

УДК 81.13

DOI: 10.33184/bulletin-bsu-2020.4.43

МОДЕЛИРОВАНИЕ ТЕРМИНОЛОГИЧЕСКОГО ПОТЕНЦИАЛА ПОНЯТИЯ «КОМПРЕССИЯ»

© Д. А. Ичкинеева

Башкирский государственный университет
Россия, Республика Башкортостан, 450076 г. Уфа, ул. Заки Валиди, 32.

Тел.: +7 (347) 273 66 72.

Email: dilaraichkineeva@gmail.com

В статье предлагается обзор научных публикаций, в которых встречается понятие «компрессия». Актуальность исследования данного термина обусловлена его интегративной и междисциплинарной природой. Для анализа использования данного понятия в работах российских ученых предлагается следующая исследовательская программа: с помощью расширенного поиска в научной электронной библиотеке eLibrary идентифицируются все статьи по языкознанию в названиях, ключевых словах и аннотациях которых встречается термин «компрессия». В информационной системе «Семограф» создается база данных контекстов, в которых встречается термин «компрессия». В качестве контекстов рассматриваются ключевые слова, которые впоследствии объединяются в семантические поля. После распределения всех компонентов по семантическим полям в информационной системе «Семограф» [setograph.com] строится семантическая карта на основе встречаемости компонентов полей в одном контексте. На основе семантической карты конструируется графосемантическая модель, т.е. в результате исследования получена интегральная модель понятийного аппарата термина «компрессия» на основе российских публикаций. Таким образом, предлагаемая методология визуализации понятийной сферы исследуемого термина при помощи модели позволяет решить задачи анализа, сравнения и оценки предметных областей.

Ключевые слова: компрессия, текст, моделирование, ключевые слова, ИС «Семограф».

В настоящее время один и тот же термин-понятие, который используется в той или иной предметной области, в связи с развитием академической мобильности, открытости научных исследований, а также стремлением к междисциплинарности получает в лингвистических исследованиях как качественное, так и количественное развитие [1–3], что определяет актуальность проведения исследований, направленных на решение задачи анализа, сравнения и оценки предметных областей научного производства. В нашем исследовании проводится анализ использования термина «компрессия» в работах российских ученых. Актуальность исследования данного термина обусловлена его интегративной и междисциплинарной природой. Термин «компрессия», пришедший в лингвистику из теории связи, в общем смысле обозначает «сжатие», сокращение объема написанного без ущерба для его содержания [4]. В современной зарубежной лингвистике компрессия рассматривается в основном в аспекте существующих автоматизированных средств компрессии огромного массива текстов, при этом исследования посвящены разработке машинных методов извлечения информации и созданию аннотированных текстов на основе вычислительных методов [5–7], в то время как в центре исследований в отечественной лингвистике находятся способы компрессии на разных языковых уровнях [8–15], в которых компрессия рассматривается как процедура перекодирования данных, производимая с целью уменьшения их объема. И в информатике, и в лингвистике компрессия основана на

устранении избыточности информации, содержащейся в исходных данных. Кроме того, следует учитывать естественную тенденцию к компрессии информации в процессе ее переработки и хранения, что представляет интерес для когнитивной лингвистики.

В нашем исследовании использована концептуально-гипертекстовая программа моделирования сферы применения понятия «компрессия» в современных лингвистических исследованиях на основе информации о нем, репрезентированной в корпусе текстов [11; 17–19].

Для определения понятийного потенциала термина «компрессия» в современной научной парадигме было проведено следующее исследование: отобрано 473 статьи, в заголовке или ключевых словах которых встречалось слово «компрессия» под рубрикой «Языкознание» (отбор производился в научной электронной библиотеке eLibrary с 2006 по 2018 гг.). На основе предположения о том, что набор ключевых слов является скомпрессированным содержанием научных статей, моделирование проводилось именно на материале ключевых слов в обозначенных статьях (всего 473 статьи). Ключевые слова статей, посвященные компрессии в сфере языкознания, были занесены в информационную систему «Семограф» (1892 ключевых слов), далее был проведен полевой анализ: ключевые слова распределены по следующим полям: *Концептосфера, Семиотическое пространство, Логические категории и операции, Текст и дискурс, Грамматика, Лексика и фразеология, Компрессия, Методы исследования, Понятийно-категориальная сфера,*

