

А. С. Пташкин

ДИСКУРС, ДИСКУРСИВНЫЙ АНАЛИЗ И ЛИНГВИСТИЧЕСКИЕ ПРОЦЕССОРЫ

Рассматривается дискурс с позиции философии, социологии, социолингвистики, культурологии, семиотики, лингвистики текста. Дискурсивный анализ описывается с точки зрения текстуального и контекстуального аспектов. Выявляются уровни анализа языкового материала с целью характеристики дискурса. Фреймы (концептуальные системы) рассматриваются как ментальные единицы, содержание которых может пониматься как дискурс. Представлен краткий обзор лингвистических процессоров MATLAB, GATE и базы онтологий WordNet, которые являются наиболее оптимальными автоматизированными средствами дискурс-анализа естественного языка. Среда Gate сочетает в себе элементы анализа корпусов, концептов, машинного обучения. Данный набор программ позволяет установить принадлежность маркеров к выявленному слоту когнитивной модели дискурсивных маркеров. Программа MATLAB помогает автоматизировать алгоритм работы с зерновыми цепями.

Ключевые слова: дискурс, дискурсивный анализ, лингвистический процессор, фрейм, дискурсивный маркер, GATE.

Анализ дискурса является междисциплинарной областью знания. Дискурс представлен в психолингвистике, семиотике, риторике, лингвистике текста, литературоведении, этнологии, антропологии, философии, социологии, политологии и пр. [1, 2].

Л. С. Бейлинсон выделяет в изучении дискурса социальный, прагматический, семантический, формально-языковой параметры [3]. Автором социально-коммуникативного подхода к дискурсу является Юрген Хабермас [4, 5]. С позиции постмодернистского подхода в рамках дискурса обнаруживаются узловые точки, которые рассматриваются «в качестве *привилегированных знаков*» [6]. В рамках культурологии «дискурс» понимается как (а) совокупность рассуждений, либо (б) любая целенаправленная речевая практика [7].

В лингвистике (стилистике, лингвистике текста, исследованиях разговорной речи, когнитивной лингвистике) дискурс часто понимается как семантическое пространство, характеризующееся когезией [8]. Термин «дискурс» введен З. Хэррисом в 1952 г. [9] как «метод изучения движения информации в дискурсе» [10].

Для лингвиста текст представляет собой в первую очередь фиксированный в письменной форме языковой материал [11], необходимый для «развертывания дискурсивного процесса» [12]. «Термин „дискурс“ означает ‘речь, процесс языковой деятельности; способ говорения’» [13].

Дискурс часто понимается как форма социального поведения, в котором учитываются культурные, исторические и прочие факторы. При этом дискурс не может выступать как отдельный текст. В целом можно сказать, что текст является компонентом дискурса. Под дискурсом также необходимо понимать «коллекцию высказываний», принадлежащих к одной и той же теме – содержание такого дискурса передает целый комплекс различных

текстов, объединенных в корпус или иное другое хранилище. Дискурс является процессом ментальным и коммуникативным одновременно, результатом которого необходимо считать текст. При этом текст выполняет функцию статическую, а дискурс процессуален.

Наблюдение, эксперимент и моделирование являются основными методами в рамках дискурсивного анализа [14]. Согласно В. Е. Чернявской, следует различать следующие уровни анализа языкового материала, характеризующие дискурсивный анализ в целом: «1) уровень (отдельного) коммуникативного акта; 2) уровень (отдельного) текста; 3) уровень глубинного макросемантического анализа текста, его смысловой структуры; 4) уровень над- и вокругтекстового анализа; 5) уровень межтекстового дискурсивного анализа» [15].

Лингвистический анализ следует начинать с определенного текста, а затем необходимо установить связность на тематическом уровне между текстами в дискурсе.

Согласно взглядам Т. А. ван Дейка, дискурсивный анализ включает в себя два аспекта: текстуальный и контекстуальный [16, 17]. Кроме того, Т. А. ван Дейк выделял знание как «когнитивный портрет» участников дискурса. Знание организовано в фреймы... Понимание дискурса может предположительно включать в себя общее содержание фрейма [18, с. 3–152].

