

# АНАЛИЗ ХУДОЖЕСТВЕННОГО ТЕКСТА С ПРИМЕНЕНИЕМ СЕМАНТИЧЕСКОЙ СЕТИ

© 2003 Т.В. Ефимова

*Воронежская государственная медицинская академия*

Наш интерес к такому формату представления знаний как семантическая сеть (далее СС) не случаен. Во-первых, основное преимущество этой модели находится в соответствии с современными представлениями об организации долговременной памяти человека, что подтверждается результатами нейропсихологических исследований [6; 5; 3; 10].

Во-вторых, данный метод представления знаний пользуется большой популярностью у исследователей в области искусственного интеллекта [19; 13; 11; 18; 7; 9], так как а) сети удобны для чтения и компьютерной обработки текста; б) их современные версии достаточно мощны, чтобы репрезентировать семантику естественного языка (подобная макроструктура является средством визуализации тематической целостности текста, то есть его глобального смыслового единства); в) специальные сетевые языки (LISP, PROLOG, NET, PROSPECTOR, TORUS, SML, FORTRAN, PASCAL, PLNLP, РЕФАЛ etc.) конкурируют с такими моделями представления знаний как фреймовые системы, логическое программирование, продукционные модели.

В-третьих, мы попытались применить сетевой метод для изучения и формализации семантического пространства художественного текста, чья длина намного превышает длину описанных в ряде работ текстов разных функциональных стилей [16; 8; 12; 22].

В своем исследовании мы в основном опирались на выводы, представленные Э.Ф. Скороходько в работе “Семантические сети и автоматическая обработка текста” [16], где такая тема как “Сетевое моделирование” впервые стала объектом монографического исследования в отечественной науке.

## Сетевое моделирование текста (СМТ)

СМТ включает три основные операции:

- 1) выявление семантических связей между элементами текста;
- 2) их представление в удобной для дальнейших исследований форме;

3) установление (с использованием этого представления) закономерностей, характеризующих изучаемый объект [16, 101].

## 1. Выявление семантических связей между элементами текста

Элементами системы семантических связей в тексте, как известно, являются три важнейшие текстовые единицы (ТЕ): слова, предложения и абзацы, – которые репрезентируют план содержания с различной степенью детализации.

Поскольку семантическая связь между единицами текста соответствует денотативному уровню отражения, в качестве денотата предложения выступают “фрагменты реальной (или художественной) действительности, отраженные в интеллекте и выраженные в тексте соответствующими языковыми средствами” [12, 46]. Таким образом, анализ семантических связей должен базироваться, прежде всего, на лингвистических категориях.

### 1.1. Семантическая связь между предложениями

Семантически связанными считаются “предложения, денотаты которых, то есть ситуации, описанные этими предложениями, связаны на предметном уровне, что позволяет рассматривать их как компоненты единой более обширной ситуации” [16, 103].

Для установления семантической связи между предложениями были использованы критерии, предложенные Э.Ф. Скороходько в монографии “Семантические сети и автоматическая обработка текста”, условно названные основным и расширенным, которые учитывают, прежде всего, номинативную связь.

В соответствии с основным критерием два предложения  $S_i$  и  $S_{i+k}$  считаются семантически связанными, если выполняется, по меньшей мере, одно из условий:

- 1) каждое из предложений  $S_i$  и  $S_{i+k}$  включают словоформу одного и того же имени существительного  $N_a$  (в нашем случае – *Гуров*, *Гурову* etc.);

- 2) предложение  $S_i$  включает некоторое имя существительное  $N_a$ , а предложение  $S_{i+k}$  – имя существительное  $N_b$ , имеющие тождественные денотаты, то есть являющиеся языковыми синонимами (*дама, женщина*);
- 3) предложение  $S_i$  включает некоторое имя существительное  $N_a$ , а предложение  $S_{i+k}$  – имя существительное  $N_b$ , причем между  $N_a$  и  $N_b$  существует парадигматическая связь гипогиперонимического типа (*собачка - шниц*);
- 4) предложение  $S_i$  включает некоторое имя существительное  $N_a$ , а предложение  $S_{i+k}$  – прилагательное или глагол с той же основой, что и  $N_a$  (*оправдание – оправдываться*);  
Согласно расширенному критерию два предложения  $S_i$  и  $S_{i+k}$  считаются семантически связанными, если выполняется, по меньшей мере, одно из условий:
  - 1) предложение  $S_i$  включает некоторое имя существительное  $N_a$ , а предложение  $S_{i+k}$  – личное или притяжательное местоимение, антецедентом которого является  $N_a$ ;
  - 2) предложение  $S_i$  включает некоторое имя существительное  $N_a$ , а предложение  $S_{i+k}$  – имя существительное  $N_b$ , причем в данном предложении  $N_a$  и  $N_b$  имеют один и тот же денотат, то есть являются контекстуальными синонимами (*муж - лакей*) [16, 118].

Со своей стороны в качестве критерия семантической связности мы приняли кореферентные наименования основных персонажей, объединив первое, второе, третье условия основного и оба условия расширенного критерия, а также четвертое условие основного критерия.

Однако максимальной силы семантическая связь между предложениями достигается в том случае, когда связанными оказываются слова, релевантные относительно заданного текста или его отрывка (имеется в виду не контекстуальная, синтагматическая, а постоянная – парадигматическая связь между значениями, например: *мол – пароход – пристань - встречающие*. Количество релевантных слов, с которым семантически связано данное слово, называется его **относительной степенью релевантности**. Множество релевантных слов текста образует фрагмент СС [16, 107].

Учет степени релевантности слов дает возможность усовершенствовать лексический критерий установления семантической связи между предложениями.

Сила семантической связи может также определяться числом совпадающих слов, их частотой (признак текстуальной значимости), тематическими характеристиками и т.д. [15].

## 1.2. Сегментирование текста

Предложение как элемент системы семантических связей репрезентирует план содержания текстов небольшой длины (20 – 30 предложений) типа сказок, притч, публицистических заметок etc., а также отрезков более длинных текстов: абзацев, параграфов etc.

