

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Астраханский государственный университет»
(Астраханский государственный университет)

Кафедра английской филологии

Письменный перевод

по книге Collocation and Term Extraction Using Linguistically
Enhanced Statistical Methods

выходные данные: Йена: Университет Фридриха Шиллера, 2008

перевод стр. с 9 по 18

для сдачи кандидатского экзамена
по иностранному языку
(английский)

Выполнила:
Савенкова Ольга Сергеевна
Кафедра современного русского языка

Астрахань – 2022 г.

<p>Chapter 2 Defining Collocations and Terms</p> <p>Since the main goals of this thesis are the definition, implementation and evaluation of statistical association measures which incorporate linguistic properties of collocations and terms, it is first mandatory to substantiate in detail the characteristic features of these linguistic expressions which have been put forth in the scientific literature.</p> <p>One can imagine that the research literature on the issue of collocations and of terms is vast and that any attempt to provide an overview will necessarily have to zoom in on the main aspects, in particular within the context of a computational approach like this one.</p> <p>As the first two sections on defining the notion of collocations (section 2.1) and the notion of terms (section 2.2) will show, these two kinds of linguistic expressions have received quite different treatments in the respective research literature.</p> <p>At first sight, this is not astonishing because from a conceptual and linguistic point of view, collocations and terms denote different linguistic entities and surface in different linguistic contexts.</p> <p>What is remarkable though (and will be discussed extensively in chapter 3) is</p>	<p>Глава 2 Определение словосочетаний и терминов</p> <p>Поскольку основными целями данной диссертации являются определение, реализация и оценка показателей статистической ассоциации, включающих в себя языковые свойства словосочетаний и терминов, в первую очередь необходимо подробно обосновать характерные черты этих языковых выражений, которые были выдвинуты в научной литературе.</p> <p>Можно себе представить, что исследовательская литература по проблеме словосочетаний и терминов огромна и что любая попытка дать обзор обязательно должна будет привлечь внимание к основным аспектам, в частности, в контексте вычислительного подхода, подобного этому.</p> <p>Как покажут первые два раздела, посвященные определению понятия словосочетания (раздел 2.1) и понятия термина (раздел 2.2), эти два вида лингвистических выражений получили совершенно разное толкование в соответствующей исследовательской литературе.</p> <p>На первый взгляд это неудивительно, потому что с концептуальной и лингвистической точки зрения словосочетания и термины обозначают разные языковые единицы и обнаруживаются в разных языковых контекстах.</p> <p>Примечательно, однако (и будет подробно обсуждаться в главе 3), что</p>
---	---

<p>the fact that the computational approaches to their automatic extraction from unrestricted text have been very similar in terms of the association measures and extraction procedures applied.</p> <p>One of the insights that this chapter aims to articulate is that, in terms of linguistic discourse, the notion of collocations preferably needs to be located in the area of general, largely subject-independent language whereas the notion of terms falls into the area of domain-specific sublanguage of a certain subject field.</p> <p>Another finding is that, from a syntactic point of view, collocations surface in different kinds of syntactic expressions whereas terms are mainly confined to noun phrases.</p> <p>However, what will also become clear in the course of this chapter is that there is indeed a linguistic property – viz. the property of limited modifiability – which both collocations and terms share and which may be derived from the discussion and insights of the respective research strands (as will be described and assessed in section 2.3).</p> <p>The linguistic frame of reference within which this linguistic property may be located is the collocational layer of Firth’s model of language description.</p>	<p>вычислительные подходы к их автоматическому извлечению из неограниченного текста были очень похожи с точки зрения применяемых мер ассоциации и процедур извлечения.</p> <p>Одна из идей, которую мы пытаемся сформулировать в этой главе, заключается в том, что с точки зрения лингвистического дискурса понятие словосочетаний желательно располагать в области общего, в значительной степени независимого от предмета языка, тогда как понятие терминов попадает в область предметно-специфического подязыка определенной предметной области.</p> <p>Другой вывод заключается в том, что с синтаксической точки зрения словосочетания появляются в различных видах синтаксических выражений, тогда как термины в основном ограничиваются именными фразами.</p> <p>Однако в ходе этой главы также станет ясно, что лингвистическое свойство действительно существует – а именно свойство ограниченной модифицируемости, которое является общим как для словосочетаний, так и для терминов и которое может быть получено в результате обсуждения и понимания соответствующих направлений исследований (как будет описано и оценено в разделе 2.3).</p> <p>Лингвистическая точка зрения, в пределах которой может располагаться это лингвистическое свойство, представляет собой</p>
---	--

