

МОДЕЛИРОВАНИЕ СЕМАНТИЧЕСКОГО ПРОСТРАНСТВА НАРРАТИВНОГО ТЕКСТА В СТАТИКЕ И ДИНАМИКЕ

Т. В. Ефимова

Воронежская государственная медицинская академия им. Н. Н. Бурденко

Поступила в редакцию 17 сентября 2011 г.

Аннотация: в работе предлагается оригинальная методика лингвистического моделирования семантического пространства текста в статике и динамике с использованием математического инструментария и средств компьютерной графики на материале наиболее сложного случая – художественного текста.

Ключевые слова: моделирование семантического пространства, кореферентные наименования, предикативные связи, статическая модель, динамическая модель, семантическая сеть.

Abstract: the paper presents the original method of linguistic modeling of the text semantic space in its statics and dynamics. Fiction texts are analyzed with the help of mathematical tools and computer graphic technique.

Key words: semantic space modeling, co-referent names, predicative relations, static model, dynamic model, semantic network.

В данной работе предлагается методика моделирования семантического пространства нарративного текста как процесса и результата его развертывания с целью выявления существенной информации.

Материалом для исследования послужил рассказ А. П. Чехова «Дама с собачкой», хорошо структурированный, небольшой по объему текст как обозримое целое для анализа, но с достаточным материалом для классификации.

Выявление семантически значимой информации текста происходит во многом благодаря использованию средств компьютерной графики. Высокая эффективность графического представления информации подтверждена многочисленными исследованиями наглядно-образного и визуального мышления на основе когнитивного подхода. Фундаментальная идея этого подхода заключается в том, что мышление есть не что иное, как манипулирование внутренними (мысленными) репрезентациями структурированных определенным образом знаний – фреймов, сетей, планов, сценариев. Исходя из этого разработаны модели представления знаний на основе семантических сетей, продукций, логики предикатов и нечетких знаний [1].

Следует отметить, что разработанное на основе когнитивного подхода новое направление в области искусственного интеллекта – *когнитивная компьютерная графика* (ККГ) – с 1988 г. является «стратегическим приоритетом национальной политики США в области развития перспективных информационных технологий, прежде всего для науки и образования.

Суть концепции ККГ очень проста: если на экране дисплея удастся визуализировать существенные свойства и отношения между объектами некоторой предметной области (даже любой степени абстракции), то такой ККГ-образ, как правило, содержит в себе информацию (на уровне графических деталей компьютерного изображения) о возможных и не всегда заранее известных следствиях этих свойств и этих отношений» [2, с. 37].

Предлагаемая нами процедура дает возможность наглядно представить данные, полученные на этапе лингвистического анализа (отчасти формализованные в виде таблиц и структурных схем), которые необходимы для решения ряда семантических задач, в частности, предполагающих в той или иной форме моделирование семантического пространства текста в статике и динамике. Речь идет, прежде всего, о визуализации полученной информации в виде графических образов (далее – ГО), которые трактуются как модели отображаемого объекта – текста, как средство представления данных исследований сложных структур, явлений, процессов [3–9].

Для построения статической модели выделяются ключевые слова (далее – КС) для получения полной информации о номинировании персонажей (действующих лиц, участников ситуации) с опорой на анализ употребительности их номинаций в тексте и на определение кореферентного тождества слов и словосочетаний. Полученные данные представляются в относительных величинах путем составления пропорций с учетом общего количества кореферентных наименований КС и связей между ними через предикацию.

Статическая модель в виде графа, состоящего из вершин и ребер, отражает как относительный вес КС, так и силу связей между ними в целом по тексту. Вершины в виде круга – это персонажи произведения с указанием их имен-гиперонимов и количества кореферентных наименований; ребра в виде стрелок отражают связи между действующими лицами через предикацию и силу этих связей (рис. 1).

Площадь круга вычисляется по формуле:

$$S = \pi R^2 = \frac{\pi D^2}{4}, \quad S = Cr,$$

где S – площадь круга, Cr – количество кореферентных наименований персонажей. Толщина стрелок на схеме пропорциональна количеству предикативных (семантико-синтаксических) связей между КС.