В данной работе под термином «дискурс-анализ» понимается сфера исследования коммуникации с учетом ее формы, функциональной стороны, контекстуальных условий, социальной среды, в рамках которой существует дискурс. Стоит согласиться с общепринятым определением дискурсивного анализа как междисциплинарного явления, разработкой которого занимаются не только лингвисты, психологи, социологи и другие, но и математики, программисты. Перспективным направле-

нием, в котором могли бы соединиться лингвистические и нелингвистические подходы к дискурсу (контент-анализ, конверсационный анализ, критический дискурс-анализ, symlog, этнографические методы, анализ способа категоризации участия и прочие), является метод автоматического дискурсивного анализа.

Метод автоматизированного дискурсивного анализа. Программы обработки связанных текстов называют лингвистическими процессорами [19, с. 272–273]. Большинство таких сред лингвистического анализа включают в себя онтологию английского языка WordNet. Например, лексема *deficiency* в тезаурусе представлена следующим образом: **deficiency** noun (6){14472871} <noun.state> [26] S: (n) lack#1 (lack%1:26:00::), **deficiency#1 (deficiency%1:26:00::)**, want#2 (want%1:26:01::) (the state of needing something that is absent or unavailable) *there is a serious lack of insight into the problem*"; *water is the critical deficiency in desert regions*"; *for want of a nail the shoe was lost*"; (2){05120433} <noun.attribute>[07] S: (n) insufficiency#3 (insufficiency%1:07:00::), inadequacy#1 (inadequacy%1:07:01::), **deficiency#2 (deficiency%1:07:00::)** (lack of an adequate quantity or number) *the inadequacy of unemployment benefits*" [20]. В данной словарной статье «(the state of needing something that is absent or unavailable)» и «(lack of an adequate quantity or number)» являются глоссами, числа (6) и (2) частотность слов (frequency count), номера {14472871} и {05120433} являются локациями в базе данных, сведения о лексемах расположены после номеров в коллекции тезауруса – <noun.state>, <noun.attribute>, [26], [7] – номера лексических файлов в системе, (lack%1:26:00::), (want%1:26:01::), (insufficiency%1:07:00::), (inadequacy%1:07:01::) – смысловые ключи (sense keys), #1, #2, #3 – порядковые номера значений. Слова разбиты по частям речи. Для всех слов представлены синсеты – понятия.

Пакет программ MATLAB позволяет автоматизировать алгоритм работы с зерновыми концептами – основными смысловыми центрами. В рамках данного алгоритма структурирования текста выявляются следующие этапы: 1) задаются зерновые концепты статьи; 2) дискурс-анализ с целью выявления «центров напряжения» (зерновых концептов); 3) добавление дополнительных концептов, выступающих посредниками между зерновыми концептами; 4) укрупнение концептов-посредников в «смысловые гнезда» [21].

Кроме того, существуют комбинированные программы с элементами корпусного анализа, анализа концептов, машинного обучения [22]. К примеру, пакет программ GATE [23]. Он является платформой для разработки среды изучения дискурсивных

маркеров [24] (*discourse markers* – verbal expressions such as *anyway*, *oh well*, or *I mean* that demarcate units of speech [25, с. 214]. *Discourse marker* n. (grammar) a word or phrase that organizes spoken language into different parts [26, с. 432].

Среда GATE применима к описанию дискурсивных маркеров. С помощью данного программного комплекса предоставляется возможность определить принадлежность маркера «к определенному функциональному слоту когнитивно-функциональной модели дискурсивных маркеров» [27, с. 204]. В работе над описанием дискурсивных маркеров можно использовать метаязык поисковых шаблонов JAPE. Jape Transducer – язык регулярных выражений над аннотациями с использованием логических операторов.

В среде GATE представлены: фреймворк для лингвистических задач, набор готовых инструментов и пр. GATE предоставляет готовые ресурсы (токенизаторы, стеммеры) и прототипы, на основе которых можно создавать собственные процессоры [28]. Из недостатков программы следует отметить отсутствие возможности задавать контекст. Кроме того, необходим метод обработки информации создаваемых модулей, а также плагин выгрузки данных сформированных корпусов.

