - Suchanfrage erforderlich				
6	29	77	3	1	2	kanal	100% der Suchanfrage schwarzer kanal				
7	24	80	3	1	2	Query	100% Suchanfrage [...] Query				
8	15	83	3	1	2	eingibt	100% Suchanfrage [...] eingibt				
9	14	87	4	-2	-2	Suchmaschinen	75% Suchmaschinen eine Suchanfrage				
10	10	91	4	1	2	bzw	100% Suchanfrage [...] bzw				
11	0	569	478	0	0	statistisch unspezifisch					

Рис. 29

С помощью анализа совместной встречаемости слов могут быть выявлены, например, устойчивые сочетания слов, до сих пор не зафиксированные в словарях. На использовании этой функции основан целый ряд проектов, представленных на странице *Wortverbindungen online* <<http://wvonline.ids-mannheim.de/>>.

При составлении фразеологического словаря эта функция также может быть полезной — например, для определения точной формы идиомы и точной формы леммы для словарной статьи. Например, в словаре [Duden 11 2002] представлен фразеологизм *Gott und die Welt*, имеющий значение ‘всё, что угодно, все, кто угодно’. Однако корпусный анализ позволяет убедиться в том, что *Gott und die Welt* встречается в основном с глаголами и существительными, обозначающими речевую деятельность, преимущественно с глаголом *reden*, так что в качестве леммы в соответствующей словарной статье должна фиксироваться скорее форма *über Gott und die Welt reden*.

## Литература

Добровольский 2009 — Добровольский Д. О. Корпус параллельных текстов в исследовании культурно-специфичной лексики // Нац-



- ональный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 383—401.
- Добровольский, Кретов, Шаров 2005 — *Добровольский Д. О., Кретов А. А., Шаров С.* Корпус параллельных текстов: архитектура и возможности использования // Национальный корпус русского языка: 2003—2005. Результаты и перспективы. М.: Индрик, 2005. С. 263—296.
- Добровольский, Падучева 2008 — *Добровольский Д. О., Падучева Е. В.* Дейксис в отсутствие говорящего: о семантике немецких дейктических элементов *hin* и *her* // Компьютерная лингвистика и интеллектуальные технологии. Труды междунар. конференции «Диалог 2008». М., 2008. С. 140—146.
- Йокояма 2012 — *Йокояма О. Б.* К метатеории перевода: перевод как дискурс // Человек о языке — язык о человеке. Сб. статей памяти Н. Ю. Шведовой. М., 2012. С. 152—159.
- Плунгян 2011 — *Плунгян В. А.* Введение в грамматическую семантику. Грамматические значения и грамматические системы языков мира. М.: Российский государственный гуманитарный университет, 2011.

- Atkins, Fillmore, Johnson 2003 — *Atkins S., Fillmore C. J., Johnson C. R.* Lexicographic relevance: Selecting information from corpus evidence // *International Journal of Lexicography*. Vol. 16. 2003. No. 3. P. 251—280.
- Cysouw, Wälchli 2007 — *Cysouw M. Wälchli B.* (eds.). Parallel texts. Using translational equivalents in linguistic typology. *Sprachtypologie & Universalienforschung* 60.2. (Special Issue). 2007.

### Словари

- БНС — Большой немецко-русский словарь по общей лексике / Е. И. Лепинг, Н. П. Страхова, Н. И. Филичева и др. Под общ. рук. О. И. Москальской. М.: Русский Язык-Медиа, 2004.
- НБНС — Новый большой немецко-русский словарь: В 3 т. / Под рук. Д. О. Добровольского. М.: Астрель, 2008—2010.
- Duden 11 2002 — Duden 11. Redewendungen und sprichwörtliche Redensarten / Wörterbuch der deutschen Idiomatik. Hrsg. und bearbeitet von G. Drosdowski / Stolze-Stubenrecht, W. Mannheim u.a., 2002. (=Duden, Bd 11)

### Текстовые источники и электронные ресурсы

- НКРЯ. Национальный корпус русского языка. URL: <http://www.ruscorpora.ru>
- DWDS. Das Digitale Wörterbuch der deutschen Sprache. URL: <http://www.dwds.de/>

Kurzübersicht zur Suche mit der Suchmaschine DDC. URL: <http://www.dwds.de/hilfe/suche/>

STTS Tag Table (1995/1999). URL: <http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html>

Syntax der zeilenorientierten Suchanfragesprache. URL: <http://www.ids-mannheim.de/cosmas2/web-app/hilfe/suchanfrage/eingabe-zeile/syntax/>

The Mannheim German Reference Corpus (DeReKo). URL: <https://cosmas2.ids-mannheim.de/cosmas2-web/>