

## ЛЕКСИКОГРАФИЧЕСКИЙ АСПЕКТ ИЗУЧЕНИЯ КОЛЛОКАЦИЙ

### LEXICOGRAPHICAL ASPECT OF THE COLLOCATION STUDY

*S. Alikova  
O. Shibkova*

#### Annotation

In this article we present a description of the paradoxes and the possibilities of study of the usual collocations on the basis of corpora and traditional data dictionaries, turn to the theoretical foundations of collocation theory. The study is based on the semiotic approach, namely semantic, lexicographic and syntactic aspects. The relevance of the research object is motivated by the realities of modern linguistics, since there have been significant changes in the features of empirical research, which allow to conduct field research without involvement of examinees via the Internet, with the corpora use.

**Keywords:** Collocation, corpus linguistics, usual phrase, lexicography, phraseology.

*Аликова Светлана Викторовна*

*К.филол.н., доцент,*

*Северо-Кавказский федеральный*

*университет, Ставрополь*

*Шибкова Оксана Сергеевна*

*Д.филол.н., профессор,*

*Северо-Кавказский федеральный*

*университет, Ставрополь*

#### Аннотация

В данной статье мы представляем описание апорий и возможностей исследования узуальных словосочетаний на базе корпусных и традиционных словарных данных, обращаемся к теоретическим основам коллокативности. Исследование базируется на семиотическом подходе, а именно на его лексикографическом, семантическом и синтаксических аспектах. Актуальность объекта исследования обусловлена реалиями современной лингвистики, поскольку произошли значительные изменения возможностей эмпирических изысканий, которые позволяют проводить полевые исследования без привлечения испытуемых, посредством интернета, с применением корпусов.

#### Ключевые слова:

лингвистический корпус, узуальное словосочетание, лексикография, фразеологизм.

Среди современных корпусных систем одной из наиболее полных и известных является немецкая система IDS "Корпус немецкого национального языка". В ее рамках создается проект "Знание о словах", лексико-лексикологическая информационная система, которая базируется на около 650 миллионов слов, всеобъемлюще документирует немецкий словарный состав и дает лингвистическое толкование. В своей конечной версии это электронное справочное издание будет содержать около 300 000 ключевых слов и, благодаря гипертекстуальной структуре, предлагать гибкую, ориентированную на запрос информацию как специалистам, так и заинтересованным непрофессионалам посредством интернета, CD-рома и – в случае необходимости – в печатной форме. По построению и возможностям использования система фундаментально отличается от привычных печатных изданий и словарей, которые были перенесены в электронную форму, но сохранили линейную структуру печатных изданий.

Рассматривая пестрый консервативный состав словарей немецкоязычных стран, представляется возможным во всех из них установить наличие комментариев

полилексемных образований, однако они носят бессистемный характер. Как правило, нет различий между информацией о коллокативном и фразеологическом типе связи в выражениях. Х. Бургер связывает данную бессистемность с отсутствием теоретической концепции для современной фразеологии. Несмотря на множество подходов в области исследования идиоматики, не существует адаптированной к лексикографическим целям модели наиболее употребляемых выражений в языке, основанной на последовательном анализе реального языкового употребления и включающей регулярные контекстуальные условия. В идиоматической лексикографии различают следующие группы представленной информации: информация о словосочетаниях в одноязычных словарях, информация о словосочетаниях в двуязычных словарях, идиоматические одноязычные и сравнительные специальные словари.

Самое большое собрание фразеологических сочетаний однозначно содержат специальные словари. То, что данные высокоидиоматичные, непрозрачные выражения неравномерно освещаются как в научном анализе, так и в лексикографических изданиях, объясняется язы-

ковой природой фразеологизмов: они отображают относительно однозначные примеры идиоматичности и имеют относительно обозримый состав языковых форм, которые хорошо поддаются классификации – без корпусов и компьютерных методов – синтаксически и семантически. По причине своей малой прозрачности они предоставляют достаточно материала для лексикологического анализа и предрасположены к тому, чтобы войти в состав словарей. В печатных словарях можно встретить коллокации без опоры на определенную теоретическую модель. Х. Бургер критически констатирует, что все лексикографические данные о коллокациях связаны с большим количеством неразрешенных проблем, что можно отнести ко всем многословным образованиям [4, S. 170].