Ключевыми компонентами программной системы следует считать модули обработки текстов, базы данных тезаурусов (или иных источников текстов) и «память» (semantic storage). Методы и цели исследования определяют набор технических средств, приложений, процессоров, необходимых для достижения задач исследования. Создание универсального лингвистического процессора в рамках дискурс-анализа является трудоемким процессом. Данная проблема связана отчасти с большим количеством методов дискурс-анализа. Одним из удачных направлений дальнейшего развития лингвистических процессоров могло бы стать соединение их функций с программами, направленными на сбор словарей и анализа текстов с возможностью создания онлайн-корпусов в автоматическом режиме, а также выгрузки данных известных корпусов по определенной теме в удобном унифицированном формате.

Лингвистические процессоры в комплексе с корпусными базами являются мощным средством дискурсивного анализа с учетом традиционных методов изучения текста. Представляется эффективным исследование определенного тематического корпуса с использованием когнитивных методик (фреймвого анализа). Так называемая рамочная структура ментальных единиц в языке реконструируется благодаря языковым средствам в тексте. Таким образом, выводится общее знание об определенном дискурсе (явлении, факте и т. д.).

Список литературы

1. Фаритов В. Т. Онтологический статус феномена дискурса // Эпистемология и философия науки. 2014. Т. XXXIX, № 1. С. 124–136.
2. Новейший философский словарь / гл. ред. А. А. Грицанов. Минск: В. М. Скакун, 1999. 896 с.
3. Бейлинсон Л. Профессиональный дискурс как предмет лингвистического изучения // Вестн. Волгоградского гос. ун-та. Серия 2: Языкознание. 2009. № 1. С. 145.
4. Русакова О. Ф. PR-дискурс: Теоретико-методологический анализ. Екатеринбург: Институт международных связей, 2008. 340 с.
5. Хабермас Ю. Вовлечение другого. Очерки политической теории / пер. с нем. Ю. Медведева. СПб.: Наука, 2008. 424 с.
6. Русакова О. Ф. Основные теоретико-методологические подходы к интерпретации дискурса // Научный ежегодник Института философии и права Уральского отделения Российской академии наук. 2007. № 7. С. 5–34.
7. Кожемякин Е. А. Дискурсивный подход к изучению культуры // Современный дискурс-анализ. 2016. Вып. 1, т. 1. URL: <http://discourseanalysis.org/ada1/st6.shtml> (дата обращения: 06.02.2016).
8. Темнова Е. В. Современные подходы к изучению дискурса // Язык, сознание, коммуникация: сб. статей / отв. ред. В. В. Красных, А. И. Изотов. М.: МАКС Пресс, 2004. Вып. 26. 168 с.
9. Harris Z. S. Discourse Analysis // Language. 1952. Vol. 28, № 1. P. 1–30.
10. Григорьева В. С. Дискурс как элемент коммуникативного процесса: прагмалингвистический и когнитивный аспекты. Тамбов: Изд-во Тамбов. гос. техн. ун-та, 2007. 288 с.
11. Щерба Л. В. Языковая система и речевая деятельность. Л.: Наука, 1974. 428 с.
12. Борботько В. Г. Принципы формирования дискурса: от психолингвистики к лингвосинергетике. Изд. 4-е. М.: Либроком, 2011. 288 с.
13. Кибрик А. А. Анализ дискурса в когнитивной лингвистике: дис. ... д-ра филол. наук. М., 2003. 90 с.
14. Кибрик А. А. Дискурсивный анализ // Энциклопедия «Фонд знаний – Ломоносов». URL: <http://www.lomonosov-fund.ru/enc/ru/encyclopedia:0127182:article> (дата обращения: 02.02.2016).
15. Чернявская В. Е. Лингвистика текста. Лингвистика дискурса: учеб. пособие. 3-е изд., стер. М.: Флинта; Наука, 2015. 208 с.
16. Дейк Т. А. ван. Язык. Познание. Коммуникация. Благовещенск: БГК им. И. А. Бодуэна де Куртенэ, 2000. 308 с.
17. Шапочкин Д. В. Метод когнитивного анализа дискурса в лингвистике // Вестн. Челябинского государственного университета. 2013. № 10 (301). С. 101–107.
18. Минский М. Фреймы для представления знаний. М.: Энергия, 1979. 152 с.
19. Тонконогов Д. Методы автоматизированного дискурсивного анализа неструктурированных текстов в заданном контексте // International Journal "Information Technologies & Knowledge". 2013. Vol. 7 (3). С. 272–276.
20. WordNet. Deficiency. URL: <http://wordnetweb.princeton.edu/perl/webwn?c=7&sub=Change&o2=1&o0=1&o8=1&o1=1&o7=1&o5=1&o9=1&o6=1&o3=1&o4=1&i=-1&h=00&s=deficiency> (дата обращения: 01.02.2016).
21. Градосельская Г. В. Сетевой анализ постсоветского информационного пространства: перспективы разработки методологии: сборник статей памяти А. Крыштановского. НИУ ВШЭ, РОС, ИС РАН. М.: НИУ ВШЭ, 2011. 557 с.
22. Hilbert M. A Text-technological Approach to Automatic Discourse Analysis of Complex Texts. Gießen: Institut für Germanistik Arbeitsbereich Angewandte Sprachwissenschaft und Computerlinguistik Justus-Liebig-Universität, 2010. 4 S.
23. Gate General Architecture overview – The University of Sheffield. URL: <http://gate.ac.uk/overview.html> (дата обращения: 04.02.2016).
24. Болсуновская Л. М., Зеремская Ю. А., Дубровская Н. В. Виды дискурсивных маркеров в русскоязычных и англоязычных научных статьях по геологии и нефтегазовому делу // Вестн. Томского гос. пед. ун-та (TSPU Bulletin). 2015. Вып. 4 (157). С. 117–123.
25. Colman A. M. The Dictionary of Psychology. Oxford: Oxford University Press, 2015. 896 p.
26. Oxford Advanced Learner's Dictionary of Current English / ed. by J. Tumbull. Oxford: Oxford University Press, 2010. 1796 p.
27. Каменский М. В. Автоматизированный анализ корпуса текстов как методологическая основа построения и верификации когнитивно-функциональной модели дискурсивных маркеров // Вестн. Ленинградского государственного университета им. А. С. Пушкина. 2013. № 4 (Т. 1. Филология). С. 200–207.
28. Кузнецов И. Разработка лингвистических процессоров в среде GATE // Mathlingvo. 2012. URL: <https://www.hse.ru/data/2012/03/23/1264653949/gate-mathlingvo-3.3.pdf> (дата обращения: 29.01.2016).