Однако для описания семантического пространства рассказа А.П. Чехова «Дама с собачкой», чья длина составляет 310 предложений, подобное членение оказывается недостаточным: предложение становится слишком мелкой единицей, а пофразная СС – слишком громоздкой. Это затрудняет анализ семантической структуры текста и определение ее типа. Поэтому в нашей работе за единицу системы семантических связей мы взяли так называемое **сложное синтаксическое целое (ССЦ)** (термин по Н.С. Поспелову [14, 53]), т.е. множество предложений, теснее связанных друг с другом, чем с остальными предложениями текста, и представляющие некоторое семантическое единство.<sup>3</sup>

Сегментирование текста на ССЦ и определение силы семантической связи между ними производилось на основании тех же критериев, что были установлены для предложений, в два этапа: сначала устанавливалась семантическая связь между предложениями, входящими в эти ССЦ, затем – семантическая связь и её сила между ССЦ. Другими словами, критерий семантической связи между ССЦ – наличие семантической связи между составляющими их предложениями.

**Сила семантической связи** как между предложениями, так и между ССЦ, эквивалентна числу **индикаторов связности**.

Следует упомянуть еще об одном критерии сегментирования текста. Речь идет о типичной особенности чеховской прозы начинать или заканчивать какой-либо смысловой кусок фразы, суммирующими впечатление, например: *“Что-то в ней есть жалкое все-таки” – подумал он и стал засыпать; Дети ему надоели, банк надоел, не хотелось никуда идти, ни о чем говорить* [21, 124].

В результате сегментирования текста по вышеперечисленным критериям выявлено **33 ССЦ**.

**1 ССЦ** включает предложения **1–6**, семантически связанные между собой посредством **лек-**

<sup>3</sup> В своей работе мы используем термин ССЦ вместо термина сгусток, употребляемого автором монографии: “Сгусток – множество предложений, тесно связанных по смыслу и образующих некое семантическое единство” [16, 144].

сического повтора (дама с собачкой 2, набережная 2, новое лицо 2); кореферентных наименований персонажей и их словоформ (Дмитрий Дмитриевич Гуров 1 - Гуров 1 он 2, молодая дама 1 - невысокого роста блондинка 1 в берете 2 - она 7, собачка 2 - белый шниц 2); совпадения основы (не узнал - без знакомых - познакомиться - 3).

Общее количество индикаторов связности (далее ИС) составляет 28.

Во 2 ССЦ входят предложения с 17 по 11, критерием семантической связи которых являются кореферентные наименования персонажей и их словоформы (Гурова: он 8 - Дмитрий 1 - Димитрий 1; жены Гурова: жена 1 - женщина 1 - она 5); лексический повтор (изменять 2); совпадения основы (женили - жена - женщина - 3). Релевантные слова (женили [его] рано, считал недалекой, боялся, не любил, изменял) служат для описания физического и морального портрета жены Гурова, а также отношения Гурова к ней.

Таким образом, общее количество ИС - 23.

1 и 2 ССЦ связаны между собой по главному действующему лицу - Гурову.

3 ССЦ объединяет с 12 по 14 предложения, критерием связности которых являются кореферентные наименования персонажей и их словоформы (Гурова: он 9, себя 2; женщин: женщины 2 - "низшая раса" 1 - они 6).

Релевантное слово - женщины и семантически связанные с ним глаголы: называть - не мог прожить (без) - знал, о чем говорить и как держать себя (с ними) - чувствовал себя свободно (с) - молчать (с ними) было легко - располагало к нему - манило - влекло.

Общее количество ИС - 20.

Связь между 2 ССЦ и 3 ССЦ осуществляется посредством общего элемента - "низшая раса", кореферентного наименования существительного женщины, а также - по главному действующему лицу - Гурову; связь с 1 ССЦ - через кореферентные наименования Гурова.

4 ССЦ включает 15 и 16 предложения, связанные между собой через лексический повтор существительного опыт (3), контекстуальные синонимы (сближение - милое, легкое приключение - целая задача - тягостное положение (4)).

Релевантные слова (учил - ускользал) сочетаются с ключевым словом опыт.

Общее количество ИС - 8.

Критерий связи 3 ССЦ и 4 ССЦ - общий элемент горький опыт, - и связь по действующему лицу - Гурову; связь 2 ССЦ осуществляется посредством кореферентных наименований главного действующего лица Гурова.

В 5 ССЦ входят предложения 17 - 35. Критерий семантической связи между ними - кореферентные наименования персонажей и их словоформы (Гуров 1 - он 12; дама в берете 1 - неизвестная женщина 1 - дама 2 - она 5; шниц 2 - тот 1 - он 3); совпадения основы и лексический повтор (рассказы 3 - сказала 2; погрозили 2; скука 1 - скучно 3).

Релевантное слово - рассказы и семантически связанные с ним глаголы: знал - презирал - сочиняются - вспомнились.

Общее количество ИС - 42.

5 ССЦ и 4 ССЦ соединяются между собой посредством кореферентных наименований общего элемента: сближение - мимолетная встреча - роман.

Связь с 1 ССЦ осуществляется через общий элемент в саду, а также через действующих лиц - Гурова и Анну Сергеевну.

6 ССЦ объединяет с 36 по 42 предложения, семантически связанные между собой посредством лексического повтора и совпадения основы (говорить 4 - разговор 1; узнал 2; замуж 1 - её муж 3); употребления синонимов (идти 2 - гуляли; рассказал 1 - объяснил 1); кореферентных наименований персонажей и их словоформ (АС 1 - она 7; Гуров 2 - он 2).

Релевантные слова данного ССЦ - говорить, рассказывать, узнавать, - служат для описания сцены знакомства главных героев.

Общее количество ИС - 29.

Связь с 5 ССЦ осуществляется посредством общих элементов - совпадения основы (обедал - обед), лексического повтора (замужем), а также по главным действующим лицам - Гурову и Анне Сергеевне.

В 7 ССЦ включены предложения 43 - 47, критерием связности которых являются кореферентные наименования персонажей и их словоформы (Гурова: он 5 - себя 1; Анны Сергеевны: она 9 - институтка ("еще так недавно была институткой") 1); синонимы (думал 2 - вспомнил 3).

Релевантные слова - думал, вспомнил.

Общее количество ИС - 22.

Связь с 6 ССЦ - через кореферентные наименования действующих лиц (Гурова и АС).

В 8 ССЦ входят предложения 48 - 52, связанные между собой на парадигматическом уровне (душно - хотелось пить - предлагал то воды с сиропом, то мороженое - некуда было деваться; неделя - день), а также посредством лексического повтора существительного день (2).