Структурная организация, Обучение, Предметная область, Значение и информация, Экстралингвистические понятия, Сфера культуры и лингвокультуры, Язык, Коммуникация. Распределение производилось в соответствии с названиями семантических полей. Для нас особый интерес представляет поле *Предметная область*, т.к. оно обобщает основные области языкознания, в которых используется понятие «компрессия», а именно: *социолингвистика, синергетика, семиотика, прагматика, прикладная и математическая лингвистика, язык и общество, психолингвистика, перевод, грамматика, общее языкознание, теория языка, сравнительно-историческое, типологическое и сопоставительное языкознание, орфоэпия, лингвосинергетика, прагматология, медиалингвистика, рекламная лингвистика, теория стиля, теория речевых жанров, теория фреймов* и т.д.

Одно и то же понятие может входить в несколько полей. Так, например, ключевое слово «Психолингвистические аспекты коммуникации» вошло в поле *Предметная область* и *Коммуникация*.

Далее в информационной системе «Семограф» была сгенерирована семантическая карта на основе совместной встречаемости слов в одном контексте (в нашем случае – наборе ключевых слов). Семантическая карта представляет собой таблицу, на пересечении строк и столбцов которой находится показатель силы семантической связи между полями. Так как некоторые связи могут носить случайный характер, был определен порог значимости (среднее + стандартное отклонение). Связи, превысившие порог значимости определены как сильные, средние и слабые. На основе полученных данных была построена графосемантическая модель понятийного потенциала термина «компрессия» в языкознании.

