

УДК 81'42: 004

## СТРАТЕГИИ СТРУКТУРИРОВАНИЯ ТЕМАТИЧЕСКОГО ПРОСТРАНСТВА ТЕКСТА<sup>1</sup>

**Константин Игоревич Белоусов**

д. филол. н., профессор кафедры теоретического и прикладного языкознания  
Пермский государственный национальный исследовательский университет  
614990, Пермь, ул. Букирева, 15. belousovki@gmail.com

Статья посвящена изучению стратегий структурирования тематического пространства текста в контексте использования современных информационных технологий. Обращение к актуальной в 60–90-е гг. XX в. проблематике обусловлено новыми возможностями, открывающимися для экспериментальной лингвистики, использующей технологии сетевой организации исследовательской деятельности (сетевой науки). Современные сетевые технологии позволяют анализировать не только результат экспериментальной деятельности информанта, но и процесс выполнения заданий, создавая тем самым онлайн-модели когнитивной деятельности. Исследование основывается на экспериментальном методе тематического картирования текста, использование которого в психолингвистических исследованиях с 2004 г. позволило сопоставить возможности и результаты применения метода в двух экспериментальных режимах: обычного эксперимента и эксперимента с применением сетевых технологий. Полученные в ходе экспериментального исследования данные (формулировки темы и микротем; слов/фрагментов текста, приписанных испытуемыми (далее – Ии.) к микротемам; временные параметры выполнения заданий, с точностью до секунды фиксирующие все действия Ии., и нек. др.) использовались для построения моделей структурирования тематического пространства текста несколькими группами Ии., выделенных на основе проанализированного материала.

**Ключевые слова:** тема текста; тематическое пространство текста; тематическое картирование текста; стратегии структурирования; онлайн-модели; информационная система «Семограф»; локальные микротемы; сквозные микротемы; анализизм; синтетизм.

Современная когнитивная лингвистика охватывает широкий спектр научных проблем, связанных с извлечением, переработкой, хранением, представлением и транслированием информации. В то же время сфера анализа методов, средств, параметров и инструментов моделирования самого процесса переработки информации (извлечения и обобщения релевантных целей анализа знаний из текстов), в том числе и с учетом специфики языковой личности субъекта (т.е. лингвоперсоналогического подхода), остается недостаточно изученной. Данная статья посвящена описанию общих концептуальных основ проводимого авторами онлайн-моделирования когнитивной деятельности эксперта-лингвиста в процессе осуществления тематического анализа.

### 1. Тема и тематическое пространство текста

О *теме* существует значительный пласт научной литературы в самых разных филологических областях. Лингвистика текста, стилистика, анализ текста, литературоведение, психолингвистика и другие, так или иначе используют это понятие. И несмотря на широкое поле примене-

ний *темы*, ее понимание в целом не обнаруживает серьезных расхождений.

В лингвистике текста *тема* понимается как «смысловое ядро текста, конденсированное и обобщенное содержание текста» [Москальская 1981: 17]. Тема целого речевого произведения есть продукт синтеза микротем, каждая из которых представляет собой тему ССЦ расшифровать (СФЕ, строфы) [Дымарский 2001; Лосева 1980; Москальская 1981 и др.]. Полагается, что микротема непосредственно выводима из ССЦ (монотематичность ССЦ): «...сверхфразовое единство монотематично. Объединение всех составляющих его предложений вокруг одной темы есть проявление его смысловой целостности или семантической изотопии текста. Переход от одной темы к другой есть пограничный сигнал, знаменующий конец одного сверхфразового единства и начало следующего сверхфразового единства» [Москальская 1981: 19]. При таком подходе к выявлению микротем и темы текста возникают некоторые проблемы. Получается, что тема текста все-таки является слагаемой из микротем. Неслучайно М. Я. Дымарский, определяя отно-

шение каждого из микрофрагментов к общей теме текста, отмечает, что «эти отношения не выходят за рамки набора стандартных логических отношений расщепления, включения, сопоставления, противопоставления. Иначе говоря, микротема каждого микрофрагмента должна прямо выводиться из гипертемы, которой подчинен весь фрагмент» [Дымарский 2001: 94]. Тема как бы делится без остатка на составляющие ССЦ (СФЕ) микротемы, что парадоксально, поскольку в лингвистике текста была обоснована невозможность безостаточного деления текста на ССЦ (см., например: [там же: 98–108]).

Но основной проблемой, на наш взгляд, является игнорирование значимых для семантики *целого текста* микротем, не локализующихся в каких-то одиночных ССЦ, а распределяющихся по всему тексту. Такие микротемы, входя в состав большого количества ССЦ, зачастую не оказывают определяющего воздействия на формирование микротемы отдельного ССЦ. В качестве подобных «текстовых» микротем могут выступать, в частности, *мотивы*, рассредоточенные по всему текстовому пространству и выявляемые только на уровне целого [Гаспаров 1994], *ассоциативные цепи текста*, создающие «многомерное содержательное пространство текстовых перекличек» [Николаева 2000: 418], *ассоциативно-смысловые поля*, материализованные лексически, связанные парадигматически и синтагматически, соотносящиеся с одним концептом текста [Болотнова 2001: 130] и др. Кроме того, мы должны исходить из возможности репрезентации в тексте всякой *двуплановости* его содержания [Гальперин 1981: 40–42], для которой предложение или ССЦ являются контекстуально недостаточными (требуется знание всего текста). Слова и сочетания слов, являясь репрезентантами семантической двуплановости текста, в составе ССЦ могут раствориться в ее теме, потерять приобретенные на уровне целого текста дополнительные смыслы. Подход «от ССЦ» не может учесть процессов многообразного «врастания» концепта в ткань текста (*textus* – ткань, сплетение, соединение). Именно поэтому подход к семантической целостности текста «от ССЦ» реализует синтагматическое (горизонтальное, сукцессивное) направление в процессе становления темы, не учитывая парадигматического (вертикального, симультанного) ее становления.