Общее количество ИС - 4.

Связь с 7 ССЦ и 6 ССЦ осуществляется посредством общего элемента - совпадения осно-

вы (знакомство – незнакомый – незнакомые), а также посредством главных действующих лиц – Гурова и Анну Сергеевну.

**9 ССЦ** включает предложения с **53** по **67**. Критерий семантической связи между ними – кореферентные наименования персонажей и их словоформы (Гуров 3 – он 5; АС 3 – она 8; они 1 – мы 1 – оба 1); контекстуальные синонимы (гуляющие – толпа – дамы и генералы); лексический повтор (вечером 2 – толпа 2 – мол 2 – пароход 2); совпадение основы (пристань – пристать; стих – утихло; гулять – гуляющие); наличие релевантных слов (мол – пароход – пристань – гуляющие – пассажиры – встречать – смотреть – гулять).

#### **Общее количество ИС – 34.**

Связь **9 ССЦ** и **8 ССЦ** осуществляется посредством общего элемента – языковых синонимов: *вихрь – ветер*. Учтена также связь по главным персонажам (Гурову и АС).

**10 ССЦ** объединяет с **68** по **103** предложения, критерием связности которых являются кореферентные наименования персонажей и их словоформы (Гуров 5 – он 4 – я 2 – вы 3 – себя 1; АС 2 – эта дама с собачкой 1 – она 17 – я 19 – ты 4 – себя 2; муж 3 – он 4 – [может быть] честный, хороший человек 1 – лакей 2; женщины 2 – они 2); совпадение основы (оправдываться 2 – оправдание 1; падение – опустили; жизнь – жить – пожить б); языковые синонимы (не уважать – презирать); лексический повтор (молодость 2 – обманула 2 – номер 2 – пожить 2 – попутал [нечистый] 2).

В целях более адекватного отражения семантической связи было допущено следующее отклонение от указанных критериев: восстановлен эллипс в предложении **77**.

Релевантные слова (не уважаете – презираете – оправдываться – обманула – хотелось пожить – покаяние) служат для описания душевного состояния Анны Сергеевны.

#### **Общее количество ИС – 90.**

Связь **10 ССЦ** и **9 ССЦ** осуществляется по главным действующим лицам, по общему элементу – синонимам: [пойдемте] к вам – у нее в номер; а также через лексический повтор местоимения оба.

Связь с **6 ССЦ** – через кореферентные наименования мужа АС.

Критерий связи **10 ССЦ**, со **2 ССЦ** и **3 ССЦ** – лексический повтор существительного женщины.

В **11 ССЦ** входят предложения **104 - 120**. Критерий семантической связи между ними – кореферентные наименования персонажей и их словоформы (Гурова: я 1- Гуров 2; мужа Анны Сергеевны: фон Дидериц 1 – муж 1- он 2 – не-

мец 1 – православный 1; Анны Сергеевны: фон Дидериц 1 – красивая молодая женщина 1 – АС 1; Гурова и АС: они 4); совпадение основы и лексический повтор (Ореанда 3 – море 3; шум – шуметь – шумит 4; горы 2 – облака 2; сидели – сидя; молчали – молчание; покой – успокоенный); контекстуальные синонимы (Ялта – домой – в город); языковые синонимы (прекрасный – красивый – сказочный – 4).

Релевантные слова – покой, сон, горы, море, облака, жизнь, смерть, спасение, цель, бытие, достоинство, таинственный, красивый, – описывают природу Ореанды, а также душевное состояние Гурова.

#### **Общее количество ИС – 53.**

Связь с **10 ССЦ** осуществляется посредством общего элемента – лексического повтора существительного муж, а также через кореферентные наименования действующих лиц – Гурова и Анны Сергеевны (оба – они); связь с **6 ССЦ** – через лексический повтор существительного муж.

В целях более адекватного отражения семантической связи было допущено следующее отклонение от указанных критериев: учтена причинно-следственная связь предложений **117, 118** и **119**: *утренняя заря → роса на траве → пора домой*.

В **12 ССЦ** включены предложения **121 - 125**, критерием связности которых являются кореферентные наименования персонажей и их словоформы (Гурова – он 5; АС – она 9; Гурова и АС – они 2); лексический повтор (море 2; не уважаете); совпадение основы (гуляли – прогулка; страстно – страстен).

Релевантные слова – встречались, гуляли, восхищались, целовал, был страстен, впечатления [были величавы и прекрасны], – служат для описания бурно развивающегося романа Гурова и Анны Сергеевны.

#### **Общее количество ИС – 27.**

Связь с **11 ССЦ** осуществляется через повтор существительного Ореанда, а с **10 ССЦ** – посредством лексического повтора глагола не уважаете, а также кореферентных наименований главных действующих лиц.

**13 ССЦ** объединяет с **126** по **129** предложения, связанные между собой посредством кореферентных наименований персонажей и их словоформ (муж 4; Гуров 1; АС 1 – она 2); языковых синонимов (уезжать – вернуться домой).

В целях более адекватного отражения семантической связи **127** и **128** предложений было допущено следующее отклонение от указанных критериев: восстановлен эллипс в предложении **128** – *Анна Сергеевна заторопилась (домой)*.

#### **Общее количество ИС – 11.**

Связь с **12 ССЦ** осуществляется по главным действующим лицам.

Критерий связи с **6 ССЦ**, **10 ССЦ** и **11 ССЦ** – общий элемент: существительное муж.

В **14 ССЦ** входят предложения **130 – 146**. Критерий семантической связи между ними – коррелятивные наименования персонажей и их словоформы (Гуров 2 – вы 4 – он 7; АС: эта молодая женщина 1 – она 19 – я 2); совпадения основы (поехал 1 – поехала 1 – поезд 2); контекстуальные синонимы (на север – домой); лексический повтор (пора 2); языковые синонимы (думать 2 – вспоминать 1 – поминать 1).

Релевантная лексика – уезжать, провожать, вагон, поезд, звонок, платформа, станция, прощаться, не следовало бы встречаться, он с ней никогда не увидится, – служит для описания сцены прощания Анны Сергеевны с Гуровым.

**Общее количество ИС – 61.**

Связь с **13 ССЦ** осуществляется по главным действующим лицам, а также через контекстуальные синонимы: вернуться домой – пора на север.