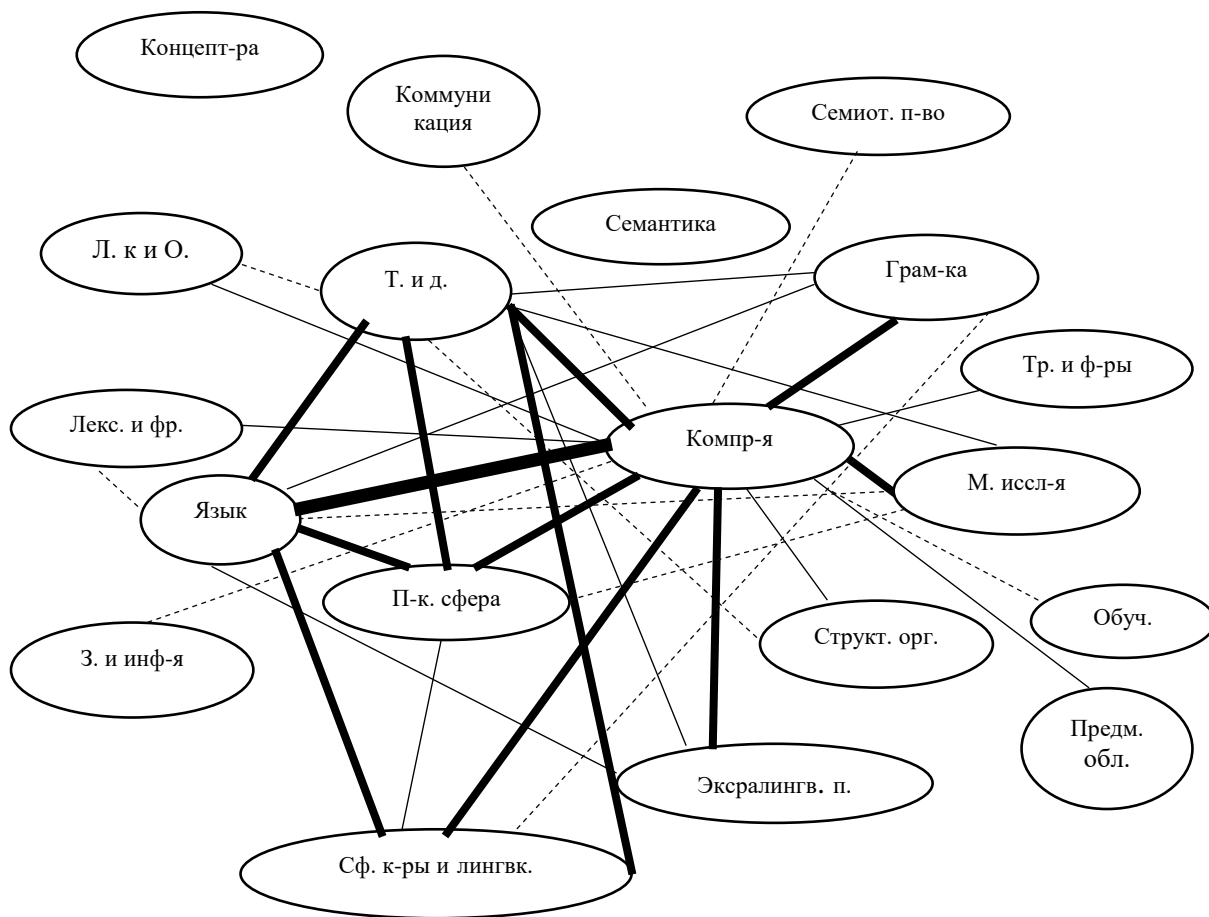


Рис. Графосемантическая модель понятийного потенциала термина «компрессия» в современных лингвистических исследованиях.

Примечание: на рисунке приняты следующие сокращения: *Концепт-ра* – Концептосфера; *Семиот. пр-во* – Семиотическое пространство; *Л.к. и О.* – Логические категории и операции; *Т. и Д.* – Текст и дискурс; *Грам-ка* – Грамматика; *Лекс. и фр.* – Лексика и фразеология; *Компр-я.* – Компрессия; *М. иссл-я* – Методы исследования; *П.-к. сфера* – Понятийно-категориальная сфера; *Структ. орг.* – Структурная организация; *Обуч.* – Обучение; *Предм. обл.* – Предметная область; *З. и. инф-я* – Значение и информация; *Тр. и ф-ры* – тропы и фигуры; *Экстралингв. п.* – Экстралингвистические понятия; *Сф. к-ры и лингвк.* – Сфера культуры и лингвокультуры.

Ядерным компонентом в данной модели является поле *Компрессия*, представленное ключевыми словами и выражениями, в семантическое поле которых входит данное понятие: «Компрессия смысла», «приемы сжатия», «языковая компрессия» и др. Самой сильной связью полученной модели является связь данного компонента модели с полем *Язык* (куда вошли понятия, связанные именно с семьей «язык»: «уровни языка», «компрессия языка», «язык газеты» и т.д., что свидетельствует о том, что понятие «компрессия» используется главным образом в работах, посвященных исследованию языка, и одним из самых распространенных контекстов употребления данного термина является «компрессия языка».

Кроме того, этот компонент имеет сильные связи с полями *Текст и дискурс*, *Понятийно-категориальная сфера*, *Сфера культуры и лингвокультуры*, *Грамматика и Методы исследования*, что также свидетельствует о применении понятия «компрессия» в данных областях языкознания. Таким образом, предложенная исследовательская программа позволяет выявлять особенности реализации термина «компрессия» в научных статьях отечественных ученых. Так, например, связь компонента *Сфера культуры и лингвокультуры* с компонентом *Компрессия* свидетельствует об использовании данного понятия в контексте лингвокультуры, ассоциативного мифотворчества, лингвоэкологии, современной немецкой поэзии, мифотворчества (ключевые слова, представляющие поле *Культура и лингвокультура*) и т.д. Сильная связь с полем *Грамматика* – о грамматической и синтаксической компрессии, эллипсисе, компрессии сложных и усложненных предложений, об использовании термина компрессии при рассмотрении таких понятий как синтаксическая структура, динамическая грамматика, генерация осмысленных предложений, специализированные синтаксические средства, аграмматизм, эвфемизмы со структурой предложения, трансформация структуры, трансформированные структуры, неполное предложение, инкорпорирование, вводные и вставные конструкции, сверхфразовое единство, грамматический уровень, сложность синтаксических структур, неличные формы глагола, атрибутивные группы, пассивные конструкции, микросинтаксис, синтаксический анализ, лингвистические (морфологические и синтаксические) механизмы, синтаксические особенности, современный синтаксис, полупредикативная конструкция, усложненное предложение, ядерные структуры и мн.др.