Первая проблема, с которой сталкиваются лексикографы, это идентификация статуса, т.е. определение того, к какому типу словосочетаний относится выражение и при каких условиях оно становится главным словом. До настоящего времени это определялось интуитивно, на основе индивидуальной выборки и не основывалось на определенном корпусе данных. И хотя классические картотечные ящики заменяются данными, собранными на основе компьютерной обработки, в большинстве случаев лексикограф на основе собственной языковой компетенции определяет, является ли данный вариант употребительным и актуальным и – в нашем случае – является ли одно выражение более типичным чем другое. В новых изданиях используются корпусы, но решение об использовании этих данных остается за лексикографом, часто для того, чтобы подтвердить собственные субъективные решения о внесении данных, основанные на многолетнем опыте. Это способствует тому, что в словари и издания попадают показательные, классические примеры и доводы. Существует так называемая коллективная память иллюстративного материала исследователей и лексикографов, которая имеет давнюю историю и нуждается в эмпирической проверке актуальности языкового употребления. Это может исправить только строго основанное на корпусах эмпирическое основание лексикографии. Изначально опора на корпусные данные таит в себе "психологический" барьер: при исследовании лексикограф сталкивается с результатами, которые противоречат его языковой интуиции, но которые являются показательными в плане языкового употребления.

В гипертекстуальном справочном издании, документирующем немецкий словарный состав с учетом всех уровней языковой системы и всех аспектов языкового употребления, в центр внимания неизбежно попадают вопросы идиоматики. По данной причине адаптированная для лексикографических целей модель должна быть развита из традиционных выражений, интегрирована к различным подходам к исследованию идиом и представить их в новом свете. Если область исследования рас-

сматривать в данном ракурсе, становятся очевидными некоторые противоречия и дезидераты: существует много несовместимых подходов и терминов, которые существуют параллельно или дают различные определения. Причины этого очевидны: почти каждое лингвистическое направление, от генеративной трансформационной лингвистики через призму контрастивной лингвистики до стилистики, сталкивается с проблемой, что помимо грамматических связей между словами существуют иные структурные отношения, которые необходимо описать и подвергнуть анализу.

Каждое направление предложило собственную исследовательскую модель, в результате чего словосочетания носят названия: фразеологизмы, коллокации, валентность или синтаксическая коокуренция, функциональные словосочетания, коммуникативные формулы и др. В контексте комплексной информационной системы необходимо, однако, объединить различные точки зрения. Особенно это относится к объединению классификационно-аналитических подходов фразеологии со статистико-синтагматическими методами коллокационного анализа, разработанными британской школой контекстуализма.

Фразеология относится к одной из наиболее тщательно разработанных лингвистических дисциплин, предполагающей необозримым множеством работ, одноязычными и контрастивными анализами и классификациями. Однако в фокусе исследования в основном находились идиомы. Хотя в специальной литературе они имеют разные определения, фразеологизмам присущи несколько общих свойств, прежде всего, это устойчивость. Они перенимают функцию отдельно взятой лексики, общее значение прямо не выводимо из суммы значений компонентов. Чем меньше прозрачность идиомы, т.е. чем больше разница между словесным и переносным значением, тем выше степень идиоматичности. В более поздних фразеологических работах выявляются проблемы рассмотрения фразеологизма как традиционного, застывшего понятия. Х. Бургер констатирует, что фразеологические исследования делают очевидным то, что фразеологизмы – полилексические единицы относительно устойчивости, которые имеют равноуровневые связи между собой и лексическим составом. "Statt dessen brauchen wir eine flexible Konzeption von Phraseologie, die den Gedanken toleriert, da? man innerhalb eines Phraseologismus stabilere und weniger stabile Elemente annehmen darf, ohne da? damit sein Status als linguistische bzw. psycholinguistische Einheit verloren ginge" [4, S. 31].

Исследование коллокаций, напротив, сравнительно молодое лингвистическое направление, рассматриваемое с позиций различных школ: школа британского контекстуализма рассматривала коллокации с точки зрения

статистико-синтагматического подхода (Sinclair 1991), также достойны упоминания филологически ориентированные, основанные на контрастивной лингвистике и лексикографии коллокационные модели (Hausmann 1985) и исследования семантической сочетаемости и понятийных реляций (Agricola 1982, Viehweger 1977).