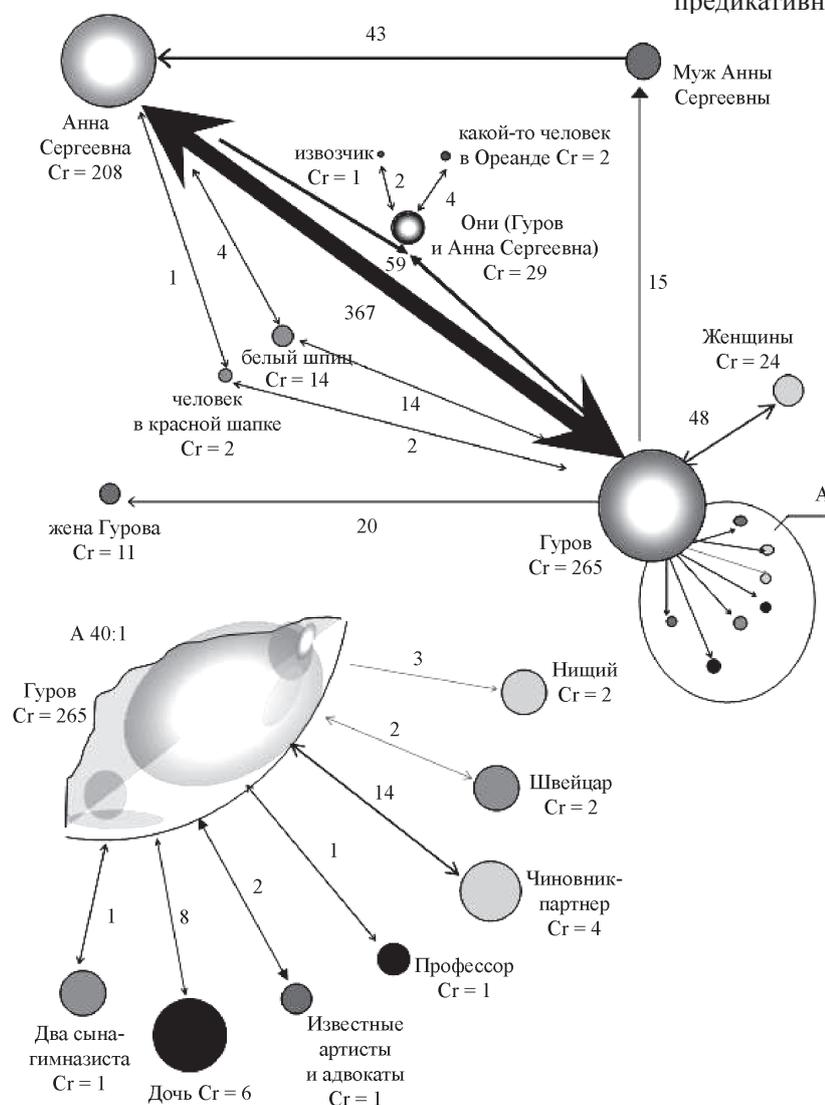


Рис. 1. Схема связей (в целом по тексту)

Как известно, каждую семантико-синтаксическую структуру формирует определенный тип предиката с относящимися к нему актантами. Каждый семантический тип предиката эксплицируется соответствующей предикатной лексикой – предикаторами, обладающими общностью грамматических и семантических признаков, предикаторами одного и того же валентного типа. Под предикатором нами, вслед за Ю. С. Степановым, понимается языковая форма предиката (глагол, прилагательное), «понятийной сущности», «явления семантики» [10, с. 321].

Выделение семантических типов предикатов и их структур осуществлялось в два этапа: на первом этапе дифференцировались конкретные высказывания, а на втором – идентифицировались и обобщались однотипные конкретные высказывания как по форме, так и по смыслу.