Взаимосвязь компонентов модели *Компрессия – Методы исследования* свидетельствует об исследовании компрессии при помощи следующих методов: сравнительный анализ, моделирование, математический метод, аналитико-синтетическая обработка и преобразование информации, ситуационно-стратегическое моделирование, анализ звучаще-

го языкового материала, статистическая обработка экспериментальных данных, педагогический эксперимент, методика ключевых слов, сопоставительный анализ, классификации стратегий и тактик компрессии текста, квантитативный анализ, метод фразеологической идентификации, метод цепного упрощения, интерпретация содержания текста, метод структурно-логических схем, системно-обобщающая технология, семантический анализ, лингвостилистический анализ, жанрово-стилистический анализ, толкование аббревиатуры, методика смысловой компрессии.

Связь с компонентом модели *Текст и дискурс* говорит об использовании понятия «компрессия» в исследованиях, посвященных следующим лингвистическим явлениям: дискурс, поэтический дискурс, понимание текста, текстообразование, медиатексты, публицистический текст, дискурсивная практика, компрессия научного текста, учебный научно-технический текст, реферат, иноязычный текст, научно-педагогический текст, современный электронный дискурс, литературный текст, репродуцированный текст, дискурсивное опущение, публичное выступление, сжатие текста, вторичные тексты, смысловое пространство текста, лингвистические особенности текста, литературно-художественные характеристики текста, компрессия как способ создания имплицитного содержания текста, аутентичный текст, ритм текста, содержание текста, смысловая структура текста, информационная обработка текста, механизмы смысловой переработки текста, прецедентные тексты, поэтический текст и др. На рисунке видно, что элемент модели *Компрессия* также связан с такими полями, как *Семантика*, *Структурная организация*, *Предметная область*, *Коммуникация*, *Значение и информация*, *Обучение*, *Лексика и фразеология*, однако эти связи представлены гораздо меньшим количеством контекстов совместной встречаемости, т.е. термин «компрессия» может использоваться в исследованиях, посвященных и обозначенным научным направлениям (напомним, что для всех показателей был определен порог значимости и данные показатели силы связи, представленные в модели, не являются случайными).

Компоненты в представленной модели могут рассматриваться не только с позиции двусторонних связей, но и фрагментов модели, представленной сразу несколькими компонентами. Так, фрагмент модели *Компрессия – Текст и дискурс – Язык – Понятийно-категориальная сфера* возможно интерпретировать как реализацию компрессии в языке и тексте посредством понятийно-категориальной сферы, что говорит о концептуальной природе исследований, посвященных рассмотрению механизмов компрессии языка и текста. Добавление к этой структуре элемента *Сфера культуры и лингвокультуры* (все элементы данного отрывка модели соединены сильными связями) очерчивает круг пред-

метной области, в рамках которой осуществляются исследования, в которых используется понятие «компрессия».