<p>This model – although from a historical perspective located in our discussion on collocations in section 2.1 – will help us define our linguistically motivated statistical association measures for both collocation and term extraction.</p>	<p>коллокационный слой модели описания языка Ферта. Эта модель, хотя и с исторической точки зрения, описанная в нашем обсуждении словосочетаний в разделе 2.1, поможет нам определить наши лингвистически мотивированные меры статистической ассоциации как для словосочетаний, так и для извлечения терминов.</p>
<p>What Firth's lexical-collocational layer is able to capture is the observation that the linguistic property of limited modifiability is manifested differently, i.e. while it surfaces syntagmatically in collocations, it is manifested paradigmatically in terms.</p>	<p>Лексико-коллокационный слой Ферта способен зафиксировать наблюдение, что лингвистическое свойство ограниченной модифицируемости проявляется по-разному, т. е. синтагматически проявляясь в словосочетаниях, парадигматически – в терминах.</p>
<h2>2.1 Defining Collocations</h2>	<h2>2.1 Определение словосочетаний</h2>
<p>That words in natural language are neither randomly combined into phrases and sentences nor that they are only constrained by the rules of syntax had been known by linguists for quite some time.</p>	<p>То, что слова в естественном языке не объединяются в фразы и предложения случайным образом и не ограничиваются только правилами синтаксиса, было известно лингвистам уже довольно давно.</p>
<p>Curiously, this basic fact about collocations and, at the same time, their rather diverse and apparently idiosyncratic behavior, has been taken out of focus by a substantial part of contemporary mainstream linguistics which has been primarily concerned with examining language from a theoretical perspective.</p>	<p>Любопытно, что этот основной факт, касающийся словосочетаний и, в то же время, их довольно иного и, по-видимому, идиосинкразическое поведение, был вычеркнут из поля зрения значительной частью современной господствующей лингвистики, которая прежде всего занималась изучением языка с теоретической точки зрения.</p>
<p>In particular, generative linguistics in the Chomskyan tradition (Chomsky (1965) or Chomsky (1995)) demote all</p>	<p>В частности, генеративная лингвистика в традиции Хомского (Хомский (1965) или Хомский (1995)) безопасно опускает все</p>

<p>lexical and syntactic idiosyncracies safely into the realm of the lexicon.</p> <p>By pointing out that “You shall know a word by the company it keeps!”, it is Firth (1957), who commonly gets the credit for first introducing the notion of collocation into contemporary linguistics (see also Bartsch (2004)) and who thus coined probably one of the most well-known slogans in 20th century linguistics.</p> <p>Still, as e.g. also Lehr (1996) points out, already Firth used to be rather vague about a precise definition of the concept, and hence it is not surprising that there has been a rather enormous conceptual diversity surrounding the idea of collocation in linguistic research up to today.</p> <p>Drawing a very rough dividing line, two lines of linguistic research may be identified in the last half-century and we will describe them in some detail in the first two subsections below.</p> <p>On the one hand, there is the structural-lexicographic approach which is mainly concerned with adequate representation forms of collocations within linguistic lexicons and dictionaries (subsection 2.1.1).</p> <p>On the other hand, there is the frequentist corpus-based approach to collocations which was initiated and significantly influenced by Firth’s linguistic research (subsection 2.1.2)</p>	<p>лексические и синтаксические идиосинкразии в область лексики.</p> <p>Указав, что «Вы узнаете слово по компании, которую оно составляет!», именно Ферт (1957) обычно получает признание за то, что он первым ввел понятие словосочетания в современную лингвистику (см. также Барч (2004)) и кто таким образом, придумал, вероятно, один из самых известных лозунгов в лингвистике 20-го века.</p> <p>Тем не менее, как, например, также указывает Лер (1996), Ферт уже довольно расплывчато относился к точному определению понятия, и, следовательно, неудивительно, что идея словосочетания в лингвистических исследованиях до сих пор сопровождалась огромным концептуальным разнообразием.</p> <p>Проводя очень грубую разграничительную черту, можно выделить два направления лингвистических исследований за последние полвека, и мы опишем их более подробно в первых двух подразделах ниже.</p> <p>С одной стороны, структурно-лексикографический подход, который в основном связан с адекватными формами представления словосочетаний в лингвистических лексиконах и словарях (подраздел 2.1.1).</p> <p>С другой стороны, существует частотный корпусный подход к словосочетаниям, который был инициирован и на который оказали существенное влияние</p>
---	--