Школа британского контекстуализма считается основателем коллокационного исследования. Она помещает в центр исследования статистическую значимость. Важный критерий – статистическая ожидаемость совместной встречаемости словосочетаний в корпусе. Д. Синклер понимает коллокацию как "the occurrence of two or more words within a short space of each other in a text" [7, P. 170]. Он руководствуется статистическим методом, в качестве базы выступает большой корпус данных, при этом "коллокация" является родовым понятием для всех повторяющихся коокуренций (идиом, клише, фразеологизмов, сложных слов и др.). Это широкое определение становится решающей проблемой идентификации выражений, т.е. вопрос, какие словосочетания считаются коллокациями и какие различия они проявляют, остается без ответа. Бесспорная заслуга британского контекстуализма состоит в исследовании коллокаций в "реальных" текстах и разработке методов исследования большого объема текста.

Помимо англосаксонской традиции выделяется учение Ф. Хаусманна, который называет коллокации "языковыми полуфабрикатами" [4, S. 118]. Постепенно коллокации прекращают рассматриваться учеными как нечто недостойное столь пристального внимания, как фразеологизмы. В своем исследовании Ф. Хаусманн различает фиксированные словосочетания, такие, как сложные слова и устойчивые выражения и нефиксированные выражения, которые включают коллокации нескольких видов. Знание коллокаций – это не лингвистическая роскошь, а элементарная необходимость, в них заложено языковое своеобразие [4, S. 9]. Это легко подтвердить на примере билингвальной коммуникации: фразеологизмы были – в противоположность коллокациям – проблемой многоязычной коммуникации, так как говорящий на втором языке вынужден в большинстве случаев использовать словарь, чтобы понять их значение. Это привело к преувеличению значимости фразеологизмов в мультязычной коммуникации. Напротив, коллокациям была отведена весьма скромная роль. Однако, несмотря на существенную роль фразеологизмов, изучающий иностранный язык вполне может изъясняться и без овладения фразеологической базой. Он может ее просто "избегать". Как правило, от иностранца и не ожидают развитой фразеологической компетенции, так как не исключена возможность нарушения стилистической нормы или коммуникативного непонимания. С коллокациями все обстоит иначе: без них продуцент текста на иностранном языке вряд ли сможет обойтись. При восприятии

иностранного текста говорящему не составит труда понять выражение типа *to do an attempt* (*einen Versuch machen*). При переводе немецкого выражения на английский он, вероятно, может сказать "to make an attempt". Комбинация *make-attempt* хотя и понятна, но не типична для англоязычной среды. Коллокации позволяют отчетливо выявить степень владения иностранным языком, хотя нарушение "коллокационных норм" не наносит такого вреда акту коммуникации, как дословно переведенное или не ситуативно примененное устойчивое выражение. В качестве аргумента в пользу коллокаций вышла статья Ф. Хаусманна "Wortschatzlernen ist Kollokationslernen" (1984), в которой он представил коллокационную модель на основе перевода с французского на немецкий. Д. Фивегер, с опорой на концепцию "слов и выражений" Э. Агриколы (1992), модель ментальной организации знаний Ф. Кликса (1984) и собственные исследования в области семантического анализа, предлагает дефиницию коллокаций с учетом структурных отношений между единицами в лексиконе. Под коллокациями следует понимать двучленные семантические отношения между так называемыми автосемантическими словами (*Autosemantika*), которые, в отличие от фразеологизмов, имеют правильную, соответствующую нормам, организацию слов-элементов [10, S. 889]. С точки зрения теории языка эта дефиниция является самой последовательной. В плане лексикографической репрезентации он склоняется к тому, чтобы не рассматривать более коллокации – в отличие от предпочитаемых ранее парадигматических данных – как случайное пояснение в словарных толкованиях, поскольку они представляют собой приоритетную информацию для пользователя [10, S. 892]. В настоящее время коллокации постепенно попадают в поле зрения лингвистических исследований. Существенным импульсом явилась концепция Х. Файльке о "идиоматическом характере языковой компетенции" [3].