В результате анализа связи номинаций действующих лиц через предикацию выявлено 11 типов предикативной связи (Р-типов): акциональный (P_{act}),

перцептивный (P_{perc}), экзистенциальный (P_{exist}), сущностный (P_{onom}), квалификативно-дескриптивный (P_{descr}), социативный (P_{soc}), possessивный (P_{poss}), коммуникативно-контактный ($P_{comm-cont}$), предикаты речи (P_{del}), движения (P_{mouv}) и состояния (P_{stat}), каждый из которых на семантико-синтаксическом уровне эксплицируется определенными структурами.

Формализация предикативной связи позволяет выявить предикативную доминанту в целом по тексту (Р-доминанту), критерием релевантности которой является самая высокая частотность Р-типа. Как показал качественный и количественный анализ Р-связи, такой Р-доминантой оказался перцептивный Р-тип. Визуализация данных в виде круговой диаграммы дает возможность продемонстрировать соотношение типов предикативной связи и Р-доминанту в целом по тексту (рис. 2).

Для построения динамической модели семантического пространства текста применена семантическая сеть (СемСеть) как инструмент содержательного членения текста и мощное наглядное средство описания знаний. Особой популярностью сетевой метод пользуется у исследователей в области искусственного интеллекта,

так как сети удобны для чтения и компьютерной обработки текста [11–19, и др.].

Идеальной математической моделью такого системно-структурного образования, как СемСеть, является граф, чьи вершины соответствуют единицам текста, а ребра – семантическим отношениям между ними: «Большое количество узлов и дуг делают графическое представление похожим на сеть из линий, поэтому оно получило название *семантической сети* (семантической, поскольку исторически такие сети были использованы, прежде всего, для представления знаний на естественном языке)» [18, с. 262].

Вершины нашего графа – это 33 сложных синтаксических целых (ССЦ), полученных в результате проведенного сегментирования текста, а ребра, соединяющие соответствующие вершины, маркируют семантическую связь между ССЦ (рис. 3).

Следует заметить, что семантическая сеть не может передавать стилистические и экспрессивно-оценочные отношения, поэтому ее следует рассматривать лишь как модель системы семантических отношений между ССЦ и модель семантической структуры текста.

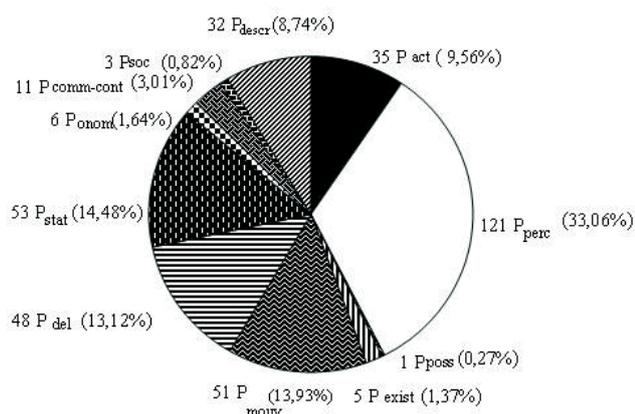


Рис. 2. Гуров – Анна Сергеевна, характер Р-связи (в целом по тексту)

Применяя сетевой метод, мы определили ряд важных количественных параметров, представляющих особую ценность для характеристики системы семантических связей в тексте, а именно: семантическую связность, характеризующую план содержания текста, его семантическую монолитность; цепочный коэффициент, определяющий среднюю длину цепочки ССЦ, в которой каждое предшествующее ССЦ непосредственно связано с последующим за ним ССЦ; функциональный вес ССЦ, параметр, определяемый числом семантических связей, в которые вступает в данном тексте рассматриваемая единица.

В результате сетевого моделирования выявлено девять частей СемСети. Для построения модели каждой из них используется тот же способ представления данных в относительных величинах, что и при создании статической модели.

Построенные для каждой части СемСети девять графов и соответствующее количество диаграмм визуализируют существенные отношения между КС и позволяют измерить динамику отношений между персонажами через количество и качество синтаксических связей их номинаций, опосредованных глаголами указанных семантических типов.