<p>and which is dedicated to an empirically grounded analysis of natural language.</p> <p>As the field of computational linguistics and natural language processing is also in need of linguistic definitions, computational linguists – if their research or application task is to extract collocations from unrestricted text – typically acknowledge that there is a wide array of diverse definitions provided by the two lines of linguistic research (subsection 2.1.3).</p> <p>Besides the property of co-occurrence, however, these only have minimal or no influence in how the algorithms and procedures for an extraction task are defined, as we will also see in more detail in section 3.1 in the next chapter.</p> <p>The consequence of this is that, in general, insights about linguistic properties of collocations are not incorporated in computational implementations.</p> <p>Obviously, this constitutes one of the gaps that this thesis aims to fill. For this purpose, we will assemble and assess the linguistic properties of collocations adopted from the various linguistic research strands in subsection 2.1.4.</p>	<p>лингвистические исследования Ферта (подраздел 2.1.2) и который посвящен эмпирически обоснованному анализу естественного языка.</p> <p>Поскольку области компьютерной лингвистики и обработки естественного языка также нуждаются в лингвистических определениях, компьютерные лингвисты – если их исследовательская или прикладная задача состоит в извлечении словосочетаний из неограниченного текста – обычно признают, что существует широкий спектр разнообразных определений, предоставляемых двумя направлениями лингвистических исследований (подраздел 2.1.3).</p> <p>Однако, помимо свойства совпадения, они имеют минимальное влияние или вообще не влияют на то, как определяются алгоритмы и процедуры для задачи извлечения, что мы также увидим в деталях в разделе 3.1 в следующей главе.</p> <p>Следствием этого является то, что, как правило, понимание лингвистических свойств словосочетаний не включается в вычислительные реализации.</p> <p>Очевидно, что это один из пробелов, которые призвана заполнить данная диссертация. С этой целью мы соберем и оценим лингвистические свойства словосочетаний, заимствованных из различных направлений лингвистических исследований в подразделе 2.1.4.</p>
--	--

<p>On the one hand, we will focus on four characteristic linguistic properties of collocations which have the capacity to be algorithmically formalized from a computational perspective.</p>	<p>С одной стороны, мы сосредоточимся на четырех характерных лингвистических свойствах словосочетаний, которые могут быть алгоритмически формализованы с вычислительной точки зрения.</p>
<p>On the other hand, we will capitalize on linguistic properties that will help us to draw a linguistic demarcation line between collocations and non-collocations and also to establish different linguistic subtypes of collocations.</p>	<p>С другой стороны, мы воспользуемся лингвистическими свойствами, которые помогут нам провести лингвистическую разграничительную линию между словосочетаниями и несловосочетаниями, а также установить различные лингвистические подтипы словосочетаний.</p>
<p>2.1.1 Defining Collocations from the Lexicographic Perspective</p>	<p>2.1.1 Определение словосочетаний с лексикографической точки зрения</p>
<p>The kinds of linguists who typically have a profound interest in examining collocations and their linguistic properties are lexicologists and lexicographers. This is of course due to the fact that lexicographers have to worry about how to represent information about collocations in a linguistic dictionary or lexicon.</p>	<p>К тем лингвистам, которые обычно проявляют глубокий интерес к изучению словосочетаний и их лингвистических свойств, относятся лексикологи и лексикографы. Это, конечно, связано с тем, что лексикографам приходится беспокоиться о том, как представить информацию о словосочетаниях в лингвистическом словаре или лексиконе.</p>
<p>This subsection will therefore describe two kinds of representative strands of work in this vein. One kind, represented by Hausmann (1985) and Mel'cuk (1995a), places their lexicographic descriptions of collocations into a broader linguistic and meaning-based framework (subsubsection 2.1.1.1) whereas the other kind, represented by Benson et al. (1986b) and Benson (1989), confines itself to more or less</p>	<p>Поэтому в этом подразделе будут описаны два вида репрезентативных направлений работы в этом ключе. Один вид, представленный Хаусманном (1985) и Мельчуком (1995a), помещает свои лексикографические описания словосочетаний в более широкую лингвистическую и смысловую структуру (подраздел 2.1.1.1), в то время как другой вид,</p>