«Вертикальное» становление темы текста основано на глобальной связности текста («...нет такого компонента, который бы не был связан хотя бы с одним другим компонентом текста...» [Мурзин 1991: 11]), – каждый его компонент вступает в многообразные и подчас противоречивые связи с другими компонентами. Объедине-

ние хотя бы двух компонентов текста, в результате чего появляется простейшая семантическая структура, может быть названо *микротемой*.

При таком подходе *тематическое пространство текста* предстает в виде *статистически* возникающих организованностей большей и меньшей степени общности. Тематическое пространство текста как целостность структурируется информантами в виде нескольких вероятностно предпочитаемых сценариев синтеза. Между такими статистическими образованиями идет борьба за доминирование, включение в себя новых и новых более элементарных образований. *Синтез побеждающих в этой конкуренции наиболее масштабных общностей и может быть назван темой текста*. Таким образом, для того чтобы определить тему текста, необходимо установить достоверно возникающие связи между компонентами, от элементарной до предельной общностей, и интерпретировать реконструированную предельную общность.

Полагаем, что наиболее релевантным методом изучения процессов синтеза темы и реконструкции тематического пространства текста является эксперимент. Экспериментальный метод, созданный для количественного моделирования тематической организации текста, получил наименование *тематического картирования текста*. Привлечение к процессу анализа текста большого числа информантов позволяет выявить доминантные внутритекстовые связи, набор которых, используя терминологию Т.М. Дридзе, можно назвать микроструктурой текста [Дридзе 1984: 87]. Кооперируясь друг с другом, внутритекстовые связи образуют группы разной степени общности: от отдельных лексем (самый нижний уровень) до групп лексем (уровни, находящиеся выше)<sup>2</sup>.

## 2. Тематическое картирование текста в режимах обычной коммуникации и сетевой науки

### 2.1. Описание метода тематического картирования текста (далее – ТКТ)

Цель эксперимента состоит в реконструкции системы семантических связей между словами текста и экспликации процесса синтеза темы текста.

Экспериментальные задания. Перед информантами ставятся задачи: 1) прочесть текст, определить его тему; 2) выделить микротемы текста и назвать их; 3) к каждой микротеме выписать слова, представляющие ее в тексте. Количество групп и слов в группах произвольно.

Условия эксперимента. Задания выполняются каждым испытуемым индивидуально. Текст подбирается объемом 150–300 словоформ. Время проведения эксперимента не ограничивается.

*Аудитория.* В качестве реципиентов выступают студенты-филологи. Нижний порог количества испытуемых установлен эмпирически и составляет 25 человек, что обусловлено появлением осмысленных структур в результате статистической обработки семантической карты, реконструируемой на основе 25 реакций.

Полученные реакции (интерпретации) состоят из лексико-семантических групп, в основе каждой из них лежит некоторый принцип (микротема). Лексемы, находящиеся в одной лексико-семантической группе, семантически связаны между собой тем принципом (гиперсемой), ради которого они и образовали данную группу.

Семантическая связь между двумя лексемами в тексте вытекает из самой его природы. Слово, становясь компонентом некоего ограниченного целого, уже не принадлежит само себе, оно теряет множество степеней свободы (возможных, но не актуализованных в данном контексте значений) ради нового синтеза – текста. Вступая в разные связи с другими словами в рамках целого, каждое слово тем самым становится компонентом разнообразных семантических подсистем, которые можно назвать микротемами, субцельностями и др.

Слова, находящиеся в одной микротеме, связаны между собой. Исходя из того что разные интерпретации порождают разные сценарии анализа и синтеза текста каждое слово может стать компонентом довольно широкого поля микро-

тем. *Сила семантической связи между двумя словами в тексте показывает, насколько часто в реакциях информантов два слова данного текста являются компонентами одной микротемы.* Семантическая связность каждого слова с остальными словами текста, рассмотренная в контексте всех микротем, отображается с помощью семантической карты (далее – С-карты). Очевидно, что связь между словами может иметь либо закономерный, либо случайный характер. Семантическая связь становится закономерной в том случае, когда два слова частотно включаются информантами в одну микротему. Порог значимости может быть установлен статистически.

#### *2.2. Особенности проведения эксперимента в режимах обычной и ИТ-коммуникации*

Эксперименты с использованием ТКТ проводились с 2004 г.; результаты исследований отражены в ряде публикаций, в частности: [Белоусов 2009]. В ходе исследования были получены результаты по тематическому картированию более двадцати текстов, на каждый из которых получены реакции 50–80 Ии. Разработка ИС «Семограф» с реализацией ТКТ позволила проводить эксперименты в информационной системе. В табл. 1 представлен сопоставительный анализ возможностей фиксации параметров эксперимента в двух экспериментальных ситуациях: обычного эксперимента и эксперимента в режиме ИТ-коммуникации (сетевой науки).

Таблица 1

**Сопоставление возможностей фиксации параметров эксперимента  
в двух экспериментальных ситуациях**

<i>Параметр</i>	<i>Обычный эксперимент</i>	<i>Эксперимент в режиме сетевой науки</i>
Размер текстов	до 150–250 слов	нет ограничений
Фиксация временных параметров выполнения заданий	нет	есть
Локализация приписываемых к микротеме слов	нет	есть
Обработка первичных данных	текст до 25–30 слов, количество Ии. 30–50	нет ограничений
Генерация выборок из собранного материала (по социально-демографическим и др. параметрам)	нет	есть
Возможность анализа других результатов одного Ии.	нет	есть

Размер текста в режиме обычного эксперимента ограничен 150–250 словами из-за временных параметров выполнения заданий (20–30 мин.). В режиме сетевой науки текст не ограничен в размерах, так как информант может вы-

полнять задания удаленно по сети. Обработка полученных данных накладывает дополнительные ограничения в размерах вследствие того, что построение С-карты для протяженных текстов потребует значительных временных затрат. Так,

данные одной анкеты информанта с тематическим анализом текста средней протяженности, состоящие из 5 микротем со средним количеством слов, привязанных к микротеме, равным 10, «требуют» от исследователя внести значения в более чем 250 ячеек С-карты. А для построения С-карты данного текста на материале 50 анкет Ии. исследователю нужно внести в таблицу более 13 тыс. значений. Поэтому на практике для ТКТ использовались тексты малых размеров.

