

DOI: 10.18721/JHSS.10205

УДК 81'13; 81'33

НАУЧНЫЙ И ТЕХНИЧЕСКИЙ ТЕКСТ И ИНФОРМАЦИЯ 4.0: КЛЮЧЕВЫЕ ЗАДАЧИ ПРИ СОЗДАНИИ СТРУКТУРИРОВАННОГО КОНТЕНТА

Л.Н. Беляева¹, В.Е. Чернявская²

¹ Российский государственный университет им. А.И. Герцена,
Санкт-Петербург, Российская Федерация

² Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

В статье рассмотрена методологическая значимость формализации лингвистического знания в связи с новыми, технически обусловленными возможностями извлечения информации из текстов в научной и технической коммуникации. Анализируются возможности лингвистического инструментария, решающего задачу автоматического профилирования информации. Теоретические основы анализа заданы выводами современной лингвистики текста, стилистики научного текста, учитываются заключения прикладной лингвистики в части инженерии знаний, т. е. методов и средств извлечения, представления, структурирования и использования знаний. Для оценки соответствия создаваемых текстов ключевым требованиям структурированного контента проанализировано 128 научных статей, представленных в рецензируемый российский журнал «Известия РГПУ им. А.И. Герцена» в 2010–2016 гг. Исследованный массив включает работы, выбранные по таблице случайных чисел из общего массива независимо от решения экспертов об их публикации и соответствующие гуманитарным наукам по направлениям языкознание и лингвистика, история, культурология, философия, экономика и юриспруденция. Результаты исследования выборки показывают, что актуальные требования к представлению информации в текстах авторами не соблюдаются, компетенции в области академического письма в условиях Информации 4.0 не сформированы вполне.

Ключевые слова: Информация 4.0, текстовая структура, анализ и синтез контента, авторская разработка структурированного контента, контент-анализ

Ссылка при цитировании: Беляева Л.Н., Чернявская В.Е. Научный и технический текст и Информация 4.0: ключевые задачи при создании структурированного контента // Научно-технические ведомости СПбГПУ. Гуманитарные и общественные науки. 2019. Т. 10, № 2. С. 53–63. DOI: 10.18721/JHSS.10205

SCIENTIFIC AND TECHNICAL TEXTS IN THE FRAMEWORK OF INFORMATION 4.0: CONTENT ANALYSIS AND TEXT SYNTHESIS

L.N. Belyaeva¹, V.E. Chernyavskaya²

¹ Herzen State Pedagogical University of Russia, St. Petersburg, Russian Federation

² Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

The paper discusses issues of formalization of linguistic knowledge in the framework providing strategic value for information. This is focused in the concept of Information

4.0. This concept reveals that information can be presented and visualized as a cloud of “information molecules”. The information molecules can be accessed directly, assembled, and changed by different users in various formats, integrated with data, other information systems, and products. Information 4.0 examines the form, production, interaction and curation of information components. The underlying assumption of the paper is that information and knowledge are represented and transferred as a text. Thereafter the text structure is the result of information transfer and the starting point of information mining and extraction. The paper reflects that Information 4.0 demands new approaches in how we work on content. New research directions and applied tasks to be solved are seen in advancing from procedures formalizing text semantics to procedures providing machine-readable automatic information structuring and text production. The methodological focus is not on the ready text structures but on linguistic tools and operations explaining how to produce texts according to given models and content. Thereafter in the framework of Information 4.0 the focus is not on the basic model of the typical publication structure IMRED, but on the notion of structured content authoring and productivist approach to the text. Based on a corpus-assisted analysis of current papers of Russian authors in social sciences and humanities the research shows typical faults in the authors’ key words.