ЛИТЕРАТУРА

1. Лукашевич Н. В. Тезаурусы в задачах информационного поиска. М.: изд-во Московского ун-та, 2011. 512 с.
2. Морозова Л. А. Терминознание: Основы и методы. М.: ГНО «Прометей» МПГУ. 2004. С. 143.
3. Виноградов В. А. Термин в научном дискурсе // Вестник Нижегородского ун-та им. Н. И. Лобачевского. 2014. №2(1). С. 368–372.
4. Большой толковый словарь русского языка / под ред. С. А. Кузнецова. СПб.: Норинт, 1998. 1534 с.
5. Brisaboa N., Farina A., Navarro, G., Parama, J. Dynamic Lightweight Text Compression // ACM Transactions on Information Systems. 2010. №28. Pp. 118–126.
6. Peng M., Gao B. L., Zhu J. H., Huang J. J., Yuan M. T., Li F. High quality information extraction and query-oriented summarization for automatic query-reply in social network // Expert Systems with applications. 2016. №44. Pp. 92–101.
7. Bentivogli L., Pianta E. Extending WordNet with Syntagmatic Information // Proceedings of International Wordnet Conference. 2004. Pp. 47–53.
8. Валгина Н. С. Теория текста. М.: изд-во МГУП «Мир книги». 1998. 210 с.
9. Никитин М. В. О компрессии речи на комбинаторно-семантическом уровне // Теория и практика лингвистического описания разговорной речи. 1976. С. 65–70.
10. Буравцова А. Н., Чопорова Е. И. Компрессия текста как вспомогательное средство в информационной деятельности человека // Международный студенческий научный вестник. 2015. №5. С. 79–81.
11. Дюндик Б. П. Компрессия и некоторые вопросы перевода // Бизнес-образование и эффективное развитие экономики. Иркутск: ИГУ, 2007. С. 282–288.
12. Елькин В. В. Диалогическая речь – основная сфера реализации языковой экономии: автореф. дис. ... канд. филол. наук. Пятигорск, 2001. 21 с.
13. Лисник М. В. Речевая компрессия как проявление закона экономии: переосмысление понятия // Вестник Ивановского гос. ун-та. 2009. №1. С. 28–33.
14. Пан Г. В. Языковая экономия и коммуникативная насыщенность текстов: автореф. дисс. ... канд. филол. наук. Ленинград, 1977. 148 с.
15. Умерова М. В. Языковая компрессия: виды и уровни реализации // Актуальные проблемы английской лингвистики и лингводидактики. Сб. научных трудов МГПУ. 2011. №10. С. 272–282.
16. Белоусов К. И., Ерофеева Е. В., Баранов Д. А., Зелянская Н. Л., Щебетенко С. А. Многопараметрический анализ лингвистических данных в информационной системе «Семограф» (на прим. исслед-я речевого поведения пользователей соц. сети) // Вестник Томского гос. ун-та. Филология. 2020. №64. С. 6–29.
17. Белоусов К. И., Баранов Д. А., Ерофеева Е. В., Зелянская Н. Л., Ичкинеева Д. А. Реализация концептуально-гипертекстовой структуры предметной области в журнале «Вопросы когнитивной лингвистики» // Вопросы когнитивной лингвистики. 2015. №2. С. 75–88.
18. Белоусов К. И., Баранов Д. А., Ерофеева Е. В., Зелянская Н. Л., Ичкинеева Д. А. Прогнозирование научной области (на мат-ле ведущего темат. журнала) // Научно-техн. информация. Сер. 2. Информационные процессы и системы. 2014. №10. С. 13–25.
19. Baranov D., Belousov K., Erofeeva E., Leshchenko Y. Semograph information system as a platform for network-based linguistic research: a case study of verbal behavior of social network users // Smart innovation, systems and technologies. 2019. T. 144. С. 313–324.

Поступила в редакцию 01.12.2020 г.

DOI: 10.33184/bulletin-bsu-2020.4.43

MODELING THE CONCEPTUAL POTENTIAL OF THE TERM “COMPRESSION”

© D. A. Ichkineeva

*Bashkir State University
32 Zaki Validi Street, 450076 Ufa, Republic of Bashkortostan, Russia.*

*Phone: +7 (347) 273 66 72.
Email: dilaraichkineeva@gmail.com*

The author of the article offers an overview of scientific publications in which the concept of “compression” is encountered. The relevance of the study of this term is due to its integrative and interdisciplinary nature. To analyze the use of this concept in the works of Russian scientists, the following research program is proposed: with the help of an extended search in the scientific electronic library, all articles on linguistics are identified in titles, keywords, and annotations of which the term “compression” is found. In the “Semograph” Information System, a database of contexts is created in which the term “compression” is encountered. Keywords are considered as contexts, which are subsequently combined into semantic fields. After the grouping of all components into semantic fields in the Information System “Semograph”, a semantic map is built based on the occurrence of components of the field in one context. The semantic map is used to construct a graph-semantic model. As a result of the study, an integral model of the conceptual apparatus of the term “compression” was obtained on the basis of Russian publications. Thus, the proposed methodology for visualizing the conceptual sphere of the term under study using the model makes it possible to solve the problems of analysis, comparison, and assessment of subject areas.