Временные параметры выполнения заданий в режиме обычного эксперимента нельзя зафиксировать (кроме самых общих, т. е. всего времени выполнения заданий). В режиме сетевой науки существует возможность временной фиксации всех действий информанта. В табл. 2 приведен фрагмент статистики пользовательских действий информанта *astafuque* в рамках тематического анализа текста И. А. Бунина «Русь».

Таблица 2

**Статистика пользовательских действий (фрагмент)**

<i>Пользователь</i>	<i>Действие</i>	<i>Объект</i>	<i>Контекст</i>	<i>Интервал</i>	<i>Микротема</i>	<i>Слова</i>	<i>Порядок слова</i>
astafuque	save	contextHas Component	19	7	Описание старухи		
astafuque	save	collocation	19	9	Описание старухи	большая	12
astafuque	save	collocation	19	3	Описание старухи	бокастая	13
astafuque	save	collocation	19	18	Описание старухи	ходит в валенках	14 15 16
astafuque	save	collocation	19	5	Описание старухи	в теплой стеганой	
astafuque	remove	collocation	19	7	Описание старухи	в теплой стеганой	
astafuque	save	collocation	19	13	Описание старухи	лицо крупное желтоглазое	21 22 23

В первом столбце отображается имя того, кто произвел действие (имя информанта/эксперта); во втором – совершенное действие (добавление или удаление категории); в третьем – объект, над которым было совершено действие; в четвертом – номер контекста (каждому информанту соответствует свой контекст); в пятом – временной интервал, характеризующий совершенное действие; в шестом – название микротемы; в седьмом – приписанные к микротеме слова; в восьмом – порядок (номер) слова в тексте.

Различия между режимами экспериментов относятся и к локализации приписанных к микротеме слов (этот параметр важен для работы со словами, имеющими повторы в тексте). В режиме обычного эксперимента информант выписывает слово из текста без указания на то, рассматривает ли он это слово как лексему или как словоформу. Другими словами, приписывает ли он к микротеме данное слово в системе всех его повторов, или же, напротив, приписывает конкретное употребление лексемы вне остальных ее реализаций. Так, в обычном эксперименте слово может выписываться как в начальной форме, так и в любой форме в составе словосочетания /

предложения [Ичкинеева 2011]. Данное обстоятельство затрудняет процесс обработки реакций. В режиме же сетевых коммуникаций все приписываемые к микротеме слова имеют локализацию, а инструменты типа «приписать все повторы слова» полностью решают проблему локализации слов.

Эксперимент, осуществляемый с применением информационных технологий и созданных на их основе программных средств, дает исследователю и другие преимущества, в частности, широкое использование метаданных (дополнительных данных, характеризующих Ии., экспериментальную ситуацию, стимульный материал и др.) для автоматической генерации выборки по интересующим значениям метаданных. Другие возможности эксперимента, проводимого с использованием программного фреймворка, связаны с представлением и обработкой данных средствами научной визуализации и моделирования.

Рассмотрим результаты тематического картирования текста И. А. Бунина «Русь», полученные в двух описываемых экспериментальных ситуациях.

Среднее количество слов, выделяемых в одной реакции в обычном эксперименте, – 72; в режиме

IT-коммуникации – 147,8; среднее количество слов в микротеме в обычном эксперименте – 11,6, в режиме IT-коммуникации – 22,95. В то же время среднее количество микротем, приходящееся на одну реакцию в обычном эксперименте, составляет 6,7, а в режиме IT-коммуникации – 6,4. Видно, что изменение экспериментальной ситуации повлекло кардинальное изменение в отборе слов, но на уровне тематической членимости текста изменений не наблюдается.

### 3. Стратегии тематического анализа и синтеза

Одним из главных недостатков экспериментальной лингвистики (и лингвистических дисциплин, непосредственно основывающихся на эксперименте) является создание идеализированного информанта, которому приписываются все результаты, извлекаемые из собранного материала десятков и сотен испытуемых. Идеализация как инструмент научного познания должна соотноситься с рамками проблемы каждого конкретного научного исследования и являться средством представления относительно однородных (в статистическом смысле этого слова) групп носителей языка.

Вариантов выделения групп из некоторой совокупности реакций может быть несколько:

1) создание сбалансированных выборок по социально-демографическим параметрам (возраст, пол, образование, специальность/сфера занятости, место рождения / проживания и др.) и/или этнической принадлежности, как это принято в исследованиях социолингвистической направленности. В данном случае сбалансированность дает возможность изучения влияния одних параметров на другие, но не отражает распределения значений этих параметров в языковой действительности (например, очевидно, что параметры пола и типа образования (гуманитарное и естественно-научное) распределяются не в равных пропорциях);

2) создание выборок на основе первичного статистического анализа собранного материала, например, с помощью кластеризации данных. Такой имманентный подход предполагает, что выделенные кластеры (т. е. группы Ии.) являются основой вычленения выборок из всего корпуса экспериментальных реакций для последующего анализа данных в пределах каждого отдельного кластера. Данный подход можно назвать когнитивным, так как закономерности организации пространства языковых экспериментальных данных выводятся из самих результатов когнитивной деятельности носителей языка;

3) синтез социолингвистического и когнитивного подходов в социокогнитивный подход. Полагаем, что независимыми параметрами в экспе-

рименте могут являться не только социально-демографические и иные характеристики, но и выделенные в результате первичного статистического анализа кластеры / группы Ии.