Keywords: Information 4.0, text structure, content analysis, text synthesis, structured content authoring

Citation: Belyaeva L.N., Chernyavskaya V.E., Scientific and Technical Texts in the framework of Information 4.0: content analysis and text synthesis, St. Petersburg State Polytechnical University Journal. Humanities and Social Sciences, 10 (2) (2019) 53–63. DOI: 10.18721/JHSS.10205

Введение, постановка проблемы

Новые приоритеты и сложные задачи научной и технической коммуникации в XXI в. обуславливают необходимость качественно иной эпистемической культуры и практики представления и распространения знания. Будем исходить из утверждения, что знание порождается, представляется и транслируется как текст в соответствии с закономерностями содержательной связности и распределения информации в текстовой структуре. Речевая/текстовая структура является объектом декодирования и источником извлечения информации. Для осознания новых векторов и новых акцентов принципиально значимыми представляются следующие факторы.

В современных условиях существенно изменились процессы движения научной информации, эволюционировала роль научных жанров в ее распределении и – в более широком контексте – меняются задачи и формат экспертизы научно-технического результата. В соответствии со сложившимся функционально и прагматически ориентированным подходом

к текстовой структуре профилирование, маркирование или, в другой терминологии, выдвижение информации и авторского результата являются задачей автора текста. Выверенность лингвистических стратегий автора по оптимальному выдвижению результата в структуре текста обеспечивает заметность результата для научного и технического сообщества. Соответственно в процессах движения информации традиционно существенная роль отводилась специфике речевого формулирования и риторике продвижения научного результата [1–7], моделям построения научного текста, типам текста/жанровым формам, связанным с аннотированием, реферированием, экспертной оценкой научного знания: аннотация, (авторский) реферат, обзорная статья, рецензия [8–12]. Дигитализация публикационной культуры, тенденция к открытому доступу (*open access*) и, главное, необходимость автоматизации процессов поиска и обработки меняет сложившиеся подходы в том отношении, что современный исследователь должен быть в большей степени ориентирован не на текст в его лингво-праг-

матическом, герменевтическом понимании генератора смыслов, структуры, подлежащей рецептивным закономерностям, но на технологию представления информации. Принципиально значимо, что эта технология должна рассматриваться только как производная от ранее принятого рассмотрения текста как структуры.

Новый подход к обмену информацией в промышленном производстве характеризуется недавно введенными терминами «Промышленность 4.0» (*Industry 4.0*) и «Информация 4.0» (*Information 4.0*) [См., например: 13]. Особенностью этого подхода является достижение максимальной гибкости производственных процессов за счет передачи оборудованию распределенных вычислений и независимых решений, принимаемых на основе цифровой информации. Уровень реализации разрабатываемых принципов новой Промышленности 4.0 зависит в том числе от методов создания, обмена и использования информации о разрабатываемом проекте, производстве, об эксплуатации конкретного прибора и/или технического устройства и о материальном обеспечении. Такая информация в условиях научной и технической коммуникации создается в виде текстов документации на всех этапах реализации проектов. От качества документов, создаваемых на исходном естественном языке и переводимых на все языки распространения продукции, зависит возможность применения высоких уровней автоматизации при их интерпретации и публикации.

Информация 4.0 характеризуется как: «молекулярная», формируемая не из структурно завершенных документов, а из информационных молекул; динамическая, т. е. непрерывно обновляемая; предоставляемая в распоряжение, а не передаваемая («*offered rather than delivered*»); универсальная (*ubiquitous*), интерактивная, доступная и удобная для поиска; спонтанная, т. е. вызываемая контекстами, профилируемая автоматически [Ср.: 14].

Проблема выбора структуры и типа текста переходит в новую плоскость, поскольку необходимо перейти от задачи формализации анализа семантики текста к задаче его создания и структурирования в соответствии с новыми контекстами использования текста и требованиями к нему. Новой исследовательской задачей нужно считать разработку алгоритмов из-

влечения информации из текстов в условиях ее автоматизированной обработки и трансфера.

Цель исследования

В рамках данной статьи цель анализа состоит в том, чтобы, первое, показать методологическую значимость формализации лингвистического знания в связи с новыми, технически обусловленными возможностями извлечения информации из текстов в научно-технической коммуникации и, второе, продемонстрировать эвристические возможности лингвистического инструментария, решающего задачу автоматического профилирования информации.