<p>“theory-free” lexicographic descriptions (subsubsection 2.1.1.2).</p>	<p>представленный Бенсоном и др. (1986b) и Бенсон (1989), ограничивается более или менее «свободными от теории» лексикографическими описаниями (подраздел 2.1.1.2).</p>
<p>2.1.1.1 The Meaning-based Lexicographic Approach to Collocations</p>	<p>2.1.1.1 Лексикографический подход к словосочетаниям, основанный на значении</p>
<p>Meaning-based approaches to collocations are characterized by their often close connection to applicative areas such as lexicography and foreign-language learning.</p>	<p>Подходы к словосочетаниям, основанные на значении, часто характеризуются их тесной связью с прикладными областями, такими как лексикография и изучение иностранных языков.</p>
<p>Mel’cuk, a prominent lexicographically oriented linguist, has embedded his approach to collocations into a complete linguistic framework, viz. Meaning-Text Theory (MTT), which attempts to account for relations between lexical items language-independently.</p>	<p>Мельчук, выдающийся лингвист, ориентированный на лексикографию, внедрил свой подход к словосочетаниям в полную лингвистическую структуру, а именно теорию смыслового текста (ТСТ), которая пытается объяснить отношения между лексическими единицами независимо от языка.</p>
<p>Within this framework, Mel’cuk (1995a) and Mel’cuk (1998) attempt to come to terms with the idiosyncrasy of collocations by embedding them into a more semantically oriented layer of description. In the Meaning-Text Theory (MTT) lexical relations are used as a means of describing so-called institutionalized lexical relations.</p>	<p>В этих рамках Мельчук (1995a) и Мельчук (1998) пытаются смириться с идиосинкразией словосочетаний, встраивая их в более семантически ориентированный слой описания. В теории смыслового текста (ТСТ) лексические отношения используются как средство описания так называемых институционализированных лексических отношений.</p>
<p>Such relations are defined as holding between two lexical items with a constant meaning linked to their combination. Although these meanings, referred to as Lexical Functions, explain</p>	<p>Такие отношения определяются как удерживание между двумя лексическими единицами постоянного значения, связанного с их сочетанием. Хотя эти значения,</p>

<p>the relations between lexical items mostly on the semantic level, phonological and syntactic descriptions are not excluded per se.</p> <p>Lexical Functions (LFs) aim at coping with the problem of lexical choices. For Mel'čuk, this boils down to go from a given semantic representation to a corresponding (deep) syntactic representation. In this process, the speaker has to select lexical units, i.e. lexical lexemes and phrasemes to build sentences.</p> <p>Although LFs are taken as a particular device to systematically describe the relations between two lexical units across various languages, they are far more encompassing than the notion of collocations. A composite formulaic notation³ including phonological, syntactic and semantic features is given to cover various syntagmatic relations between lexical items.</p> <p>Thereby, it is assumed that all languages, in different ways, realize the meanings postulated by LFs and that the main difference lies in the language-specific ways in which the combination of given lexical items is used to arrive at various LF meanings.</p> <p>There are 36 syntagmatic LFs which are distinguished by their syntactic part of</p>	<p>называемые лексическими функциями, объясняют отношения между лексическими единицами в основном на семантическом уровне, фонологические и синтаксические описания не исключаются сами по себе.</p> <p>Лексические функции (ЛФ) направлены на решение проблемы лексического выбора. Для Мельчука это сводится к переходу от данной семантической репрезентации к соответствующей (глубокой) синтаксической репрезентации. В этом процессе говорящий должен выбрать лексические единицы, то есть лексические лексемы и фраземы, для построения предложений.</p> <p>Хотя ЛФ рассматриваются как особый способ систематического описания отношений между двумя лексическими единицами в разных языках, они гораздо шире, чем понятие словосочетаний. Составная формульная нотация, включающая фонологические, синтаксические и семантические признаки, дается для охвата различных синтагматических отношений между лексическими единицами.</p> <p>Таким образом, предполагается, что все языки по-разному реализуют значения, постулируемые ЛФ, и что основное различие заключается в специфических для языка способах использования комбинации данных лексических единиц для получения различных значений ЛФ.</p> <p>Существует 36 синтагматических ЛФ, различающихся по</p>
---	--