Как отмечалось выше, эксперимент в режиме сетевой науки имеет ряд преимуществ перед обычным психолингвистическим экспериментом, в том числе и в аспекте группировки реакций Ии. Изучение стратегий тематического анализа текста подготовленными информантами (как правило, филологами) предполагает выделение групп Ии. на основе анализа результатов их деятельности. Рассмотрим сказанное на примере анализа текста И. А. Бунина «Русь» [Бунин 1992: 198]. Предложения в тексте пронумерованы для удобства интерпретации.

#### РУСЬ

1. Старуха приехала в Москву издалека.  
2. Свой северный край называет Русью.  
3. Большая, бокастая, ходит в валенках, в теплой стеганой безрукавке. 4. Лицо крупное, желтоглазое, в космах толстых седых волос, – лицо восемнадцатого века.

5. Спросил ее как-то:

– А сколько вам лет будет?

6. – Семьдесят семь, господин милый.

7. – А вы, дай Бог не сглазить, еще совсем хоть куда.

8. – А что ж мне? 9. Это года не велики.

10. Наш родитель до ста лет дожил.

11. Чаю она не пьет, сахару не ест. 12. Пьет горячую воду с черным хлебом, с селедкой или солеными огурцами.

13. – Вы никогда, небось, и не хворали?

14. – Нет, трясовица была на мне, порча на мне была. 15. Мужа страшилась: как он ко мне с любовным чувством, меня и начинало трясти, корежить. 16. Сжечь бы ее, ту, что напустила на меня это!

17. Слово «сжечь» одно из ее любимых.

18. Про большевиков говорит очень строго:

– Не смеют они так про Бога говорить. 19. Бог наш, а не их. 20. Сжечь бы их всех!

21. Ее рассказы о родине величавы. 22. Леса там темны, дремучи. 23. Снега выше вековых сосен. 24. Бабы, мужики шибко едут в лубяных санках, на кубастых лохматых коньках, все в лазоревых, крашеного холста тулупах со стоячими аршинными воротами из жестокого псиневого меху и в таких же шапках. 25. Морозы грудь насквозь прожигают. 26. Солнце на закате играет как в сказке: то блещет лиловым, то кумачовым, а то все кругом рядит в золото или зелень. 27. Звезды ночью – в лебяжье яйцо.

В качестве информантов выступили 34 филолога (бакалавры, магистры, преподаватели ПГНИУ). Эксперимент проводился в ИС «Семограф». Отдельные результаты (С-карта, С-граф), полученные на материале всех экспериментальных реакций, отчасти представлены в статье [Белоусов 2014]. Здесь же остановимся на выявлении стратегий тематического анализа текста.

На первом этапе исследования генерируется таблица  $N \times M$ , где  $N$  – графические слова текста, приведенные в порядке их следования в тексте;  $M$  – количество микротем во всех реакциях Ии. При этом каждая микротема дополнительно маркируется на принадлежность информанту. На пересечении строки и столбца располагается ячейка с числом, значение которого может быть либо 0 (отсутствие слова в микротеме), либо 1 (наличие слова в микротеме).

Затем между всеми столбцами (микротемами) вычисляются семантические расстояния:

$$r_{ij} = \frac{\sum_{k=1}^M a_{ik} + a_{jk}}{\sum_{k=1}^M a_{ik} + a_{jk} + 1}$$

где  $r_{ij}$  – семантическое расстояние между  $i$ -й и  $j$ -й микротемами;  $a_{ik}$  и  $a_{jk}$  – значения (0 или 1)  $i$ -й и  $j$ -й микротем, относящиеся к  $k$ -му слову текста. Иными словами, семантическое расстояние между двумя микротемами понимается равным отношению суммы всех пар значений в том случае, когда оба члена пары имеют ненулевое значение, к сум-

ме всех возможных комбинаций, в которых хотя бы один член пары имеет ненулевое значение.

В результате генерируется матрица семантических расстояний  $M \times M$  (в нашем случае –  $209 \times 209$ ) между всеми микротемами текста.

На втором этапе генерируется матрица сходства между Ии. Поскольку каждая микротема приписана информанту, появляется возможность редуцировать матрицу микротем  $M \times M$  до матрицы информантов  $K \times K$  (в нашем случае –  $34 \times 34$ ). В ячейке  $b_{ds}$  располагается показатель того количества микротем, которые превышают пороговые значения, устанавливаемые для определения семантической близости между микротемами.

Полученная матрица  $K \times K$  называется матрицей сходства.

На третьем этапе матрица сходства преобразуется в матрицу различий.

$$r_{ds} = 1 - \frac{b_{ds}}{A_d \cap A_s}$$

где  $r_{ds}$  – значение коэффициента в матрице различий между информантами  $D$  и  $S$ ;  $b_{ds}$  – значение показателя сходства (количество близких микротем между информантами  $D$  и  $S$ );  $A_d$  – общее количество микротем, которые выделил информант  $D$ ;  $A_s$  – общее количество микротем, которые выделил информант  $S$ .

На четвертом этапе осуществляется кластеризация матрицы различий [Боровиков 2003], результаты которой представлены на рис. 1.