Чтобы получить целостную картину пошаговой динамики Р-связи по частям СемСети и в целом по тексту, а также поведения каждого из Р-типов в отдельности, вся полученная информация представляется на графиках, репрезентирующих динамику самых частотных Р-типов (P_{perc} , P_{stat} , P_{mouv} , P_{del}), среднечастотных Р-типов (P_{act} , P_{descr} , $P_{comm-cont}$) и наименее частотных Р-типов (P_{exist} , P_{onom} , P_{soc} , P_{poss}); 11 кривых последнего графика соответствуют всем 11-ти типам Р-связи (рис. 4).

Таким образом, примененная нами процедура по извлечению текстовых референтов, смысловых единиц – предикатов, их классификации по типам, кван-

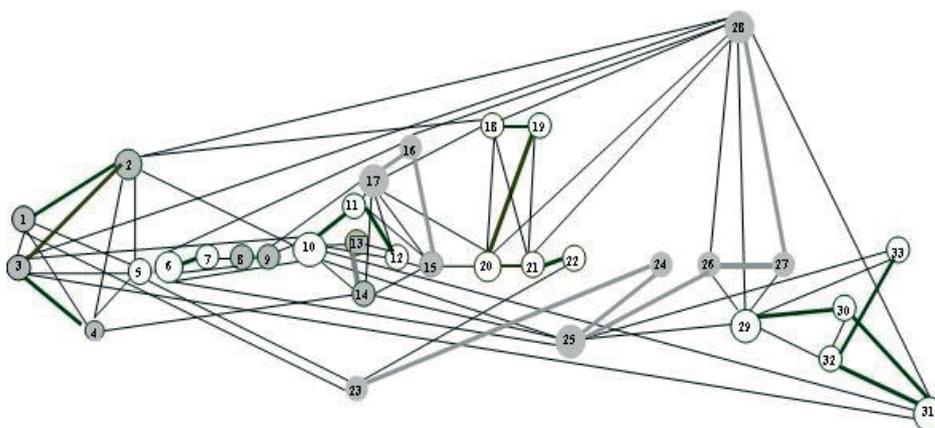


Рис. 3. Семантическая сеть рассказа «Дама с собачкой»