Вторая довольно серьезная проблема: необходимо определить, под каким ключевым словом устойчивые выражения должны быть представлены в печатных словарях. Часто они состоят более чем из одной смысловоединицы. По мнению Х. Бургера, это решается большей частью "прагматически", когда устойчивые словосочетания размещаются по первым смысловым единицам или входящим в состав существительным, другие составляющие единицы помещаются с соответствующей сноской. Следовательно, коллокации будут располагаться под основой, которая часто выражена существительным. Насколько это сомнительно, иллюстрирует следующий пример: комбинация слов *schutteres Haar* будет находиться под основой *Haar*, хотя справочный материал к коллокатору "schutter" требуется в большей степени. В данном простом случае *schutteres Haar* можно обойтись двойным комментарием: как под словом *Haar*, так и под словом *schutter* с перекрестными

ссылками. Однако данная практика не применима к более сложным сочетаниям по причине ограниченного объема словарной статьи. Ф. Хаусманн, описывая практику составления немецкого универсального словаря, подтверждает во многих случаях внесение коллокаций как начиная с основы, так и с коллокатора и признает, что это не всегда возможно по причине экономии места [4, S. 121].

Последствия вынужденного сжатия в печатной лексикографии проявляются в области прагматилистических комментариев, которые важны именно для идиом по причине высокой контекстуальной обусловленности. Стилиевые ссылки и пометки (gehoben, umgangssprachlich) ограничены несколькими словами. Частично это однословные комментарии, которые не соответствуют современному состоянию уровня исследований в области прагматики, лингвистики текста и стилистики. Данные однословные комментарии внушают, что якобы существуют устойчивые, ингерентные прагматические признаки и коннотации, свойственные лексемам. При этом именно эти данные в большей степени контекстуально обусловлены и могут быть отнесены к лексеме вариативно, что возможно представить только в более подробных комментариях из одного и более предложений.

Однословные маркеры по праву подвергаются критике в металексикографии, однако это не отразилось на практической словарной работе.

Следующая, до настоящего времени не разрешенная проблема – вопрос так называемой нормальной формы фразеологизма: в какой форме должна быть представлена коллокация в словарной статье. Сокращенные до обязательной (не вариативной) части основные формы представляют собой лексикографические конструкты и значительно отличаются в предыдущих словарях и собраниях идиом. В данном случае корпусные данные могут преподнести информацию, как представлена основная форма и какие модификации возможны.

Многие упущения классической печатной лексикографии могут быть устранены посредством возможностей компьютерной лексикографии. Имеются многочисленные попытки создания алгоритмов по нахождению коллокаций и фразеологизмов в корпусах. Проекты, имеющие целью выявление коллокаций различных языков посредством статистической и машинной обработки, ориентированы на развитие соответствующих автоматических инструментов и оказание помощи специалистам и заинтересованным лицам.

#### ЛИТЕРАТУРА

1. Agricola E. Ein Modellwörterbuch lexikalisch-semanticcher Strukturen. In: Agricola, E./Schildt, J./Viehweger, D. (Hg.): Wortschatzforschung heute. Aktuelle Probleme der Lexikologie und Lexikographie. Leipzig, 1982. S. 9–22.
2. Burger H. Phraseologie. Eine Einführung am Beispiel des Deutschen. Berlin, 1998.
3. Feilke H. Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik. Frankfurt a. M., 1996.
4. Hausmann F. Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. In: Bergenholtz, H./Mugdan, J. (Hg.): Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch. Tübingen, 1985. S. 118–129.
5. Klix F. Über Wissensrepräsentation im menschlichen Gedächtnis. In: Klix, F. (Hg.): Gedächtnis – Wissen – Wissensnutzung. Berlin, 1984. S. 9–73.
6. Schemann H. PONS Deutsche Idiomatik. Die deutschen Redewendungen im Kontext. Stuttgart/Dresden, 1993.
7. Sinclair J. Corpus, Concordance, Collocation. Oxford, 1991.
8. Steyer K. Usuelle Wortverbindungen des Deutschen // Deutsche Sprache Jg. 28. 2000. S. 101–125.
9. Viehweger D. Probleme der semantischen Analyse. (= studia grammatica; XV). Berlin, 1977.
10. Viehweger D. Probleme der Beschreibung semantischer Vereinbarkeitsrelationen im allgemeinen einsprachigen Wörterbuch. In: Hausmann F.J., Reichmann O., Wiegand H.E., Zgusta L. (Hg.): Wörterbücher. Ein internationales Handbuch zur Lexikographie. Berlin/New York, 1989. S. 888–893.
11. Zimmer R. Äquivalenzen zwischen Französisch und Deutsch. Theorie – Korpus Indizes. Ein Kontextwörterbuch. Tübingen, 1990.

© С.В. Аликова, О.С. Шибкова, (kopaeva1979@yandex.ru), Журнал «Современная наука: актуальные проблемы теории и практики».