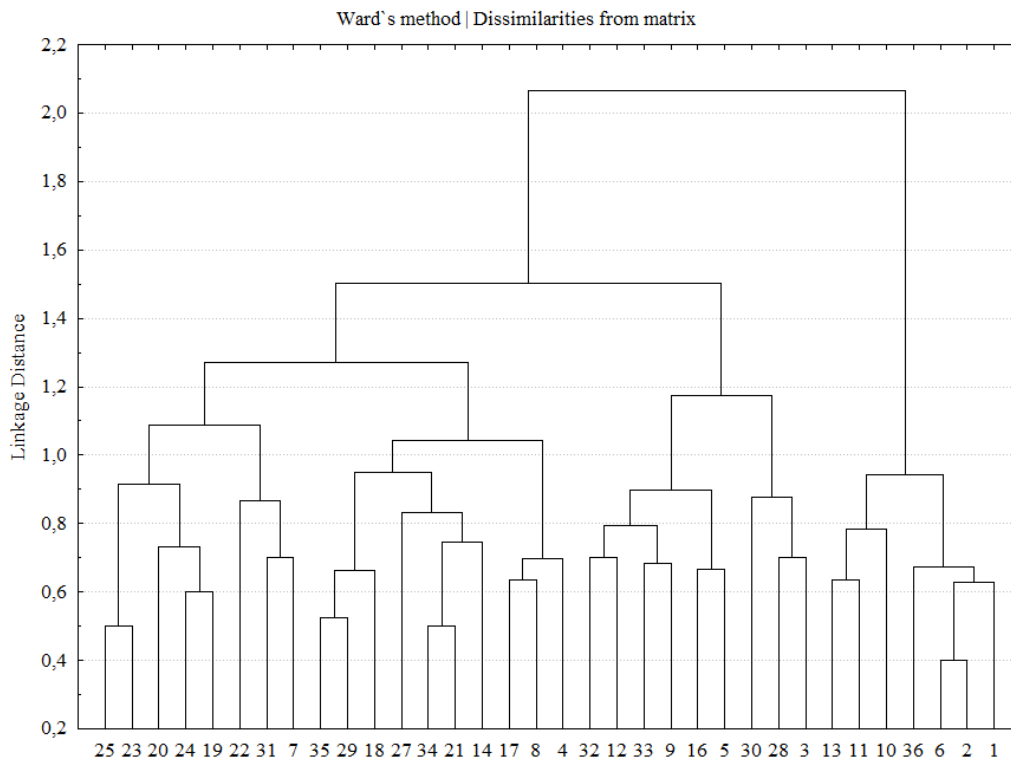


Рис. 1. Кластеризация Ии.

Результаты кластеризации позволяют выделить четыре относительно равнообъемных кластера:

Кластер 1 составляют Ии.: 1, 2, 6, 36, 10, 11, 13.

Кластер 2 составляют Ии.: 3, 28, 30, 5, 16, 9, 33, 12, 32.

Кластер 3 составляют Ии.: 4, 8, 17, 14, 21, 34, 27, 18, 29, 35.

Кластер 4 составляют Ии.: 7, 31, 22, 19, 24, 20, 23, 25.

Наиболее близки между собой кластеры 3 и 4, а наиболее удален от остальных кластер 1.

*Пятый этап* исследования посвящен определению специфики каждого из четырех кластеров. Одним из возможных направлений исследования специфики выделенных кластеров является анализ распределения микротем в линейном пространстве текста. Так как микротемы репрезентируются в тексте на уровне слов и каждое слово имеет локализацию в линейном пространстве

текста, то и микротемы становятся привязанными к этим же участкам текста, т. е. локализируются в нем. Однако в силу относительно большого размера данного текста представляется возможным использовать в качестве актуализированной микротематической единицы не слово, а предложение, в котором данное слово локализовано.

В результате исходные таблицы распределения слов в микротемах преобразуются в таблицы меньшей размерности распределения микротем в предложениях данного текста. Для каждого кластера Ии. генерируется отдельная таблица, которая затем используется для выявления параметров распределения микротем и кластеризации тематического пространства. В табл. 3 представлены некоторые характеристики распределения микротем в тексте; результаты кластеризации тематической структуры текста для каждой группы (кластера) Ии. представлены на рис. 2–5.

Таблица 3

**Параметры распределения микротем в тексте**

<i>Кластер</i>	<i>Параметры</i>	<i>Lm</i>	<i>V<sub>2/1</sub></i>
Кластер 1		11,42	1,57
Кластер 2		3,91	1,78
Кластер 3		5,93	2,60
Кластер 4		2,76	2,80

*Примечание.* Lm – средняя протяженность микротемы (в предложениях); V<sub>2/1</sub> – отношение количества микротем, локализованных в П14-П27, к количеству микротем, локализованных в Название-П13.

Параметр Lm дает представление о средней протяженности микротемы в реакциях Ии. каждой группы (кластера). Видно, что микротемы, характеризующие аналитическую деятельность Ии. Кластера 1, выделяются на общем фоне: их протяженность в 2–6 раз превышает протяженность микротем в реакциях Ии. других кластеров. Такая протяженность микротем свидетельствует о доминировании сквозных микротем, связывающих отдельные части текста, над микротемами локальными, являющимися темами (подтемами) отдельных текстовых фрагментов (ССЦ, абзацев и под.). Протяженность микротем Кластера 2 и Кластера 4, напротив, обусловлена доминированием стратегий локальной микротематической связи; микротемы Кластера 3 можно охарактеризовать как в равной степени находящиеся под влиянием локальных и глобальных (сквозных) стратегий, структурирующих тематическое пространство текста.

Параметр V<sub>2/1</sub> отражает неравносность тематической структуры текста более частую локализацию микротем во второй половине текста (П14-П27). Однако можно заметить, что если в

Кластере 1 и Кластере 2 соотношение микротематического «веса» второй половины текста к «весу» первой его половины относительно невелико, то в Кластере 3 и Кластере 4 микротематическое доминирование второй части текста над первой возрастает почти в три раза.

Результаты кластеризации тематической структуры текста на уровне предложений, репрезентирующих микротемы (рис. 2–5), свидетельствуют о разных стратегиях понимания текста (и результате его понимания).

Так, например, на рис. 2 тематическое пространство делится на два больших кластера, меньший из которых образуют предложения П2 и П21-П27, в нем описывается «Русь» в несобственно-прямой речи Старухи; больший, делящийся на два кластера, содержит описание Старухи от третьего лица (П1, П4, П17 – характерные, онтологические черты, П3, П11, П12, П18 – описание внешнего вида, характерного поведения) и описание Старухи в диалоге с автором, при этом «диалогический» кластер можно подразделить на кластер вопросов автора (П5, П13, П7) и кластеры ответов Старухи как на непосредственные

вопросы (П8, П9, П10), так и на задаваемые персонажем себе (П19, П20). Название текста включено в кластер вопросов автора, что можно интерпретировать как понимание названия Ии. данной группы в контексте структурированной диалогичности. Интересно, что П1 вынесено из кластера «Русь» в кластер «Старуха», в то время как в структуре тематического пространства

Кластера Ии. 2 (рис. 3) и Кластера Ии. 3 (рис. 4) данное предложение включено в такой же отдельный кластер, что уже не позволяет говорить о формировании данного кластера в рамках не-собственно-прямой речи.