<p>speech. Mel'čuk (1996) provides some examples and their English realizations:</p> <p>Verbal Lexical Functions:</p> <p>1. Degrad [Lat. degradare (to degrade, worsen)]</p> <p>a. Degrad (clothes) = to wear off b. Degrad (house) = to become dilapidated c. Degrad (temper) = to fray</p> <p>Adjectival Lexical Functions:</p> <p>2. Magn [Lat. magnus (big, great)]</p> <p>a. Magn (belief) = staunch b. Magn (thin[person]) = as a rake</p> <p>3. Bon [Lat. bonus (good)]</p> <p>a. Bon (aid) = valuable b. Bon (proposal) = tempting</p> <p>Nominal Lexical Functions:</p> <p>4. Centr [Lat. centrum (the center/culmination of)]</p> <p>a. Centr (crisis) = the peak (of the crisis) b. Centr (desert) = the heart (of the desert)</p> <p>As can be seen from the above examples, the semantic radar of LFs is far more extensive and comprehensive than just natural language expressions that are typically assumed to fall under the notion of collocation.</p>	<p>синтаксической части речи. Мельчук (1996) приводит несколько примеров и их английскую реализацию:</p> <p>Глагольные лексические функции:</p> <p>1. Degrad [лат. degradare (деградировать, ухудшаться)]</p> <p>а. Degrad (одежда) = изнашиваться б. Degrad (дом) = приходить в упадок в. Degrad (характер) = раздражаться</p> <p>Лексические функции прилагательных:</p> <p>2. Magn [лат. magnus (большой, великий)]</p> <p>а. Magn (вера) = стойкий б. Magn (худой [человек]) = как грабли</p> <p>3. Bon [лат. bonus (хороший)]</p> <p>а. Bon (помощь) = ценный б. Bon (предложение) = заманчивый</p> <p>Номинальные лексические функции:</p> <p>4. Centr [лат. centrum (центр/кульминация)]</p> <p>а. Centr (кризис) = пик (кризиса) б. Centr (пустыня) = сердце (пустыни)</p> <p>Как видно из приведенных выше примеров, семантический радар ЛФ гораздо более обширен и всеобъемлющ, чем просто выражения естественного языка, которые обычно считаются подпадающими под понятие словосочетания.</p>
--	--

<p>In particular, Mel'čuk's MTT is aimed at providing a complete linguistic framework for the mapping from the content or meaning of an utterance to its form or text, with collocations being one particular (i.e., idiosyncratic) lexical surface realization.</p>	<p>В частности, ТСТ Мельчука нацелена на обеспечение полной лингвистической основы для отображения содержания или значения высказывания на его форму или текст, при этом словосочетания являются одной конкретной (то есть идиосинкразической) реализацией лексической стороны.</p>
<p>The overall lexicographic goal of MTT is the creation of so-called Explanatory Combinatorial Dictionaries (ECDs) (cf. (Bartsch, 2004)) displaying the combinatorial properties of word combinations in a language.</p>	<p>Общей лексикографической целью ТСТ является создание так называемых толково-комбинаторных словарей (ТКС) (ср. (Барч, 2004)), отображающих комбинаторные свойства словосочетаний в языке.</p>
<p>In the area of German linguistics, research on collocations is founded on a completely different conceptualization, i.e. one derived from a phraseological-semantic point of view.</p>	<p>В области немецкого языкознания исследования словосочетаний основаны на совершенно иной концептуализации, т. е. выведенной с фразеологически-семантической точки зрения.</p>
<p>In particular Hausmann (1985) and Burger (2003), besides focusing on prescriptive correctness of collocational language use, categorize collocations according to the semantic specificity of their constituents.</p>	<p>В частности, Хаусманн (1985) и Бургер (2003), помимо сосредоточения внимания на предписывающей правильности использования словосочетаний, классифицируют словосочетания в соответствии с семантической спецификой их составляющих.</p>
<p>Thus, content words (i.e. verbs, adjectives, nouns) play a central role as components of collocations. The different constituents in a collocation do not have an equal status, but rather, their relationship is a directed one. The collocational base is defined as the dominant constituent while the collocate is dominated by the base.</p>	<p>Таким образом, содержательные слова (то есть глаголы, прилагательные, существительные) играют центральную роль как компоненты словосочетаний. Различные составляющие в словосочетании не имеют равноправного статуса, а их отношения носят направленный характер. Коллокационная основа определяется как доминирующая</p>