Связь с **10 ССЦ** осуществляется посредством общего элемента – лексический повтор существительного воспоминание, а также через синонимы: похождение, приключение, встреча.

Критерий связи с **4 ССЦ** – лексический повтор существительного приключение.

**15 ССЦ** включает предложения с **147 по 150**, критерием семантической связи между которыми являются лексический повтор (белый 3); парадигматическая связь между словами (по зимнему – морозы – топить [печи] – снег – иней).

**Общее количество ИС – 8.**

Связь с **14 ССЦ** осуществляется посредством общих элементов – контекстуальных синонимов: дом – север – Москва; с **12 ССЦ** – связь по главным персонажам (Гурову и АС); с **11 ССЦ** – через лексический повтор существительных море и горы.

В **16 ССЦ** включены предложения **151 – 154**, критерием связности которых являются коррелятивные наименования Гурова (Гуров 1 – он 9); совпадение основы (москвич – Москва – московские [жизнь, газеты]); релевантная лексика (рестораны – клубы – званые обеды – юбилей – съесть [целую порцию солянки]).

**Общее количество ИС – 19.**

Связь с **15 ССЦ** осуществляется через повтор существительного Москва, посредством совпадения основы (морозы – морозный), с **6 ССЦ** – посредством общего элемента – существительного москвич.

Учтена также связь по главному персонажу – Гурову.

**17 ССЦ** объединяет с **155 по 163** предложения, критерием связности которых являются коррелятивные наименования персонажей и их словоформы (Гурова – он 12; АС 3 – она 4); совпадение основы и лексический повтор (в памяти 3 – воспоминания 2 – вспоминал 1); контекстуальные синонимы (вспоминал – видел – воскресало в памяти).

Релевантная лексика (воспоминания разгорались – вспоминать – (АС) шла за ним (Г) повсюду – следила – казалась красивее – глядела на него – слышал ее дыхание) отражает воспоминания Гурова о курортном романе.

**Общее количество ИС – 21.**

Связь с **16 ССЦ** – посредством коррелятивных наименований Гурова.

Связь **17 ССЦ** и **15 ССЦ** осуществляется по главным действующим лицам, а также по общим элементам – существительному дети, а также посредством совпадения основы: зима – по-зимнему; с **14 ССЦ** – через коррелятивные наименования Анны Сергеевны.

Критерий связи с **11 ССЦ** – лексический повтор существительного горы и совпадение основы: раннее утро – утренняя заря; с **12 ССЦ** – лексический повтор существительного поцелуй.

Связь с **9 ССЦ** – через лексический повтор существительных мол и паролод.

В **18 ССЦ** входят предложения **164 – 171**. Критерий семантической связи между ними – коррелятивные наименования Гурова и их словоформы (Дмитрий 1 – ты 1 – он 1); совпадение основы (о любви 2 – любил 1), синонимы (поделиться – говорить); лексический повтор (говорить 3).

В целях более адекватного отражения семантической связи **166** и **165** предложений было допущено следующее отклонение от указанных критериев: восстановлен эллипс в предложении **166** – Но дома нельзя было говорить о своей любви, а вне дома – не с кем. – Не с жильцами же (говорить) и не в банке.

Релевантная лексика (сильное желание поделиться, говорить о своей любви) служит для описания душевного состояния Гурова.

**Общее количество ИС – 20.**

Связь с **17 ССЦ** осуществляется по главным действующим лицам – Гурову и Анне Сергеевне, а также – посредством лексического повтора существительного воспоминание.

Критерий связи с **2 ССЦ** – лексический повтор существительного жена.

**19 ССЦ** включает предложения с **172 по 175**, связанные между собой через коррелятивные наименования персонажей и их словоформы (Гурова: он 1 – Дмитрий Дмитрич 1 – вы 1; чиновника-партнера: партнер 1 – чиновник 2 – вы

1); восстановленный эллипс предложения 174.

**Общее количество ИС – 7.**

Связь **19 ССЦ** и **16 ССЦ** осуществляется посредством общего элемента - *докторский клуб*.

Критерий связи с **18 ССЦ** - причинно-следственная связь: *томило сильное желание поделиться с кем-нибудь → он не удержался и сказал (своему партнеру, чиновнику), а также коррелятивные наименования: Анна Сергеевна – очаровательная женщина.*

В **20 ССЦ** входят предложения **176 – 183**, критерием семантической связности которых являются коррелятивные наименования главного действующего лица и их словоформы (*Гурова: он 3*); лексический повтор (*Гуров 2; возмутить 2; ночь 3; всё об одном 2*); гипогиперонимическая связь существительных (*дни и ночи – жизнь*); совпадение основы (*разговоры – говорить; уйти – идти*); контекстуальные синонимы (*обжорство – пьянство – жизнь – чепуха – сумасшедший дом – арестантские роты*).

Релевантные слова (*дикие нравы, незаметные дни, ненужные дела, бескрылая жизнь, сидеть в сумасшедшем доме или в арестантских ротах*) отражают кризис сознания Гурова.

**Общее количество ИС – 35.**

Критерий связи с **19 ССЦ** – восстановленный антецедент словосочетания *эти слова – А давеча вы были правы: осетрина-то с душком!*

Связь **20 ССЦ** и **18 ССЦ** осуществляется посредством общего элемента – существительного *банк*.

**20 ССЦ** связан с **17** и **15 ССЦ** через общий элемент – существительное *дети*.

**21 ССЦ** объединяет с **184** по **186** предложения. Критерий семантической связи между ними - коррелятивные наименования Гурова их словоформы (*он 3*); синонимы (*повидаться – устроить свидание – поговорить; собраться в дорогу – уезжать*); восстановленный эллипс предложения 185.

**Общее количество ИС – 9.**

Критерий связи с **18 ССЦ** – повтор существительного *жена*.

Связь с **19 ССЦ** – через коррелятивное наименование *Анны Сергеевны – очаровательная женщина*.

Связь с **20 ССЦ** - по главному персонажу - Гурову.

В **22 ССЦ** включены предложения **187 - 189**, критерием связности которых являются коррелятивные наименования Гурова (*он 2*); совпадение основы (*швейцар 2*); релевантная лексика (*гостиница – занять – номер – швейцар*).