Пташкин А. С., кандидат филологических наук, доцент.

Национальный исследовательский Томский политехнический университет.

Пр. Ленина, 30, Томск, Россия, 634050.

Томский государственный педагогический университет.

Ул. Киевская, 60, Томск, Россия, 634061.

E-mail: pt.alexandr@gmail.com

Материал поступил в редакцию 29.02.2016.

A. S. Ptashkin

DISCOURSE, DISCOURSE ANALYSIS AND LINGUISTIC PROCESSORS

This article deals with the analysis of the discourse in the light of philosophy, sociology, sociolinguistics, culturology, semiotics, text linguistics. The discourse analysis has been depicted from the point of text and context aspects. The levels of analysis of language material for discourse characterization have been distinguished. Frames (concept systems) have been considered as the mental units; their contents are being seen as the discourse. The brief review of linguistic processors MathLab, Gate and WordNet data-base has been done; they are defined as the most effective and automated means of discourse analysis of natural languages. The environment Gate includes the elements of corpus and concept analyses and machine learning. These programmes help to define the slot position of markers in their cognitive model. MathLab automates the algorithm in the analysis of the core concepts.

Key words: *discourse, discourse analysis, linguistic processor, frame, discourse marker, GATE.*

References

1. Faritov V. T. Ontologicheskiy status fenomena diskursa [Ontological status of discourse phenomenon]. *Epistemologiya i filosofiya nauki – Epistemology and Philosophy of Science*, 2014, vol. XXXIX, no. 1, pp. 124–136 (in Russian).
2. *Noveyshiyy filosofskiy slovar'* [New Philosophical Dictionary]. Minsk, 1999. 896 p. (in Russian).
3. Beylinson L. S. Professional'nyy diskurs kak predmet lingvisticheskogo izucheniya [Professional discourse as the subject of linguistic study]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2: Yazykoznanie – Science Journal of Volgograd State University. Linguistics*, 2009, no. 1, p. 145 (in Russian).
4. Rusakova O. F., Rusakov V. M. *PR-diskurs: Teoretiko-metodologicheskyy analiz* [PR-discourse: theoretical and methodological analysis]. Ekaterinburg, UrO RAN, Institut mezhdunarodnykh svyazey Publ., 2008. 340 p. (in Russian).
5. Habermas Yu. *Vovlecheniye drugogo. Ocherki politicheskoy teorii* [The inclusion of the other: studies in political theory]. St. Petersburg, Nauka Publ., 2008. 424 p. (in Russian).
6. Rusakova O. F. Osnovnyye teoretiko-metodologicheskiye podkhody k interpretatsii diskursa [Main theoretical and methodological approaches to discourse interpretation]. *Nauchnyy ezhegodnik Instituta filosofii i prava Ural'skogo otdeleniya Rossiyskoy akademii nauk – Science Year-Book of the Institute of Philosophy and Law of the Russian Academy of Sciences, Ural Branch*, 2007, no. 7, pp. 5–34 (in Russian).