<p>In particular, the base is the semantically autonomous part, which, however, needs the collocate to obtain its full meaning. This is illustrated in the following preposition-noun-verb (PNV) collocations from German (and their English translations):</p> <p>5.</p> <p>a. “zur Verfügung stellen” (to make available)</p> <p>b. “in Erwägung ziehen” (to take into consideration)</p> <p>Here, the collocational base “Verfügung” (availability) is completed by the meaning of the collocate “stellen” (to place) and, in the English translation, the meaning of the collocational base “available” is completed by the meaning of the collocate “make”.</p> <p>Central to Hausmann’s definition of collocations is the directionality from the base to the collocate in that the base as the dominant constituent is the element which is semantically more stable and which thus exerts a stronger influence in a way that it can dominate the collocate.</p> <p>Hence, collocations consist of at least two component parts, with at least one</p>	<p>составляющая, в то время как коллокат преобладает в основе.</p> <p>В частности, основа является семантически автономной частью, которая, однако, нуждается в словосочетании для получения своего полного значения. Это иллюстрируется следующими сочетаниями предлог-существительное-глагол (ПСТ) из немецкого языка (и их английскими переводами):</p> <p>5.</p> <p>a. «zur Verfügung stellen» (сделать доступным)</p> <p>б. «in Erwägung ziehen» (принимать во внимание)</p> <p>Здесь словосочетание «Verfügung» (доступность) дополняется значением словосочетания «stellen» (помещать), а в английском переводе значение словосочетания «доступный» дополняется значением словосочетания «сделать».</p> <p>Центральное место в определении словосочетаний, данное Хаусманном, занимает направленность от основы к словосочетанию в том смысле, что основа как доминирующая составляющая является элементом, который семантически более стабилен и, таким образом, оказывает более сильное влияние таким образом, что может доминировать над словосочетанием.</p> <p>Следовательно, словосочетания состоят как минимум из двух составных частей, причем по</p>
---	--

<p>component part either having kept or lost its literal meaning.</p> <p>Another central distinction in Hausmann's conception of collocations is concerned with the degree of fixedness between the different constituents of a collocational expression.</p> <p>On the one hand, there are fixed word combinations under which mainly idioms can be found and for which the above definition for base and collocate hardly applies.</p> <p>These fixed expressions are referred to as fully idiomatic expressions in which every component is void of its literal meaning, as is exemplified by the following idiomatic expressions:</p> <p>6.</p> <ul style="list-style-type: none"> a. "ins Gras beißen" (literal: to bite the grass; actual: to bite the dust, die) b. "auf der Hand liegen" (literal: to lie on the hand; actual: to be obvious) <p>In contrast, less fixed partly idiomatic expressions (teilidiomatisierte Wendungen) are expressions in which some component part, typically the base in Hausmann's conception, still keeps its literal meaning, such as the nouns "Druck" (pressure) and "Geltung" (importance) in the following examples:</p>	<p>крайней мере одна составная часть либо сохранила, либо утратила свое буквальное значение.</p> <p>Другое центральное различие в концепции словосочетаний Хаусмана связано со степенью постоянства между различными составляющими словосочетания.</p> <p>С одной стороны, существуют фиксированные словосочетания, под которыми можно обнаружить в основном идиомы и к которым вряд ли применимо вышеприведенное определение основы и словосочетания.</p> <p>Эти фиксированные выражения называются полностью идиоматическими выражениями, в которых каждый компонент лишен своего буквального значения, что иллюстрируется следующими идиоматическими выражениями:</p> <p>6.</p> <ul style="list-style-type: none"> а. «ins Gras beißen» (буквально: кусать траву; на самом деле: быть убитым, умереть) б. «auf der Hand liegen» (буквально: лежать на руке; на самом деле: быть очевидным) <p>Напротив, менее фиксированные частично идиоматические выражения (teilidiomatisierte Wendungen) представляют собой выражения, в которых некоторая составная часть, как правило, основа в концепции Хаусмана, все еще сохраняет свое буквальное значение, например, существительные «Druck»</p>
---	--

<p>7.</p> <p>a. “unter Druck geraten” (to get under pressure)</p> <p>b. “zur Geltung kommen” (to become important)</p> <p>These are also the types of expressions that Hausmann (1985) refers to as collocations. On the other end of the continuum there are free word combinations where all components keep their literal meaning, thus making the expression fully compositional:</p> <p>8.</p> <p>a. “auf einen Baum klettern” (to climb up a tree)</p> <p>b. “an die Zukunft glauben” (to believe in the future)</p> <p>In the linguistic classification task to derive a gold standard for German preposition-noun-verb (PNV) collocations (as described in subsection 4.5.2.3), a distinction along these lines has turned out to be quite operational for the human classification of collocation candidates.</p> <p>We will return to this issue in subsection 2.1.4 below in which we assemble our adopted linguistic properties of collocations.</p> <p>2.1.1.2 Other Lexicographic Accounts of Collocations</p>	<p>(давление) и «Geltung» (важность) в следующих примерах:</p> <p>7.</p> <p>a. «unter Druck geraten» (попасть под давление)</p> <p>б. «zur Geltung kommen» (стать важным)</p> <p>Это также типы выражений, которые Хаусманн (1985) называет словосочетаниями. На другом конце континуума находятся свободные словосочетания, в которых все компоненты сохраняют свое буквальное значение, что делает выражение полностью композиционным:</p> <p>8.</p> <p>a. «auf einen Baum klettern» (взобраться на дерево)</p> <p>б. an die Zukunft glauben (верить в будущее)</p> <p>В задаче лингвистической классификации по получению золотого стандарта для немецких словосочетаний предлог-существительное-глагол (ПСТГ) (как описано в подразделе 4.5.2.3) различие в этом направлении оказалось вполне применимым для человеческой классификации кандидатов на словосочетания.</p> <p>Мы вернемся к этому вопросу в подразделе 2.1.4 ниже, в котором мы собираем наши принятые лингвистические свойства словосочетаний.</p> <p>2.1.1.2 Другие лексикографические описания словосочетаний</p>
--	---