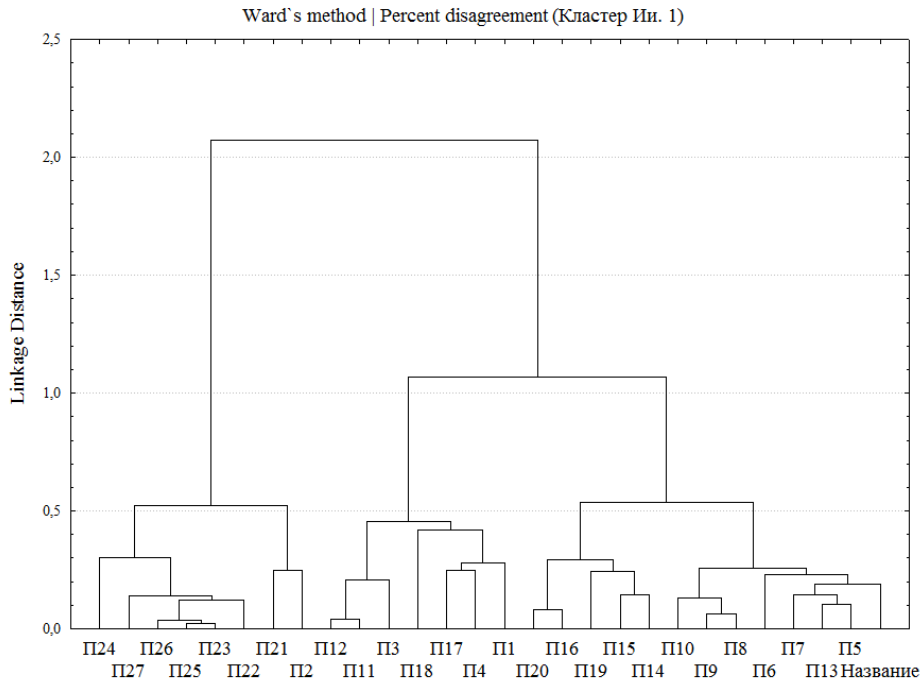


Рис. 2. Результаты кластеризации тематической структуры текста на уровне предложений, репрезентирующих микротемы (для Кластера Ии. 1)

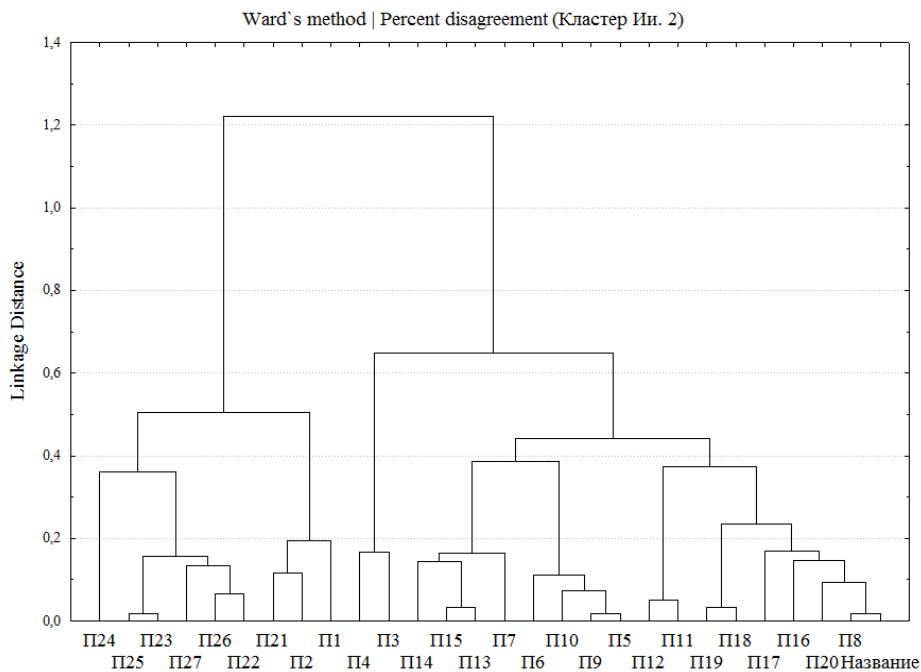


Рис. 3. Результаты кластеризации тематической структуры текста на уровне предложений, репрезентирующих микротемы (для Кластера Ии. 2)



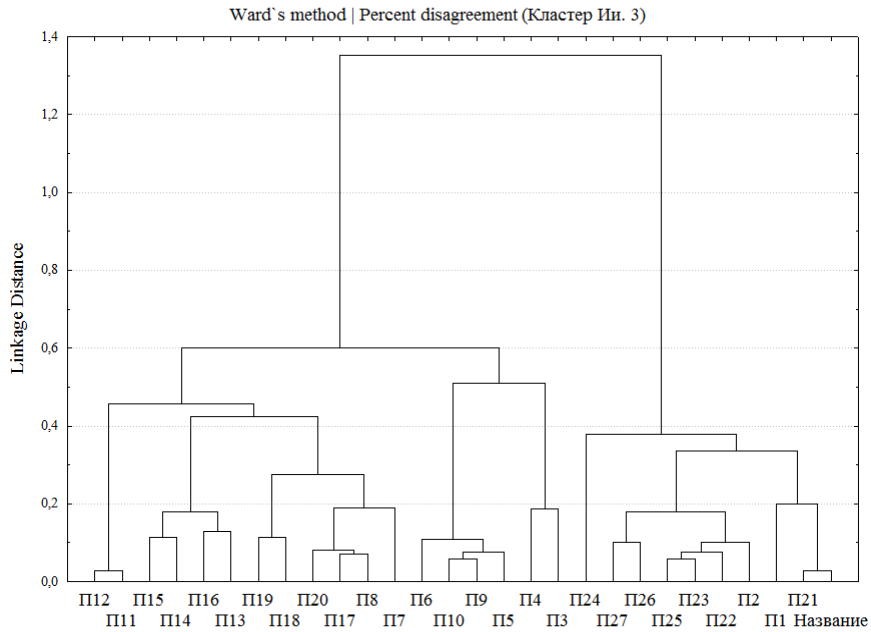


Рис. 4. Результаты кластеризации тематической структуры текста на уровне предложений, репрезентирующих микротемы (для Кластера Ии. 3)