**Общее количество ИС – 8.**

Связь с **21 ССЦ** осуществляется через повтор существительного (*город*) *С*.

Учтена также связь по главному персонажу - Гурову.

В **23 ССЦ** входят предложения с **190** по **202**, связанные между собой через лексический повтор и коррелятивные наименования персонажей и их словоформы (*Гуров 3 - он 7; Анна Сергеевна 2; шпиц 2 – собака 1*), повтор релевантных существительных (*дом 3; забор 6*); совпадение основы (*игра – играла*).

**Общее количество ИС – 26.**

Критерий связи с **22 ССЦ** – повтор существительных *Старо-Гончарная (улица)* и *дом*; с **1 ССЦ** и **5 ССЦ** – коррелятивные наименования шпица (*собака – знакомый белый шпиц*).

**24 ССЦ** объединяет предложения с **203** по **211**, критерием связности которых являются коррелятивные наименования персонажей и их словоформы (*Гурова – он 6; Анны Сергеевны – дама с собачкой 1 – она 1*); совпадение основы и лексический повтор (*вот 2; сидел 2 – сиди; спал – проснувшись – выспался*); также учтена гипогиперонимическая связь существительных (*театр – представления*).

**Общее количество ИС – 18.**

Связь с **23 ССЦ** осуществлена посредством повтора существительного *номер*.

**25 ССЦ** включает предложения с **212** по **258**. Критерий семантической связи между ними - коррелятивные наименования персонажей и их словоформы (*Гуров 5 – Дмитрий Дмитрич 1 – милый, добрый, дорогой 1 – он 13 – себе 2 – вы 12; Анна Сергеевна 5 – дорогой, близкий человек 1 – маленькая женщина с вульгарною лорнеткою в руках 1 – его радость, горе, единственное счастье 1 – она 19 – вы 1; муж 2 – молодой человек с бакенами 1 – лакей 1 – он 3*); лексический повтор и совпадение основы (*театр 3; оркестр 3; входила – вошла; лестница 3; значок 2; поймите 3; лакей – лакейски-скромное - лакейский; сердце 2; сел 2 – сидевший – сидела; уезжайте 2 – уехать 1 – приеду 2 – приехали 1; делаете 2; входит 3; несчастливая 2 – несчастна 1; страдаю – страдать; (искал) глазами – глядела – взглянул; расстанемся 2*); контекстуальные синонимы (*публика – провинциальная толпа*); наличие релевантных слов (*театр – галерка – оркестр – настраиваться – люстры – ложа – занавес – места – занимать – публика – скрипки*).

**Общее количество ИС – 125.**

Связь **25 ССЦ** и **24 ССЦ** осуществляется посредством общего элемента – существительного *театр*.

Критерий связи с **6 ССЦ** и **10 ССЦ** – наличие общего элемента: существительного *муж*; с **14**

**ССЦ** – лексический повтор (*на станции*).

Учтена также связь по главным персонажам (Гурову и АС).

В **26 ССЦ** входят предложения **259 – 262**, критерием семантической связи которых являются корреферентные наименования действующих лиц и их словоформы (*Гурова – он 1; АС – она 3*); лексический повтор (*Гуров 2; муж 2; Москва 2*); совпадение основы (*приезжать – уезжала – едет – приехав*).

**Общее количество ИС – 14.**

Критерий связи с **25 ССЦ** – наличие общего элемента: словосочетания *приехать в Москву, а также корреферентные наименования Гурова* (*Гуров – к нему*).

**27 ССЦ** объединяет с **263 по 269** предложения, критерием связности которых являются корреферентные наименования персонажей и их словоформы (*Гуров 1 – папа 1 – он 5; Анна Сергеевна: она 1; его дочь 2*), лексический повтор (*шел 2 – снег 2 – тепло 2*); совпадение основы (*зимнее [утро] – зима*); синонимы (*говорил – объяснил*); релевантная лексика (*дочь – проводить – гимназия – дорога – тепло – мокрый снег*).

**Общее количество ИС – 29.**

Связь с **26 ССЦ** осуществляется посредством общих элементов – корреферентных наименований персонажа (*человек в красной шапке – посылный*).

Учтена также связь по главным персонажам (Гурову и АС).

**28 ССЦ** включает предложения с **270 по 274**. Критерий семантической связи между ними – корреферентные наименования Гурова и их словоформы (*он 7 – себя 1*); лексический повтор и совпадение основы (*тайна 3 – тайно 2; жизнь 4 – знали 3*); контекстуальные синонимы (*жизнь – оболочка*).

**Общее количество ИС – 20.**

Связь с **27 ССЦ** осуществляется посредством общего элемента – *говорил*, с **26 ССЦ** – через повтор отрицательной формы глагола *знать* (*никто не знал*); с **16 ССЦ** – посредством повтора существительных *клуб* и *юбилей*.

Критерий связи с **6 ССЦ** и **20 ССЦ** – повтор существительного *банк*, с **2 ССЦ** и **21 ССЦ** – повтор существительного *жена*, а с **18 ССЦ** – повтор существительных *банк* и *жена*.

Связь с **3 ССЦ** осуществляется через общий элемент – *низшая раса*.

Учтена также связь по главному персонажу – Гурову.

В **29 ССЦ** входят предложения **275 – 290**, критерием семантической связи которых являются лексический повтор и корреферентные наименования действующих лиц и их словоформы

(*Гуров 1 – он 12; Анна Сергеевна 2 – она 7*); лексический повтор и совпадение основы (*плакала 2 – поплачет 1; кончится – конец*); языковые синонимы (*сказать – говорить*); релевантные слова (*любовь – обожала – привязывалась*); восстановленный эллипс предложения **279** – *Ну, (расскажи), как живешь там? – погоди, сейчас скажу...*

**Общее количество ИС – 32.**

Критерий связи с **25 ССЦ** – совпадение основы (*кончилось – до конца – иметь какой-нибудь конец*); с **26 ССЦ** – наличие общего элемента *Славянский базар*.

Связь с **27 ССЦ** осуществляется посредством лексического повтора (*дочь, проводить в гимназию*), а с **28 ССЦ** – через повтор наречия *тайно* и существительного *жизнь*.

Учтена также связь по главным персонажам (Гурову и Анне Сергеевне).