<p>This subsection reviews two accounts of collocations which may be mainly described as applicational as they are primarily concerned with collecting and representing collocational entries in a lexicon or dictionary.</p> <p>Whereas the first account (Benson et al., 1986b) still offers some theoretical underpinnings, the second one (Dudenredaktion, 2002) is purely applicative in nature but needs to be discussed as it constitutes the only such type of work for the German language.</p> <p>The first dedicated and large-scale lexicographic study of collocations was undertaken for the English language by Benson et al. (1986b), Benson (1989) and Benson (1990), which led to the publication of the BBI Combinatory Dictionary of English: A Guide to Word Combinations (in short: BBI) (Benson et al., 1986a).</p> <p>Benson et al. (1986a) outline the motivation for a dictionary of word combinations and the kinds of information included in it. The goal is to provide information on the general combinatorial possibilities of an entry word.</p> <p>Various types of combinatorial preferences are listed, such as e.g. whether there are any combinatorial preferences of verbs for nouns (e.g. “[to adopt, enact, apply] a regulation”) or</p>	<p>В этом подразделе рассматриваются два типа словосочетаний, которые в основном можно охарактеризовать как прикладные, поскольку они в первую очередь связаны со сбором и представлением словосочетаний в лексиконе или словаре.</p> <p>В то время как первый отчет (Бенсон и др., 1986b) по-прежнему предлагает некоторые теоретические обоснования, второй (Dudenredaktion, 2002) носит чисто прикладной характер, но требует обсуждения, поскольку представляет собой единственный такой тип работы для немецкого языка.</p> <p>Первое специализированное и крупномасштабное лексикографическое исследование словосочетаний было предпринято для английского языка Бенсоном (Бенсон и др. (1986b), Бенсон (1989) и Бенсон (1990)), что привело к публикации Комбинированного словаря английского языка BBI: Руководство по словосочетаниям (сокращенно: BBI) (Бенсон и др., 1986a).</p> <p>Бенсон и др. (1986a) обрисовывают мотивацию создания словаря словосочетаний и виды информации, включенной в него. Цель состоит в том, чтобы предоставить информацию об общих комбинаторных возможностях входящего слова.</p> <p>Перечислены различные типы комбинаторных предпочтений, например, существуют ли какие-либо комбинаторные предпочтения глаголов для существительных</p>
---	--

what the possible adverbial combinations (i.e. modifications) of a verb are (e.g. “to regret [deeply, very much]”).

These combinatorial preferences are classified into two types of collocations, i.e., grammatical collocations and lexical collocations.

Grammatical collocations are phrases consisting of a dominant word (e.g. noun, adjective, verb) and a preposition or grammatical structure such as an infinitive or a clause, as exemplified by the following expressions:

9.
 - a. “account for”
 - b. “adjacent to”
 - c. “dependent on”
 - d. “the fact that + clause”

Lexical collocations, on the other hand, are classified by the BBI approach according to their part-of-speech patterns, such as verb-(preposition)-noun, adjective-noun or noun-noun, as exemplified by the following expressions:

10.
 - a. “compose music”
“launch a missile”
“set an alarm” (verb-noun pattern)
 - b. “strong tea”

(например, «[принимать, вводить в действие, применять] правило») или каковы возможные комбинации наречий (т. е. модификации) глагола (например, «сожалеть [глубоко, очень]»).

Эти комбинаторные предпочтения подразделяются на два типа словосочетаний: грамматические словосочетания и лексические словосочетания.