7. Kozhemyakin E. A. Diskursnyy podkhod k izucheniyu kul'tury [Discourse approach to culture study. Modern discourse analysis]. *Sovremennyy diskurs-analiz – Modern Discourse Analysis*, 2016, issue 1, vol. 1. URL: <http://discourseanalysis.org/ada1/st6.shtml> (accessed 6 February 2016) (in Russian).
8. Temnova E. V. Sovremennyye podkhody k izucheniyu diskursa [Modern approaches to discourse study]. *Yazyk, soznaniye, kommunikatsiya: sb. statey* [Language, consciousness, communication: collection of articles]. Moscow, MAKSPress Publ., 2004. 168 p. (in Russian).
9. Harris Z. S. Discourse Analysis. *Language*, 1952, vol. 28, no. 1, pp. 1–30.
10. Grigor'eva V. S. *Diskurs kak element kommunikativnogo protsessa: pragmalingvisticheskiy i kognitivnyy aspekty* [Discourse as an element of a communicative process: pragmalinguistic and cognitive aspects]. Tambov, Izd-vo Tamb. gos. tekhn. un-ta Publ., 2007. 288 p. (in Russian).
11. Shcherba L. V. *Yazykovaya sistema i rechevaya deyatel'nost'* [Language system and speech activity]. Leningrad, Nauka Publ., 1974. 428 p. (in Russian).
12. Borbot'ko V. G. *Printsipy formirovaniya diskursa: ot psikholingvistiki k lingvosinergetike* [Principles of discourse forming: from psycholinguistics to linguosynergetics]. Moscow, Librokom Publ., 2011. 288 p. (in Russian).
13. Kibrik A. A. *Analiz diskursa v kognitivnoy lingvistike*. Dis. dokt. filol. nauk [Discourse analysis in cognitive linguistics. Diss. doct. of philol. sci.]. Moscow, 2003. 90 p. (in Russian).
14. Kibrik A. A. Diskursivnyy analiz [Discourse analysis]. *Entsiklopediya "Fond znaniy – Lomonosov"* [Encyclopedia "Fund of Knowledge – Lomonosov"]. URL: <http://www.lomonosov-fund.ru/enc/ru/encyclopedia:0127182:article> (accessed 2 February 2016) (in Russian).
15. Chernyavskaya V. E. *Lingvistika teksta. Lingvistika diskursa* [Textual linguistics. Linguistics of discourse]. Moscow, Flinta, Nauka Publ., 2015. 208 p. (in Russian).
16. Deyk T. A. van. *Yazyk. Poznanie. Kommunikatsiya* [Language. Perception. Communication]. Blagoveshensk, BGK im. I. A. Boduena de Kurtene Publ., 2000. 308 p. (in Russian).
17. Shapochkin D. V. Metod kognitivnogo analiza diskursa v lingvistike [Methods of cognitive analysis of discourse in linguistics]. *Vestnik Chelyabinskogo gosudarstvennogo universiteta – Vestnik ChelGU*, 2013, no. 10 (301), pp. 101–107 (in Russian).
18. Minskiy M. *Freymy dlya predstavleniya znaniy* [Frames for knowledge representation]. Moscow, Energiya Publ., 1979. 152 p. (in Russian).
19. Tonkonogov D. Metody avtomatizirovannogo diskursivnogo analiza nestruturirovannykh tekstov v zadannom kontekste [Method of automated discourse analysis of the texts with no structure in a certain context]. *International Journal "Information Technologies & Knowledge"*, 2013, vol. 7 (3), pp. 272–276 (in Russian).
20. *Word Net. Deficiency*. URL: <http://wordnetweb.princeton.edu/perl/webwn?c=7&sub=Change&o2=1&o0=1&o8=1&o1=1&o7=&o5=1&o9=&o6=1&o3=1&o4=1&i=-1&h=00&s=deficiency> (accessed 1 February 2016).