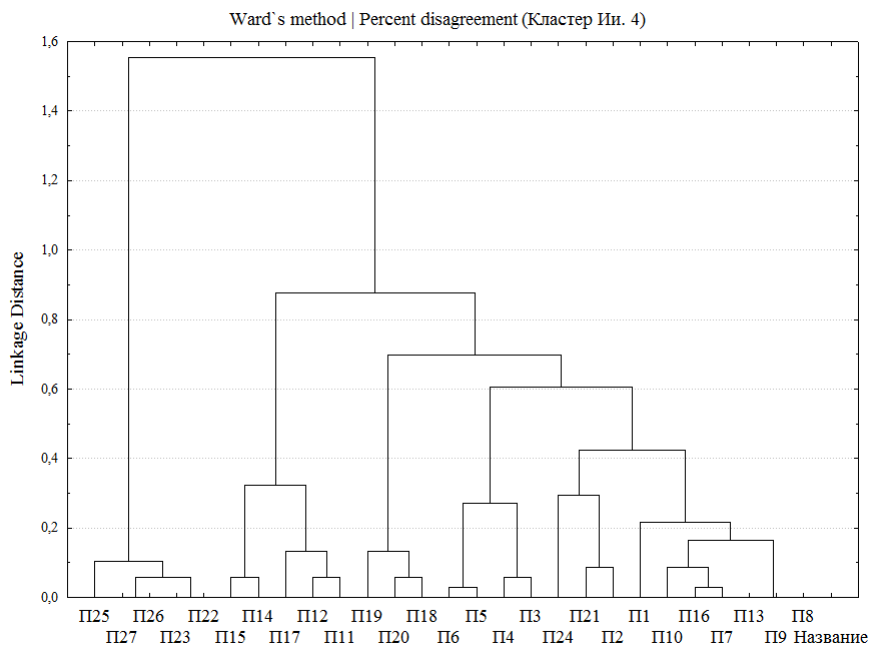


Рис. 5. Результаты кластеризации тематической структуры текста на уровне предложений, репрезентирующих микротемы (для Кластера Ии. 4)

Наиболее неравновесная тематическая структура представлена на рис. 5 (Кластер Ии. 4). Максимальная дискретизация тематического пространства и доминирование второй половины текста ( $L_m = 2,76$ ;  $V_{2/1} = 2,8$ ) приводят к такому сценарию структуризации, при которой на минимальную структурную основу структурными «слоями» добавляются остальные части тематического пространства. Это приводит к тому, что

распадается даже такой устойчивый кластер, как «Русь».

Кластер Ии. 2 показателен тем, что реализуемая его представителями стратегия структурирования тематического пространства текста «от уровня локальных микротем» оказывается почти столь же успешной, что и стратегия «глобальных (сквозных) микротем», используемая информантами Кластера 1.

Отдельно следует остановиться на соотношении полученных результатов с действием когнитивных стилей аналитичности и синтетичности (см., например: [Ичкинеева 2011]). Выявленные стратегии структурирования тематического пространства слабо коррелируют с показателями количества выделенных Ии. микротем. Так, средние показатели количества микротем Кластера 1, Кластера 2 и Кластера 3 варьируют в диапазоне  $6,4 \div 7,3$ . В каждом из данных кластеров присутствуют реакции как аналитиков (больше 9), так и синтетиков (меньше 4). В то же время Кластер Ии. 4 почти полностью состоит из синтетиков (среднее количество выделенных Ии. микротем 4,3). Протяженность микротем для данного кластера наименьшая среди остальных кластеров –  $L_m = 2,76$ , что также характерно для синтетиков, использующих локальный микротематический контекст для структурирования тематического пространства. При этом протяженность микротемы, измеренная в словах, составляет 18,6, что превышает значения всех остальных кластеров, кроме Кластера 1. Информанты, входящие в Кластер 4, приписывают к микротеме протяженные фрагменты текста, не выделяя в них значимую информацию. Этим объясняется и тот факт, что Ии., составившие Кластер 4, в среднем затрачивали 19,8 мин. на выполнение всей работы, в то время как обычно информанту требуется  $30 \div 40$  мин. на анализ данного текста.

### Заключение

Экспериментальное исследование стратегий структурирования тематического пространства текста имеет перспективы при обращении к методам сетевой организации исследовательской деятельности (сетевой науки). Автоматизация процессов сбора и обработки экспериментальных реакций позволяет, с одной стороны, значительно снизить трудозатратность и сделать эксперимент удобным средством получения знаний о процессах и механизмах порождения и понимания речи и текста, а с другой стороны, использовать преимущества, которые имеет экспериментальный подход перед другими. Современные сетевые технологии дают возможность анализировать не только результат экспериментальной деятельности информанта, но и процесс выполнения заданий, создавая тем самым онлайн-модели когнитивной деятельности.

### Примечания

<sup>1</sup> Исследование выполнялось при финансовой поддержке Российского гуманитарного научного фонда (проект № 12-34-01087).

<sup>2</sup> Деривационный принцип структурирования цельности текста рассматривается в работах Н. И. Жинкина, А. А. Леонтьева, Т. М. Дридзе, Л. В. Сахарного, Л. Н. Мурзина и др.