Грамматические словосочетания – это фразы, состоящие из доминирующего слова (например, существительного, прилагательного, глагола) и предлога или грамматической структуры, такой как инфинитив или предложение, как показано в следующих выражениях:

9.
 - а. «учитывать»
 - б. «рядом с»
 - в. «зависит от»
 - д. «тот факт, что + пункт»

С другой стороны, лексические словосочетания классифицируются с помощью подхода ВБИ в соответствии с частями речи, такими как глагол-(предлог)-существительное, прилагательное-существительное или существительное-существительное, как показано в следующих выражениях:

10.
 - а. «сочинять музыку»
«запустить ракету»
«установить будильник»
(форма глагол-существительное)

<p>“chronic alcoholic” (adjective-noun pattern) c. “a swarm of bees” “a flock of sheep” (noun-noun pattern)</p> <p>Although some of these expressions describe lexically determined co-occurrences and thus are more in line with what is commonly understood as collocation, it can be seen that others again are fairly compositional from a semantic perspective in that all constituents still keep their literal meaning and thus probably would not be labelled “collocation” by approaches such as Hausmann’s (outlined in subsection 2.1.1.1 above).</p> <p>A look at the intended audience, however, explains the extensiveness of the BBI approach to word combinations and collocations since Benson et al. (1997) explicitly target their dictionary towards foreign-language learners. Still, the BBI dictionary is the most comprehensive lexicographic resource of word combinations in any language to date and thus deserves attention.</p> <p>As far as dictionaries and lexicographic resources for German-language collocations and idiomatic expressions are concerned, Volume 11 of the Duden series (Dudenredaktion, 2002) may be regarded as the main representative.</p>	<p>б. «крепкий чай» «хронический алкоголик» (образец прилагательное-существительное) в. «рой пчел» «стадо овец» (образец существительное-существительное)</p> <p>Хотя некоторые из этих выражений описывают лексически определенные совпадения и, таким образом, больше соответствуют тому, что обычно понимается как словосочетание, можно увидеть, что другие опять-таки довольно композиционны с семантической точки зрения, поскольку все составляющие по-прежнему сохраняют свое буквальное значение и, таким образом, вероятно, не будет называться коллокацией в рамках таких подходов, как подход Хаусмана (описанный в подразделе 2.1.1.1 выше).</p> <p>Однако взгляд на целевую аудиторию объясняет широту подхода ВБИ к словосочетаниям и словосочетаниям, поскольку Бенсон и др. (1997) явно ориентируют свой словарь на изучающих иностранный язык. Тем не менее, словарь ВБИ на сегодняшний день является наиболее полным лексикографическим ресурсом словосочетаний на любом языке и поэтому заслуживает внимания.</p> <p>Что касается словарей и лексикографических ресурсов для немецкоязычных словосочетаний и идиоматических выражений, то 11-й том серии Duden (Dudenredaktion, 2002) можно считать основным</p>
---	--

This dictionary, however, differs from the BBI dictionary in several respects.

As already the title “Redewendungen” (figures of speech, sayings) suggests, the focus of this volume is rather on idiomatic speech figures than on the allowable and preferred combinatorial properties of words. Hence, each entry in the dictionary is accompanied by etymological information rather than by lexico-grammatical one.

Still, in practice this dictionary has a broader definition of what is considered to fall under the notion of “Redewendungen”. In their introductory remarks, for example, the Duden editorial staff points out that, besides idiomatic expressions, they also count Funktionsverbgefüge (support verb constructions) to the class of collocations.

As will be seen later on, these types of syntactic constructions will play a prominent role in the set of collocational candidates in our experimental study described in subsection 4.5.2. In particular, we will focus on their surface realization as preposition-noun-verb (PNV) constructions.

представителем. Этот словарь, однако, отличается от словаря BBI в нескольких отношениях.

Как уже предполагает название «Redewendungen» (фигуры речи, высказывания), в этом томе основное внимание уделяется идиоматическим фигурам речи, а не допустимым и предпочтительным комбинаторным свойствам слов. Следовательно, каждая статья в словаре сопровождается этимологической, а не лексико-грамматической информацией.

Тем не менее, на практике этот словарь имеет более широкое определение того, что считается подпадающим под понятие «Redewendungen». В своих вступительных замечаниях, например, редакция Duden указывает, что, помимо идиоматических выражений, они также относят к классу словосочетаний Funktionsverbgefüge (опорные глагольные конструкции).

Как будет видно позже, эти типы синтаксических конструкций будут играть заметную роль в наборе потенциальных коллокаций в нашем экспериментальном исследовании, описанном в подразделе 4.5.2. В частности, мы сосредоточимся на их поверхностной реализации в виде конструкций предлог-существительное-глагол (ПСГ).