21. Gradoseĭskaya G. V. *Setevoy analiz postsovetskogo informatsionnogo prostranstva: perspektivy razrabotki metodologii – Sbornik statey pamyati A. Kryshchanovskogo* [Net analysis of post-Soviet information space: perspectives of methodology development – collection of articles in memory of A. Kryshchanovskogo]. Moscow, NIU VShE Publ., 2011. 557 p. (in Russian).
22. Hilbert M. *A Text-technological Approach to Automatic Discourse Analysis of Complex Texts*. Gießen, Institut für Germanistik Arbeitsbereich Angewandte Sprachwissenschaft und Computerlinguistik Justus-Liebig-Universität, 2010. 4 S.
23. *Gate General Architecture overview – The University of Sheffield*. URL: <http://gate.ac.uk/overview.html> (accessed 4 February 2016).
24. *The Dictionary of Psychology*. Oxford, Oxford University Press, 2015. 896 p.
25. Bolsunovskaya L. M., Zeremskaya Yu. A., Dubrovskaya N. V. Vidy diskursivnykh markerov v russkoyazychnykh i angloyazychnykh nauchnykh stat'yakh po geologii i neftegazovomu delu [Types of discourse markers in Russian and English science articles in the sphere of geology, oil and gas]. *Vestnik Tomskogo gosudarstvennogo pedagogicheskogo universiteta – TSPU Bulletin*, 2015, vol. 4 (157), pp. 117–123 (in Russian).
26. *Oxford Advanced Learner's Dictionary of Current English*. Oxford, Oxford University Press, 2010. 1796 p.
27. Kamenskiy M. V. Avtomatizirovanny analiz korpusa tekstov kak metodologicheskaya osnova postroeniya i verifikatsii kognitivno-funktional'noy modeli diskursnykh markerov [Automated analysis of corpus texts as the methodological basis of constructing and verification of cognitive and functional model of discourse markers]. *Vestnik Leningradskogo gosudarstvennogo universiteta im. A. S. Pushkina – Pushkin Leningrad State University Bulletin*, 2013, no. 4, pp. 200–207 (in Russian).
28. Kuznetsov I. *Razrabotka lingvisticheskikh protsessorov v srede GATE* [Development of linguistic processors in GATE]. Mathlingvo. 2012. URL: <https://www.hse.ru/data/2012/03/23/1264653949/gate-mathlingvo-3.3.pdf> (accessed 29 January 2016) (in Russian).

Ptashkin A. S.

National Research Tomsk Polytechnic University.

Pr. Lenina, 30, Tomsk, Russia, 634050.

Tomsk State Pedagogical University.

Ul. Kievskaya, 60, Tomsk. Russia, 634061.

E-mail: pt.alexandr@gmail.com