### Список литературы

Белоусов К. И. Теория и методология полиструктурного синтеза текста. М.: Флинта: Наука, 2009. 216 с.

Белоусов К. И. Экспериментальная лингвистика и сетевая наука // Социо- и психолингвистические исследования. 2014. Вып. 2. С. 21–31.

Болотнова Н. С. Филологический анализ текста. Томск: ТГПУ, 2001. Ч. I. 129 с.

Боровиков В. STATISTICA. Искусство анализа данных на компьютере: для профессионалов. СПб.: Питер, 2003. 688 с.

Бунин И. А. Окаянные дни: Дневники, рассказы, воспоминания, стихотворения. Тула: Приок. кн. изд-во, 1992. 319 с.

Гальперин И. Р. Текст как объект лингвистического исследования. М.: Наука, 1981. 139 с.

Гаспаров Б. М. Литературные лейтмотивы. Очерки по русской литературе XX века. М.: Наука: Вост. лит., 1994. 304 с.

Дридзе Т. М. Текстовая деятельность в структуре социальной коммуникации. М.: Наука, 1984. 270 с.

Дымарский М. Я. Проблемы текстообразования и художественный текст. На материале русской прозы XIX–XX вв. М.: Эдиториал УРСС, 2001. 328 с.

Ичкинеева Д. А. Аналитическая и синтетическая стратегии членения семантического пространства текста // Филология и человек. 2011. №4. С. 171–178.

Лосева Л. М. Как строится текст. М.: Просвещение, 1980. 94 с.

Москальская О. И. Грамматика текста. М.: Высш. шк., 1981. 183 с.

Мурзин Л. Н., Штерн А. С. Текст и его восприятие. Свердловск: Изд-во Урал. ун-та, 1991. 172 с.

Николаева Т. М. От звука к тексту. М.: Языки рус. культуры, 2000. 680 с.

### References

Belousov K. I. Eksperimental'naja lingvistika i setevaja nauka [Experimental linguistics and network science]. Socio- i psiholingvističeskie issledovanija [Socio and psycholinguistic research]. 2014. Iss. 2. P. 21–31.

Belousov K. I. Teorija i metodologija polistrukturnogo sinteza teksta [Theory and methodology of polystructural synthesis of the text]. Moscow, Flinta: Nauka Publ., 2009. 216 p.

*Bolotnova N. S.* Filologičeskij analiz teksta [Philological analysis of the text]. Tomsk, TSPU Publ., 2001. Part I. 129 p.

*Borovikov V.* STATISTICA. Iskusstvo analiza dannykh na kompjutere: dlja professionalov [STATISTICA. Art of the analysis of data on the computer: For professionals]. St. Petersburg, Piter Publ., 2003. 688 p.

*Bunin I. A.* Okajannye dni: Dnevnik, rasskazy, vospominanija, stikhotvorenija [Damned days: Diaries, stories, memoirs, poems]. Tula, Priokskoje Publ., 1992. 319 p.

*Dridze T. M.* Tekstovaja dejatel'nost' v strukture social'noj kommunikacii [Text activity in structure of social communication]. Moscow, Nauka Publ., 1984. 270 p.

*Dymarskij M. Ja.* Problemy tekstoobrazovanija i khudozhestvennyj tekst. Na materiale russkoj prozy XIX–XX vv. [Problems of organization of the text and literary text. On material of the Russian prose of the XIX–XX centuries]. Moscow, Ehditorial URSS Publ., 2001. 328 p.

*Gal'perin I. R.* Tekst kak objekt lingvističeskogo issledovanija [Text as object of linguistic research]. Moscow, Nauka Publ., 1981. 139 p.

*Gasparov B. M.* Literaturnye lejtmotivy. Očerki po russkoj literature XX veka [Literary leitmotif. Sketches on the Russian literature of the XX century]. Nauka: Vostochnaja literature Publ., 1994. 304 p.

*Ichkineeva D. A.* Analitičeskaja i sintetičeskaja strategii chlenenija semantičeskogo prostranstva teksta [Analytical and synthetic strategy of partitioning the semantic space of the text]. Filologija i čelovek [Philology and person]. 2011. No 4. P. 171–178.

*Loseva L. M.* Kak stroitsja tekst [As the text is under construction]. Moscow, Prosveshhenie Publ., 1980. 94 p.

*Moskal'skaja O. I.* Grammatika teksta [Grammatika of the text]. Moscow, Vysshaja škola Publ., 1981. 183 p.

*Murzin L. N., Shtern A. S.* Tekst i ego vosprijatie [Text and its perception]. Sverdlovsk, Ural. Univ. Publ., 1991. 172 p.

*Nikolaeva T. M.* Ot zvuka k tekstu [From a sound to the text]. Moscow, Jazyki russkoj kul'tury Publ., 2000. 680 p.

## STRATEGIES OF STRUCTURING THEMATIC SPACE OF TEXT

### Konstantin I. Belousov

Professor in the Department of Theoretical and Applied Linguistics  
Perm State University

The article considers strategies of structuring thematic space of text in the context of use of modern information technology. Broaching a range of problems urgent in the 1960-1990s is connected with the new opportunities of experimental linguistics that uses a technology of research activities networking (network science). Modern network technologies make it possible to analyze not only the result of an informant's experimental activity, but also its process itself, thus allowing for creation of an online model of cognitive activity. The study is based on the experimental method of thematic mapping of text, used in psycholinguistic research since 2004, which made it possible to compare the potential and results of applying the method under two experimental conditions: a conventional experiment and an experiment with network technologies. The data obtained in the pilot study (wording of the topic and micro-topics; words / blocks of text attributed to the micro-topics by informants; assignments timing showing all the informants' actions accurate to the second, etc.) have been used to construct models of structuring thematic space of text by several groups of informants picked out on the basis of the analyzed material.

**Key words:** topic of text; thematic space of text; thematic mapping of text; structuring strategies; an online model; information system Semograph; local micro-topics; prevailing micro-topics; analytical cognitive style; synthetic cognitive style.