**30 ССЦ** объединяет с **291 по 296** предложения, семантически связанные между собой посредством лексического повтора и корреферентных наименований персонажей и их словоформ (*Гурова; он 5 – себя 1; Анны Сергеевны: она 3*); контекстуальных синонимов (*Анна Сергеевна – эта жизнь*); лексического повтора (*плечи 2*); языковых синонимов (*постарел – сесть – блекнуть – вянуть – подурнел*).

В целях более адекватного отражения семантической связи **291** и **292** предложений было допущено следующее отклонение от указанных критериев: восстановлен эллипс в предложении **292** – [...] *увидел себя в зеркале. – (Он увидел, что) голова его уже начала седеть.*

**Общее количество ИС – 17.**

Связь с **29 ССЦ** осуществляется посредством общих элементов – корреферентных наименований персонажей Гурова и АС, а также – посредством синонимов *обожала – любит*.

В **31 ССЦ** включены предложения **297 – 299**, критерием связности которых являются корреферентные наименования Гурова (*человек 1 – он 5*), *женщин* (*они 3*), а также совпадение основы (*любили – любовь – (не) любил – 4*).

**Общее количество ИС – 13.**

Связь с **30 ССЦ** осуществляется через повтор глагола *любить*; посредством общего элемента (повтора и корреферентных наименований существительного *женщины*) – связь с **28 ССЦ**, **10 ССЦ** и **3 ССЦ**.

Учтена также связь по главному персонажу – Гурову.

В **32 ССЦ** входят предложения **300 – 303**. Критерий семантической связи между ними – лексический повтор и корреферентные наименования действующих лиц и их словоформы (*Гурова: – он 6; Анны Сергеевны 1 – она 1; они 3 –*



друг друга 3 – как близкие и родные люди 1 – как муж и жена 1 – нежные друзья 1 – две перелетные птицы 1 – самец и самка – оба 1); лексический повтор и совпадение основы (полюбил – любили – любовь; простили – прощали).

**Общее количество ИС – 24.**

Связь с **30 ССЦ** осуществляется посредством общего элемента – голова начинала сидеть – голова стала седой, а с **29 ССЦ** и **31 ССЦ** – через глагол любить и его словоформы, а также посредством совпадения основы: *любовь – любить*.

Учтена также связь по главным персонажам – Гурову и Анне Сергеевне.

33 ССЦ объединяет с 304 по 310 предложения, семантически связанные между собой посредством лексического повтора и кореферентных наименований персонажей и их словоформ (Гурова; он 2 – себя 1; Анны Сергеевны: моя хорошая 1; они 1 – себя 1); лексического повтора (как 5; начинается 2); синонимов (говорили – советовались – спрашивал; избавить себя – освободиться); совпадения основы (поговорим – говорил – говорили).

Критерий связи с 25 ССЦ – лексический повтор (до конца), с 32 ССЦ – кореферентные наименования Гурова (ему – он).

Связь с 29 ССЦ осуществляется посредством общего элемента – лексического повтора (перестань, конец); глагола плакать и его словоформ, а также языковых синонимов (прятаться – скрываться).

Общее количество ИС – 24.

Учтена также связь по главным персонажам – Гурову и Анне Сергеевне.

## 2. Представление системы семантических связей в тексте

Идеальной математической моделью такого системно-структурного образования как СС 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)» [17, 262].

Семантическая сеть, которая с некоторыми упрощениями передает содержание текста на естественном языке, может рассматриваться как его нелинейный аналог. Являясь плоским графом, то есть двумерной фигурой, она не имеет ограничений на число связей, в которые вступает элемент этой фигуры – вершина.

Вершины нашего графа – это 33 ССЦ, полученные в результате ранее проведенного сегментирования текста, а ребра, соединяющие соответствующие вершины, маркируют семантическую связь между ССЦ, установленную по вышеперечисленным критериям. (См. граф Семантическая сеть рассказа “Дама с собачкой”)

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

## 3. Установление (с использованием этого представления) закономерностей, характеризующих изучаемый объект

### 3.1. Семантическая связность текста

Являясь удобным инструментом исследования системных и структурных свойств текста, сетевой метод представления знаний позволяет определить ряд количественных параметров, характеризующих систему семантических связей в нем.

Один из основных параметров – семантическая связность (С), характеризующая план содержания текста, его семантическую монолитность. Она определяется по формуле:

$$C = \frac{2n}{m(m-1)},$$

где n – число пар ССЦ, семантически связанных между собой; m – общая длина (число ССЦ) текста [16, 122].

Определим семантическую связность анализируемого нами текста. Его длина (на графе – число вершин) составляет 33 ССЦ. Число пар семантически связанных друг с другом ССЦ (на графе – число ребер) равно 83. Следовательно, величина семантической связности текста при вышеперечисленных критериях составляет 0,115.

### 3.2. Цепочечный коэффициент

Еще одним важным параметром плана содержания является так называемый цепочечный коэффициент (Z). Он характеризует среднюю длину цепочки ССЦ, в которой каждое предшествующее ССЦ непосредственно связано с последующим за ним ССЦ.

Цепочечный коэффициент определяется по формуле:

$$Z = \frac{\sum_{i=1}^S Z_i}{S},$$



где  $Z_i$  – длина  $i$ -й цепочки, то есть количество ССЦ, семантически связанных между собой и непосредственно примыкающих друг к другу;  $S$  – количество цепочек в тексте [16, 127].

На графе Дамы имеются девять цепочек, включающие ССЦ 1–4, 5–7, 8–9, 10–12, 13–14, 15–17, 18–22, 23–28, 29–33. Следовательно, цепочечный коэффициент равен 3,7.

Следует отметить, что этот параметр, как и семантическая связность, напрямую зависит от функционального стиля: для специального научного текста он равен приблизительно 7,4; для научно-популярного – 6,6; для художественного и публицистического – 5,3 [4, 1971].

### 3.3. Функциональный вес ССЦ

Функциональный вес ССЦ ( $\varphi$ )<sup>4</sup> характеризует локальную и общетекстовую значимость ССЦ. Этот параметр определяется числом семантических связей, в которые вступает в данном тексте рассматриваемая единица.