Методология

Методологически анализ проведен с учетом теоретических выводов современной лингвистики текста, стилистики научного текста, количественных методов, прикладной лингвистики в части инженерии знаний, т. е. методов и средств извлечения, представления, структурирования и использования знаний. Лингвистическая теория переосмыслена в направлении от декларативных («знаю, что») к процедурным («знаю, как») знаниям. Приоритет объяснительного подхода соответствует современному направлению развития науки о языке, которая может и должна анализировать и объяснять эффекты использования лингвистического инструментария речевых структур в разных коммуникативных практиках.

Результаты исследования

Теоретические основания и методология работы с Информацией 4.0. Формализация, кодификация лингвистической интерпретации, будучи первоначально связанной со структурно-ориентированным подходом в языкознании, получила развитие в последние десятилетия XX в. в направлении точных методов анализа языка в инженерной и компьютерной лингвистике, в естественно-научных экспериментах в когнитивной лингвистике, в нейролингвистике [См. подробнее об этом: 15–17]. Использование математических методов и новых возможностей современной вычислительной техники привело к новому витку исследований в области вероятностно-статистического моделирования процессов восприятия и производства речи. Исследования, основанные на

применении методов вероятностного прогнозирования и лингвостатистики к огромному эмпирическому материалу, позволили показать, что в отличие от большинства искусственных систем переработки, хранения и передачи информации язык представляет собой открытую динамическую неравновесную метасистему.

Для работы специалиста в контексте Информации 4.0 важен не анализ уже готовых текстовых структур, а построение текста по заданным, заранее определенным структурным моделям и контенту. Следовательно, информация, представленная на естественном языке (как правило, на языке контролируемом [18]) в виде научной и/или технической документации, должна быть подготовлена для использования в различных ситуациях, должна динамично приспосабливаться к различным сценариям производства, эксплуатации и материального обеспечения. Информация должна быть представлена так, чтобы ею можно было обмениваться на любых этапах реализации проекта. Это меняет подход к структуре текста и понимание роли отдельных компонентов в формировании текстовой информации, с опорой на которые можно извлечь существенную информацию [См. подробнее об этом: 19, 20].

Так, в соответствии с прагматически ориентированным подходом к текстовой структуре в рамках грамматики и семантики текста предполагается, что в тексте должно быть структурное и содержательно-смысловое подчеркивание определенных текстовых компонентов, несущих наиболее значимую, с точки зрения автора, информацию, их акцентирование в тексте через систему формальных и семантических средств привлечения внимания читателя. При таком подходе эталонной моделью построения и восприятия научного и/или технического текста считается принятая формула IMRAD (*introduction, method, results and discussion*/введение, метод, результаты и обсуждение).

Исследование структуры текстов для извлечения из них информации стало особой задачей с того времени: поток информации (научной, специальной, медийной) расширился так, что его оперативная и качественная переработка оказалась действительно сложной для тех, кому эта информация необходима (для специалистов и аналитиков, в задачу которых входят извлечение и обработка данных и

знаний). Анализ текстов с точки зрения их содержания получил широкое распространение при автоматизации информационного поиска (ИП), начавшейся еще в 1970-е гг. Задачей ИП был отбор текстов из заранее созданного или постоянно пополняемого массива по теме текста, определяемой запросом пользователя или извлекаемой из заранее зафиксированного набора тем, по наличию конкретной фактографической информации и т. п. При этом предполагалось, что система ИП должна автоматически извлекать из базы текстов именно те, которые соответствуют конкретному запросу/теме [См. подробнее об этом: 21–26]. Одной из важных процедур, реализация которой необходима для решения этой задачи, было предварительное индексирование и запросов, и самих текстов, что предполагало предварительное формулирование тем, которые интересны пользователю, или оперативное извлечение тех тем, которые можно выделить из текста. Тем самым можно утверждать, что темы (*topics*) соотносимы с системой организации тематических «молекул» текста при автоматизации его создания.