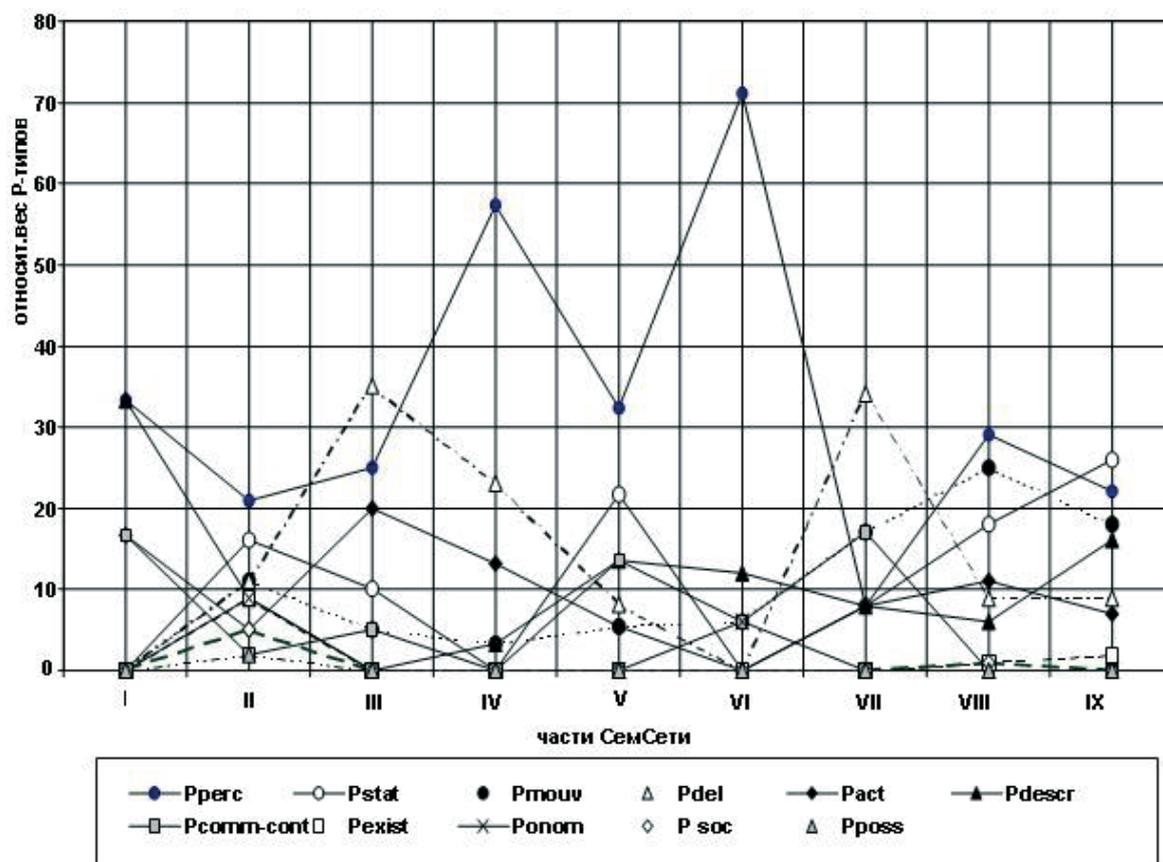


Рис. 4. Динамика Р-связи в целом по тексту

тификации данных, ранжированию по частоте встречаемости, выявлению Ст-доминанты и Р-доминанты для каждой части СемСети и по тексту в целом, – позволяет существенно углубить содержательный анализ текста, сделать его, возможно, более объективным: выявить динамику и характер отношений персонажей, сделать более очевидным намерение автора акцентировать те или иные стороны ситуации. Достоверность выводов, полученных формальным способом, подтверждается при их сопоставлении с результатами, полученными литературоведами посредством неформального, интуитивного анализа [20–26].

Так, статическая модель рассказа, с одной стороны, выявляет связи между КС; демонстрирует приоритетность такого персонажа, как Гуров, визуальным способом доказывая, что именно он является объектом внимания автора, а не дама с собачкой. С другой стороны, мы полагаем, что такое представление образа содержания художественного произведения не противоречит замыслу автора. Подтверждением тому служат слова самого А. П. Чехова: «Поневоле, делая рассказ, – пишет он своему издателю Суворину, – хлопчешь, прежде всего, о его рамках: из массы героев берешь только одно лицо – жену или

мужа, – кладешь это лицо на фон и рисуешь только его; его и подчеркиваешь, а остальные разбрасываешь по фону, как мелкую монету, и получаешь нечто вроде небесного свода: одна большая луна и вокруг нее масса маленьких звезд» [25, с. 186].

В результате формализации предикативной связи в целом по тексту установлено, что Р-доминантой являются P_{perc} (1-я позиция) и P_{stat} (2-я позиция), что подтверждает мнение А. П. Чудакова о том, что в произведениях писателя «большое место занимают конструкции с глаголами восприятия и словами категории состояния» [23, с. 5].

Сетевое моделирование позволило визуализировать систему семантических связей между единицами этой системы – ССЦ, семантическую структуру такого объекта повышенной сложности, как художественный текст, и определить ее тип – это кусочно-нелинейная, т.е. включающая несколько фрагментов, незамкнутая – не имеющая связи между первым и последним ССЦ; ветвящаяся структура. Это объясняется «многоплановостью художественного произведения, в котором сюжет развивается в нескольких направлениях» [18, с. 140].

Применяя сетевой метод, мы определили, что максимальный функциональный вес имеет 17-е ССЦ:

«Пройдет какой-нибудь месяц, и Анна Сергеевна, казалось ему, покроется в памяти туманом и только изредка будет сниться с трогательной улыбкой, как снились другие. Но прошло больше месяца, наступила глубокая зима, а в памяти все было ясно, точно расстался он с Анной Сергеевной только вчера. И воспоминания разгорались все сильнее. Доносились ли в вечерней тишине в его кабинет голоса детей, приготовлявших уроки, слышал ли он романс, или орган в ресторане, или завывала в камине метель, как вдруг воскресало в памяти все: и то, что было на молу, и раннее утро с туманом на горах, и пароход из Феодосии, и поцелуи. Он долго ходил по комнате, и вспоминал, и улыбался, и потом воспоминания переходили в мечты, и прошедшее в воображении мешалось с тем, что будет. Анна Сергеевна не снилась ему, а шла за ним всюду, как тень, и следила за ним. Закрывши глаза, он видел ее, как живую, и она казалась красивее, моложе, нежнее, чем была; и сам он казался себе лучше, чем был тогда, в Ялте. Она по вечерам глядела на него из книжного шкафа, из камин, из угла, он слышал ее дыхание, ласковый шорох ее одежды. На улице он провожал взглядом женщин, искал, нет ли похожей на нее...» [27, с. 451]. Такой результат идет вразрез с традиционной точкой зрения, согласно которой особую значимость имеет известный монолог возмущенного Гурова о «куцей и бескрылой жизни». Подобного мнения придерживается и В. Б. Катаев, известный исследователь творчества А. П. Чехова: «Нет необходимости считать этот эпизод едва ли не более важным, чем вся история любви и перемены в отношениях Гурова с «низшей расой». Так считают интерпретаторы, стремящиеся обосновать важный общественный смысл рассказа «Дама с собачкой». Но в чеховском мире гораздо более глубокой и подлинной общественной значимостью, чем любая возмущенная или протестующая фраза героя, обладает неизменная устремленность автора на различение истинного и ложного, «настоящего» и «ненастоящего» в человеческих идеях и делах» [26, с. 266].

Анализ динамической модели содержания текста дает возможность наглядно доказать, что динамика проявляется только в отношениях главных героев произведения (связь через предикацию с другими персонажами статична).

Процедура по выявлению R-доминанты, а также предикатная динамика, наблюдаемая, прежде всего, внутри таких R-типов, как P_{perc} , P_{stat} , P_{mouv} , P_{act} , P_{descr} позволяют определить вектор действия, который направлен не во внешний мир главных героев: глобальные метаморфозы происходят в их внутреннем мире – и, таким образом, уяснить авторскую страте-

гию размещения информации. Эта стратегия нацелена в конечном итоге на реализацию главной задачи – отразить процесс поиска истины, анализа ложных стереотипов, переход персонажа от прежнего состояния к новому.

Предлагаемая процедура создает необходимые предпосылки для решения прикладных задач, связанных с компьютерной обработкой текста в рамках литературо- и переводоведения.

ЛИТЕРАТУРА

1. Пospelов Д. А. Моделирование рассуждений / Д. А. Пospelов. – М. : Наука, 1989. – 215 с.
2. Агеев В. Н. Семиотика / В. Н. Агеев. – М. : Весь мир, 2002. – 256 с.
3. Боумэн У. Графическое представление информации / У. Боумэн. – М. : Мир, 1971. – 228 с.
4. Валькман, Ю. Р. Когнитивные графические метафоры : когда, зачем, почему и как мы их используем / Ю. Р. Валькман // Знания-Диалог-Решение (KDS-95) : тр. междунар. конф. – Ялта, 1995. – С. 261–272.
5. Валькман Ю. Р. Видеообразы в операциях исследовательского проектирования / Ю. Р. Валькман // Искусственный интеллект-96 (КИИ-96) : тр. междунар. конф. – Казань, 1996. – С. 118–123.
6. Валькман Ю. Р. Интеллектуальные технологии исследовательского проектирования : формальные системы и семиотические модели / Ю. Р. Валькман. – Киев : Port-Royal, 1998. – 250 с.
7. Валькман Ю. Р. Анализ понятия «графический образ» / Ю. Р. Валькман, Ю. Н. Книга // Компьютерная лингвистика и интеллектуальные технологии : тр. междунар. сем. (Диалог'2002). – Протвино, 2002. – С. 41–52.
8. Зенкин А. А. Когнитивная компьютерная графика / А. А. Зенкин. – М. : Наука, 1991. – 187 с.
9. Зенкин А. А. Когнитивная компьютерная графика : некоторые вопросы методологии применения в интеллектуальных системах / А. А. Зенкин // Искусственный интеллект-94 (КИИ-94) : тр. нац. конф. с междунар. участием. – Рыбинск, 1994. – С. 100–105.
10. Степанов Ю. С. К универсальной классификации предикатов / Ю. С. Степанов // Изв. АН СССР. Сер. лит. и яз. – 1980. – Т. 39, № 4. – С. 311–323.
11. Ермаков А. Е. Автоматическое извлечение фактов из текстов досье. Опыт установления анафорических связей / А. Е. Ермаков // Компьютерная лингвистика и интеллектуальные технологии : тр. междунар. конф. (Диалог'2007). – Бекасово, 2007. – С. 172–177.
12. Кузнецов И. П. Семантические представления / И. П. Кузнецов. – М. : Наука, 1986. – 268 с.
13. Кузнецов И. П. Семантико-ориентированные системы на основе баз знаний / И. П. Кузнецов, А. Г. Мацкевич. – М. : Связьиздат, 2007. – 173 с.
14. Кузнецов И. П. Особенности извлечения знаний из текстов семантико-ориентированным лингвистиче-