Функциональный вес ССЦ определяется по формуле:

$$\varphi = N (m - m_{\max}),$$

где  $N$  – число текстуальных семантических связей, в которые вступает ССЦ (на текстовой сети – число ребер, инцидентных вершине;  $m$  – число ССЦ всего текста (на сети – количество вершин);  $m_{\max}$  – число ССЦ в самом длинном связном отрезке, который образовался после удаления рассматриваемого ССЦ (на сети – число вершин в максимальном по величине связном элементе графа после удаления рассматриваемой вершины) [16, 128].

В этой формуле составляющая  $N$  – локальная,  $(m - m_{\max})$  – общетекстовая значимость.

В анализируемом тексте были выделены следующие ССЦ с  $N_{\max}$ : у 15 ССЦ, 18 ССЦ и 20 ССЦ  $N=6$ ; у 2 ССЦ, 3 ССЦ, 6 ССЦ, 25 ССЦ и 29 ССЦ  $N=7$ ; у 17 ССЦ  $N=8$ ; у 10 ССЦ и 28 ССЦ  $N=10$  (на графе – это вершины более крупного размера).

Общетекстовая значимость составляет для 2 ССЦ –  $(33-31)=2$ ; для 3 ССЦ –  $(33-30)=3$ ; для 6 ССЦ –  $(33-27)=6$ ; для 10 ССЦ –  $(33-23)=10$ ; для 15 ССЦ –  $(33-18)=15$ ; для 17 ССЦ –  $(33-16)=17$ ; для 18 ССЦ –  $(33-17)=16$ ; для 20 ССЦ –  $(33-$

19)=14; для 25 ССЦ –  $(33-24)=9$ ; для 28 ССЦ –  $(33-27)=6$ ; для 29 ССЦ –  $(33-28)=5$ .

Следовательно, функциональный вес для 2 ССЦ равен 12 ( $\varphi_2=14$ ); для 3 ССЦ –  $\varphi_3=21$ ; для 6 ССЦ –  $\varphi_6=42$ ; для 10 ССЦ –  $\varphi_{10}=100$ ; для 15 ССЦ –  $\varphi_{15}=90$ ; для 17 ССЦ –  $\varphi_{17}=136$ ; для 18 ССЦ –  $\varphi_{18}=96$ ; для 20 ССЦ –  $\varphi_{20}=84$ ; для 25 ССЦ –  $\varphi_{25}=63$ ; для 28 ССЦ –  $\varphi_{28}=60$ ; для 29 ССЦ –  $\varphi_{29}=35$ .

Таким образом,  $\varphi_{\max}$  имеет 17 ССЦ ( $\varphi_{17}=136$ ): “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 растаял 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 камина, из угла, он слышал ее дыхание, ласковый шорох ее одежды. На улице он провожал взглядом женщин, искал, нет ли похожей на нее...” [27, 5].

Понятие функционального веса и локальной значимости может использоваться в качестве одного из основных критериев при автоматической обработке текстов – реферировании и индексировании.

### 3.4. Семантическая структура текста

Сегментирование текста по указанным выше критериям и форма построенной нами СС “Дамы...” позволили определить тип семантической структуры текста по семантической связи между ССЦ<sup>5</sup> – это кусочно-нелинейная, т. е. включающая несколько фрагментов: 1-5, 18-21, 26-29, 29-31; незамкнутая, т. е. не имеющая связи между первым и последним ССЦ; ветвящаяся структура. Как отмечает автор, это объясняется

<sup>4</sup> В своей монографии Э.Ф. Скороходько определяет семантическую связность текста, цепочечный коэффициент и функциональный вес, взяв за единицу текста предложение. Мы же в своем исследовании эти параметры определяем для ССЦ.

<sup>5</sup> О типах семантических структур тестов см. Скороходько 1983, 138 – 140.

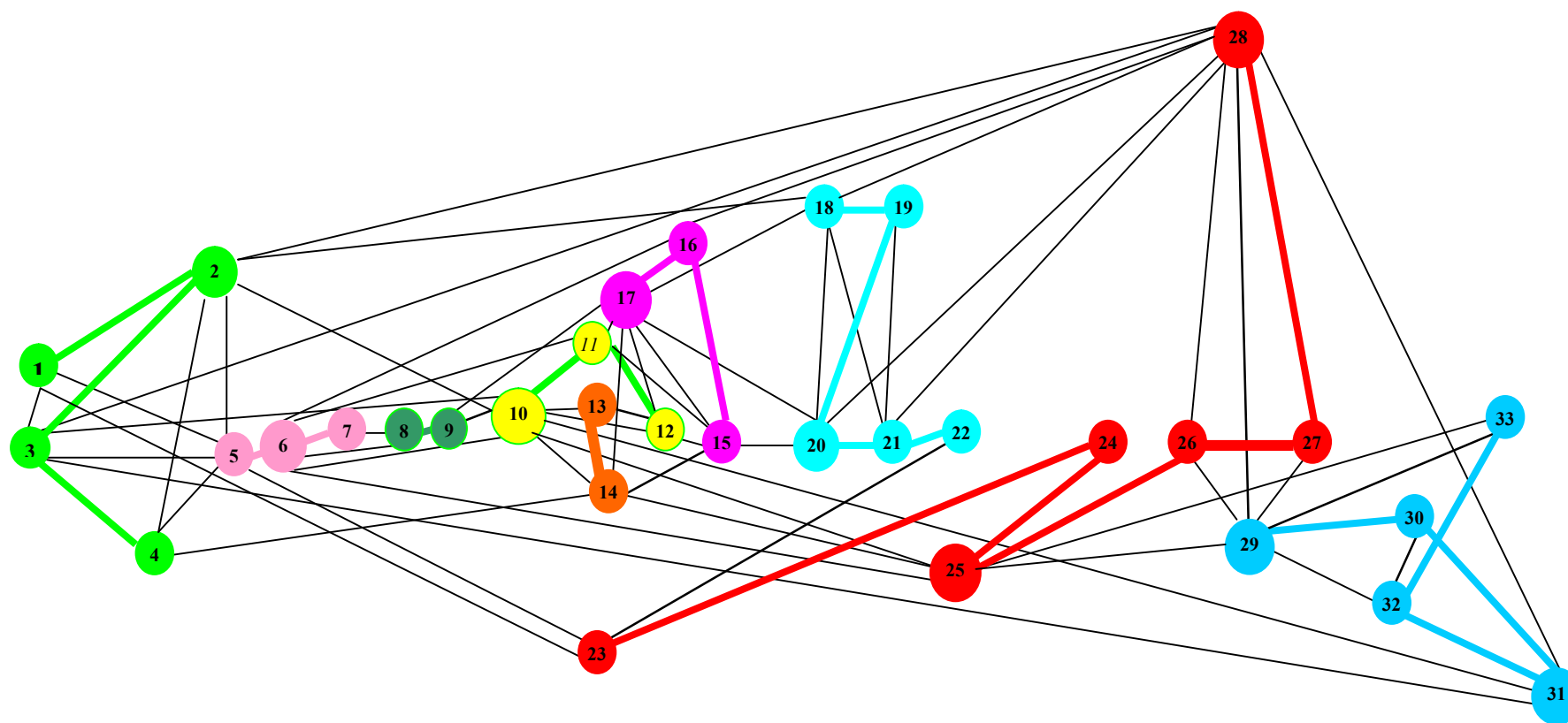
многоплановостью художественного произведения, в котором сюжет развивается в нескольких направлениях [16, 140].