В основе выделения тем лежит предельно ясное предположение о том, что в структуре текста можно выделить особые (ключевые) единицы, в современной терминологии – ключевые термины (*key terms*), которые достаточно полно описывают его основное значение. Выбор ключевых слов на современном этапе рассматривается как основополагающая задача формулирования полученного результата [27]. Степень принадлежности термина или терминологического словосочетания к ключевым единицам, характеризующим проблематику конкретного текста (*keyness, aboutness of a text*), является важным параметром для дальнейшего использования текста специалистами и распространения результатов. Все современные метрики, с помощью которых выделяются ключевые единицы, опираются на эту идею. Меняются только конкретные подходы и процедуры, использующие все более специализированные методы статистического анализа и огромные массивы оцифрованных текстовых данных (*Big Data*).

Возможности автоматизированной обработки текстов по-новому ставят задачу выдвижения результата и его алгоритмической мар-

кированности. В основе нового подхода лежит анализ продуктивности (*productivist approach*), при котором степень детализации тематических разделов определяется задачами создания научной и технической документации и потенциально отделена от самого содержания, т. е. от тех тем, которые реально обсуждаются в тексте [См., например: 19].

В рамках подхода Информация 4.0 основным становится понятие *авторская разработка структурированного контента* (*structured content authoring*). Она означает структурирование содержания на части, называемые тематическими разделами (*topics*). Впоследствии они автоматически собираются с помощью карт (*maps*), чтобы создать окончательный вариант контента, предназначенный для использования в определенной функции и определенном типе документации. Тематические разделы должны жестко соответствовать темам текста. Тогда «молекулы информации» могут алгоритмически маркироваться и быть основой для создания текстов разных типов, использование которых, в свою очередь, может быть автоматизировано.

В таком контексте значимым эвристическим методом работы с текстом становится контент-анализ, являющийся инструментом исследования содержания текста, который определяет фактическое содержание (контент) текста и его внутренние характеристики. Этот инструмент основан на оценке содержания отдельного текста или совокупности текстов с опорой на содержащиеся в документах слова, словосочетания, символы или предложения, понятия, темы, и на количественном определении этого присутствия. Контент-анализ вообще рассматривается как относительно точный, формализованный метод исследования, основанный на эмпирических данных, извлекаемых из текстов, и необходимый как предварительный этап разработки системы создания текстов [28]. Решение стандартных задач и контент-анализа, и информационного поиска традиционно опирается на небольшие наборы конкретных ключевых терминов (ключевых слов) для извлечения соответствующей запросу информации относительно коллекции текстов, либо формируемой заранее как банк текстов, либо представляющей собой постоянно пополняемый поток текстов.

Эмпирические данные и обсуждение. Для оценки соответствия реальных текстов описанным выше требованиям было проанализировано 128 научных статей, представленных в рецензируемый российский журнал «Известия РГПУ им. А.И. Герцена» в 2010–2016 гг. В исследуемый массив были включены работы, выбранные по таблице случайных чисел из общего массива независимо от решения экспертов об их публикации и соответствующие гуманитарным наукам по направлениям языковедение и лингвистика, история, культурология, филология, экономика и юриспруденция. Анализ осуществлялся с применением инструментов корпусного менеджера AntConc – Word List, Concordance и Collocates, что позволило анализировать состав частотного списка, ближайшее окружение и сочетаемость слов в корпусе. Общий объем анализируемого массива научных текстов без учета аннотаций и ключевых слов составил 270 102 орфографических токена, количество разных словоупотреблений 48 741.

Результаты исследования выборки показывают, что актуальные требования к представлению информации в текстах авторами не соблюдаются, ни так называемая молекулярность, ни согласованность терминологии, ни соответствия структурным требованиям к научным статьям не выдерживаются вполне.