Keywords: compression, text, modeling, keywords, information system “Semograph”.

Published in Russian. Do not hesitate to contact us at bulletin_bsu@mail.ru if you need translation of the article.

REFERENCES

1. Lukashovich N. V. Tezaurusy v zadachakh informatsionnogo poiska [Thesauri in information search problems]. Moscow: izd-vo Moskovskogo un-ta, 2011.
2. Morozova L. A. Terminoznanie: Osnovy i metody [Terminology: Basics and methods]. Moscow: GNO «Prometei» MPGU. 2004. Pp. 143.
3. Vinogradov V. A. Vestnik Nizhegorodskogo un-ta im. N. I. Lobachevskogo. 2014. No. 2(1). Pp. 368–372.
4. Bol'shoi tolkovyi slovar' russkogo yazyka [Comprehensive dictionary of the Russian language]. Ed. S. A. Kuznetsova. Saint Petersburg: Norint, 1998.
5. Brisaboa N., Farina A., Navarro, G., Parama, J. ACM Transactions on information Systems. 2010. No. 28. Pp. 118–126.
6. Peng M., Gao B. L., Zhu J. H., Huang J. J., Yuan M. T., Li F. Expert Systems with applications. 2016. No. 44. Pp. 92–101.
7. Bentivogli L., Pianta E. Proceedings of International Wordnet Conference. 2004. Pp. 47–53.
8. Valgina N. S. Teoriya teksta [Text theory]. Moscow: izd-vo MGUP «Mir knigi». 1998.
9. Nikitin M. V. Teoriya i praktika lingvisticheskogo opisaniya razgovornoj rechi. 1976. Pp. 65–70.
10. Buravtsova A. N., Choporova E. I. Mezhdunarodnyi studencheskii nauchnyi vestnik. 2015. No. 5. Pp. 79–81.
11. Dyundik B. P. Biznes-obrazovanie i effektivnoe razvitie ekonomiki. Irkut-sk: IGU, 2007. Pp. 282–288.
12. El'kin V. V. Dialogicheskaya rech' – osnovnaya sfera realizatsii yazykovoi ekonomii: avtoref. dis. ... kand. filol. nauk. Pyatigorsk, 2001.
13. Lisnik M. V. Vestnik Ivanovskogo gos. un-ta. 2009. No. 1. Pp. 28–33.
14. Pan G. V. Yazykovaya ekonomiya i kommunikativnaya nasyshchennost' tekstov: avtoref. diss. ... kand. filol. nauk. Leningrad, 1977.
15. Umerova M. V. Aktual'nye problemy angliiskoi lingvistiki i lingvodidaktiki. Sb. nauchnykh trudov MGPU. 2011. No. 10. Pp. 272–282.
16. Belousov K. I., Erofeeva E. V., Baranov D. A., Zelyanskaya N. L., Shchebetenko S. A. Vestnik Tomskogo gos. un-ta. Filologiya. 2020. No. 64. Pp. 6–29.
17. Belousov K. I., Baranov D. A., Erofeeva E. V., Zelyanskaya N. L., Ichkineeva D. A. Voprosy kognitivnoi lingvistiki. 2015. No. 2. Pp. 75–88.
18. Belousov K. I., Baranov D. A., Erofeeva E. V., Zelyanskaya N. L., Ichkineeva D. A. Nauchno-tekhn. informatsiya. Ser. 2. Informatsionnye protsessy i sistemy. 2014. No. 10. Pp. 13–25.
19. Baranov D., Belousov K., Erofeeva E., Leshchenko Y. Smart innovation, systems and technologies. 2019. Vol. 144. Pp. 313–324.

Received 01.12.2020.