### Заключение

Подобная модель плана содержания текста, являющаяся аппроксимацией изучаемого объекта [1], позволяет решать ряд теоретических и прикладных задач, особенно связанных с проектированием лингвистического обеспечения систем искусственного интеллекта: типологическое исследование текстов; анализ архитектоники литературного произведения; создание языков описания документов; информационный поиск, автоматическое реферирование, индексирование и экстрагирование: “Хотя сети (или графы) давно применялись в лингвистике для изображения тех или иных системных связей между языковыми единицами, в настоящее время семантическая сеть как метод лингвистического исследования, сохраняя иллюстративную функцию, превратилось в мощное средство изучения языка и речи, в инструмент познания и творчества” [16, 3].

По словам Ю.Д. Апресяна, модель призвана не только эксплицитно отображать моделируемый объект, но и объяснять факты, предсказывать неизвестное ранее поведение объекта, а также обеспечивать возможность количественной оценки тех или иных его аспектов [2].

Граф 4. Семантическая сеть рассказа А.П. Чехова «Дама с собачкой»



## ЛИТЕРАТУРА

1. Андреев Н.Д. Возможный путь моделирования семантики языка/Н.Д. Андреев. - М.: ВИНТИ, 1961. - 26 с.
2. Апресян Ю.Д. Современные методы изучения значений и некоторые проблемы структурной лингвистики/Ю.Д. Апресян //Проблемы структурной лингвистики. - М.: Изд-во АН СССР, 1963. - С. 102-150.
3. Ахутина Т.В. Порождение речи. Нейролингвистический анализ порождения синтаксиса/ Т.В. Ахутина. - М.: МГУ, 1989. - 215 с.
4. Белза М.И. К вопросу о некоторых особенностях семантической структуры связных текстов/ М.И. Белза //Семантические проблемы автоматизации информационного поиска. - Киев.: Наук. думка, 1971. - С.58-73.
5. Брагина Н.Н. Доброхотова Т.А. Функциональные асимметрии человека/ Н.Н. Брагина, Т.А. Доброхотова. - М.: Медицина, 1981- 287 с.
6. Глазerman Т.Б. Психофизические основы нарушений мышления при афазии/ Т.Б. Глазerman. - М.: Наука, 1986 - 230 с.
7. Ермаков А.Е., Плешко В.В. Ассоциативная модель порождения текста в задаче классификации / А.Е. Ермаков, В.В. Плешко // Информ. Технологии. - 2000. - №12.
8. Кузнецов И.П. Семантические представления/ И.П. Кузнецов. - М.: Наука, 1986. - 268 с.
9. Куо К.М. Макдональд Дж.Э. Формальная методология приобретения и представления знаний/ К.М. Куо, Дж.Э. Макдональд // ТИИЭР. - т.74. - №10. - М.: Октябрь, 1986. - С.145-155.
10. Лурия А.Р. - Основы нейропсихологии/ А.Р. Лурия. - М.: МГУ, 1973. - 374 с.
11. Мельчук И.А. Опыт теории лингвистических моделей "Смысл - Текст". Семантика, синтаксис/ И.А. Мельчук. - М.: Школа "Языки русской культуры", 1999.- 350 с.
12. Новиков А.И. Семантика текста и его формализация/ А.И. Новиков. - М.: Наука, 1983.- 361 с.
13. Попов Э.В. Динамические интеллектуальные системы в управлении и моделировании/ Э.В. Попов. - М.: МИФИ, 1996.
14. Поспелов Н.С. Сложное синтаксическое целое и основные особенности его структуры/ Н.С. Поспелов// Сб. инта рус.яз., доклады и сообщения. - Вып.2. - М.-Л., 1948. - С.53.
15. Рылова Т.Н. Некоторые формальные критерии выявления семантических связей между предложениями текста/Т.Н. Рылова// Семантические проблемы автоматизации информационного поиска. - Киев: Наук. думка, 1971.- С. 73-84.
16. Скороходько Э.Ф. Семантические сети и автоматическая обработка текста/ Э.Ф. Скороходько. - Киев: Наук. думка, 1983. - 217 с.
17. Скрэг Г. Семантические сети как модели памяти/ Г. Скрэг// НЗЛ. - т. III. - М.: Прогресс, 1999. - С. 259-302.
18. Харламов А.А., Ермаков А.Е., Кузнецов Д.М. Технология обработки текстовой информации с опорой на семантическое представление на основе иерархических структур из динамических нейронных сетей / А.А. Харламов, А.Е. Кузнецов, Д.М. Кузнецов // Информ. Технологии. - 1998. - № 2. - М. - С.26-32.
19. Цейтин Г.С. Программирование на ассоциативных сетях / Г.С. Цейтин // ЭВМ в проектировании и производстве. - Л.: Машиностроение. Вып. 2, 1985. - 196 с.
20. Чехов А.П. Дама с собачкой. - Электронный текст. - 8 с.
21. Чудаков А.П. Чехов. Единство видения/ А.П. Чудаков// Слово - вещь - мир. От Пушкина до Толстого. - М.: Современ. писатель, 1992. - С. 105-131.
22. Ligozat G.Représentation des connaissances et linguistique.- Acquis avancés de l'informatique. Paris:Armand Collin, 1994. - С.29-41.