Особенностью исследуемого массива является его явная перегруженность стоп-словами, которые в частотном словаре занимают первые 35 рангов, а затем в словаре эти слова также имеют высокие ранги, что подтверждает активное использование различных связочных конструкций, служащих для поддержания логики изложения (см. табл. 1). Первым знаменательным словом с рангом 36 является словоформа *время*, встретившаяся в текстах с частотой $F = 367$. Рассмотрение конкорданса показывает, что чаще всего эта словоформа встречается в стандартных конструкциях типа *во время* ($F = 62$), *в настоящее время* ($F = 57$), *в то же время* ($F = 40$), *за это время* ($F = 25$); единственным тематическим словосочетанием является коллокация *пространство и время* ($F = 14$).

Поскольку основным источником информации о принадлежности текста определенной теме являются ключевые слова и аннотации, были получены частотные словари аннотаций (общий объем 6248 токенов, 3144 разных сло-

Таблица 1

Частотный список словоформ массива текстов статей (фрагмент)

Frequency list of wordforms of the text array of articles (Fragment)

Ранг	Частота	Словоформа	Ранг	Частота	Словоформа
1	11254	в	26	503	только
2	9685	и	27	497	г.
3	3872	с	28	484	во
4	3323	на	29	469	он
5	2294	не	30	445	можно
6	1959	как	31	415	же
7	1907	что	32	412	которые
8	1868	к	33	377	между
9	1598	а	34	375	этом
10	1453	по	35	371	все
11	1365	для	36	367	время
12	1138	из	37	358	было
13	1056	о	38	353	более
14	862	или	39	353	до
15	837	его	40	352	т.
16	799	от	41	351	у
17	760	это	42	336	образом
18	730	то	43	336	том
19	697	но	44	328	может
20	654	их	45	320	ее
21	640	при	46	309	быть
22	590	так	47	308	были
23	560	за	48	297	был
24	545	также	49	291	однако
25	519	является	50	291	они

воупотребления, 1974 леммы) и ключевых слов (общий объем 1166 токенов, 857 разных словоупотреблений, 739 лемм) (см. табл. 2).

Появление в качестве самых частых слов лексических единиц *статья* и *рассматривать*, как и их отсутствие в словаре ключевых слов, вполне естественно, поскольку в аннотациях привычным клише являются формы типа *в статье затрагиваются; анализируются; рассматриваются; уточняются* и т. п. ($F = 67$), *статья посвящена* ($F = 24$).

Средняя длина аннотации в исследуемом массиве составляет 57 токенов, изменяясь от минимальной длины в 10 токенов до максимальной в 126. Анализ соответствий лемм в аннотации и ключевых словах основывался на предположении, что ключевые слова должны

быть использованы в аннотации, однако это предположение не подтвердилось: из 739 разных лемм, использованных в ключевых словах, 204 леммы не встретились в аннотациях.

Рассмотрим особенности самой длинной аннотации и ключевых слов к соответствующей статье (см. табл. 3). Не оценивая качество и адекватность перевода на английский язык, отметим несоответствие набора ключевых слов и текста аннотации (совпадения выделены жирным шрифтом). Введение в состав ключевых слов терминов, не поддерживаемых аннотацией, может привести к неверному отнесению статьи к конкретной теме. Кроме того, сама аннотация представляет собой скорее реферат, поскольку не актуализирует конкретный фокус исследования именно в этой статье. Поэтому

Таблица 2

Частотный список лемм в аннотациях в сравнении с ключевыми словами (фрагмент)
Frequency list of lemmas in annotations in comparison with keywords (Fragment)

Номер в списке	Лемма	Аннотации		Ключевые слова	
		Ранг	Частота	Ранг	Частота
1	статья	3	117		
2	рассматривать	6	54		
3	анализ	9	33	29	5
4	основной	16	29		
5	проблема	18	28	549	1
6	автор	19	27		
7	основа	21	27	142	2
8	процесс	22	27	66	3
9	современный	23	27	17	6
10	исследование	24	26	113	2
11	развитие	25	25	67	3
12	посвятить	27	24		
13	культура	29	22	2	13
14	образование	33	21	8	7
15	формирование	35	21		
16	представлять	37	20		
17	система	38	20	10	7
18	социальный	39	20	1	16
19	художественный	40	20	27	5
20	изучение	42	18		
21	особенность	43	17	492	1
22	различный	44	17		
23	деятельность	45	16	56	3
24	работа	46	16		
25	анализировать	47	15		

Примечание. В таблице приведен список из 25 самых частотных лемм из аннотаций с информацией об их корреляции в ключевых словах.