ским процессором Semantix / И. П. Кузнецов, Д. А. Ефимов. – Режим доступа: <http://www.dialog-21.ru/>

15. *Москин Н. Д.* Применение нечетких теоретико-графовых моделей в задачах моделирования и поиска песенных мотивов / Н. Д. Москин // Искусственный интеллект-2010 : тр. XII нац. конф. по искусственному интеллекту с междунар. участием (КИИ-2010). – М. : Физматлит, 2010. – Т. 1. – С. 243–251.

16. *Поспелов Д. А.* Моделирование рассуждений / Д. А. Поспелов. – М. : Наука, 1989. – 215 с.

17. *Поспелов Д. А.* Серые и/или черно-белые / Д. А. Поспелов // Прикладная эргономика. Рефлективные процессы : спец. вып. – 1994. – № 1. – С. 29–43.

18. *Скороходько Э. Ф.* Семантические сети и автоматическая обработка текста / Э. Ф. Скороходько. – Киев : Наук. думка, 1983. – 217 с.

19. *Скрэг Г.* Семантические сети как модели памяти / Г. Скрэг // НЗЛ. – М. : Прогресс, 1999. – Т. III. – С. 259–302.

Воронежская государственная медицинская академия имени Н. Н. Бурденко

Ефимова Т. В., кандидат филологических наук, доцент кафедры иностранных языков

E-mail: efimova-t-v@mail.ru

Тел.: (4732) 269-77-32

20. *Барлас Л. Г.* Язык повествовательной прозы Чехова. Проблемы анализа / Л. Г. Барлас. – Ростов н/Д : Изд-во Рост. ун-та, 1991. – 208 с.

21. *Мелетинский Е. М.* О литературных архетипах / Е. М. Мелетинский. – М. : Наука, 1994. – 244 с.

22. *Чудаков А. П.* Поэтика Чехова / А. П. Чудаков. – М. : Наука, 1971. – 281 с.

23. *Чудаков А. П.* Об эволюции стиля прозы Чехова / А. П. Чудаков // Славянская филология. – М. : Изд-во МГУ, 1963. – Вып. V. – С. 5–28.

24. *Чудаков А. П.* Чехов. Единство видения / А. П. Чудаков // Слово – вещь – мир. От Пушкина до Толстого. – М. : Современ. писатель, 1992. – С. 105–131.

25. *Бердников Г. П.* Чехов (ЖЗЛ) / Г. П. Бердников. – М. : Молодая гвардия, 1974. – 512 с.

26. *Катаев В. Б.* Проза Чехова : проблемы интерпретации / В. Б. Катаев. – М. : Изд-во МГУ, 1979. – 290 с.

27. *Чехов А. П.* Рассказы и повести / А. П. Чехов. – Воронеж : Изд-во Воронеж. гос. ун-та, 1982. – 480 с.

Voronezh State Medical Academy named after N. N. Burdenko

Efimova T. V., Candidate of Philology, Associate Professor of Foreign Languages Department

E-mail: efimova-t-v@mail.ru

Tel.: (4732) 269-77-32