Таблица 3

Сопоставление названия, ключевых слов и аннотации (фрагмент)
Comparison of the name, keywords and annotations (Fragment)

Описание на русском языке	Описание на английском языке
Эвристический метод концептуальной структуризации знаний при формализации слабоструктурированных ситуаций на основе когнитивных карт	Heuristic method for conceptual structurization of knowledge in the course of ill-structured situations formalization based on cognitive map
слабоструктурированная ситуация, когнитивная карта, формализация, риски из-за человеческого фактора, концептуальная структуризация экспертных знаний, SWOT-схема (анализ), сценарий (16 токенов)	ill-structured situation, cognitive map, formalization, risks due to human factors, conceptual structurization of expert knowledge, SWOT-scheme, scenarios (17 токенов)

Окончание табл. 3

Описание на русском языке	Описание на английском языке
<p>Концептуальная структуризация относится к ранним этапам формализации знаний разнородной группы экспертов о развитии слабоструктурированных (проблемных) ситуаций (СС) на основе когнитивных карт. Роль этого этапа, как правило, недооценивается, а к формализации на основе когнитивных карт (и других моделей) относятся как к формальной процедуре выбора факторов и установления связей. Опыт решения практических задач по анализу слабоструктурированных ситуаций показывает успешность применения концептуальных схем (SWOT, PEST и ряда уникальных схем) на этапе структуризации знаний. Применение SWOT-схемы, с одной стороны, позволяет выделить основные предметные области и их укрупненные взаимосвязи, значимые при анализе развития исследуемой ситуации, а с другой – осложняется из-за ряда трудностей и недостатков. В статье предлагается краткий обзор методов построения когнитивных карт, общая схема процесса формализации с выделением значимых этапов и метод структуризации знаний о развитии СС на основе SWOT-схемы (126 токенов)</p>	<p>In the analysis of ill-structured situation (SS), the principal is to identify and study has disputed factors, the impact of which only experts express assumption. The article provides the method for selecting the important factors based on SWOT-schemes and the construction of scenarios analysis of the situation, knowledge of which are presented in the form of formal cognitive maps. (59 токенов)</p>

Примечание. В таблице жирным шрифтом отмечены лексические единицы, совпадающие в названии, ключевых словах и тексте аннотации. Анализ аннотации показывает, что она в подобной форме не может рассматриваться как молекула текста.

объем аннотации на английском языке меньше аннотации на русском языке практически вдвое.

Проведенный анализ в рамках выборки свидетельствует, что исследователи гуманитарных областей знаний не демонстрируют сформированных компетенций в области академического письма и соответственно не ориентированы на работу в рамках Информации 4.0

Заключение

Изучение, а затем и моделирование языковых механизмов – одна из важнейших и пока еще не решенных задач современной лингвистики. Ее решение может быть, по нашему мнению, связано с другим, казалось бы, не соотносимым напрямую с этим процессом – созданием текста из набора заранее созданных единиц контента. При внешней разнонаправленности задач контент-анализа и задач создания научных и технических текстов они тесно связаны концептуально. Разработка структу-

рированного контекста и его тематическое моделирование являются основой как процессов анализа научного и технического текста, так и процессов создания документации.

Одновременно с этим утверждением считаем необходимым подчеркнуть, что лингвистическое обоснование и компьютерная реализация методов и приемов, необходимых для создания «молекул» текста в рамках концепции и инструментария Информации 4.0, опирается на многолетние сложившиеся исследования структуры текстов и их семантики. Последние проводились и проводятся в рамках решения теоретических и прикладных задач контент-анализа и информационного поиска. Решение современных новых задач возможно только и именно через необходимый баланс знаний, инструментов и методов, позволяющих соотносить эти два направления практически.

Перспективы и рекомендации в прикладном аспекте следует видеть в решении задач подготовки новых типов специалистов в обла-

сти обработки текстов. Специалистам, работающим с новыми формами представления информации, соответствует английский термин *language worker*, который можно приблизительно перевести как «специалист в области переработки текстов» [См. подробнее об этом: 29]. Такой термин используется как объединяющая номинация для терминологов, переводчиков, всех тех, кто создает техническую документацию (технических писателей/*technical authors*,

technical writers), специалистов по передаче технической информации (*technical communicators*). От современных специалистов в области обработки текстов требуются знания типологии специальных и технических текстов на родном и иностранных языках; умения создавать все типы специальных текстов как на родном, так и на иностранном языке; умения переводить тексты с учетом различий в требованиях к специальным текстам в конкретных культурах [30].

СПИСОК ЛИТЕРАТУРЫ

1. **Bazerman Ch.** Shaping Written Knowledge: The Genre and Activity of the experimental Article in Science. Madison: Univ. of Wisconsin Press, 1988.
2. **Bazerman C., Paradies J.** (Eds.). Textual Dynamics of the Professions: Historical and contemporary studies of writing in professional communities. Madison: Univ. of Wisconsin Press, 1991.
3. **Gross A.G.** The Rhetoric of Science. Cambridge, MA: Harvard Univ. Press, 1990.
4. **Kreuzbauer G., Gratzl N.** Persuasion und Wissenschaft: aktuelle Fragestellungen von Rhetorik und Argumentationstheorie. Wien: LIT, 2007.
5. **Tognini-Bonelli E., del Lungo Canicotti G.** (Eds.) Strategies in Academic Discourse. John Benjamins Publ., 2005.
6. **Чернявская В.Е.** Выдвижение нового знания в структуре публикации // Филол. науки. Науч. доклады высш. школы. 2017. № 5. С. 99–107. DOI 10.20339/PhS.5-17.099.
7. **Чернявская В.Е.** Научный дискурс: выдвижение результата как коммуникативная и языковая проблема. М.: УРСС : Либроком, 2017. 144 с.
8. **Chernyavskaya V.** “Monoculture of the Mind” and Cultural Diversity in Knowledge Dissemination // J. of Siberian Federal Univ. Humanities & Social Sciences. 2017. Vol. 10, no. 2. Pp. 219–227. DOI: 10.17516/1997-1370-0024.
9. **Noguchi J.** The Science Review Article. An opportune Genre in the Construction of Science. Bern: Peter Lang, 2006.
10. **Porter A.L., Kongthon A., Lu J.C.** Research profiling: improving the literature review // Scientometrics. 2002. Vol. 53, no. 3. Pp. 351–370.
11. **Van Leeuwen T.N., Visser M.S., Moed H.F., Nederhof T.J., van Raan A.F.J.** The Holy Grail of science policy: exploring and combining bibliometric tools in search of scientific excellence // Scientometrics. 2003. Vol. 57, no. 32. Pp. 257–280.
12. **Swales J.M., Feak Ch.B.** Abstracts and the Writing of Abstracts. Univ. of Michigan Press (Michigan Series in English for Academic & Professional Purposes), 2009.
13. **Gollner J.** Information 4.0 for Industry 4.0 // Towards a European Competence Framework // tekomp Jahrestagung und tcworld conference in Stuttgart. Zusammenfassungen der Referate. Stuttgart: tcworld GmbH Verantwortlich, 2016. Pp. 93–94.
14. **Gallon R.** Information 4.0, the Next Steps // Towards a European Competence Framework // tekomp Jahrestagung und tcworld conference in Stuttgart. Zusammenfassungen der Referate – Stuttgart: tcworld GmbH Verantwortlich, 2016. Pp. 95–97.
15. **Беляева Л.Н., Чернявская В.Е.** Доказательная лингвистика: метод в когнитивной парадигме // Вопросы когнитивной лингвистики. 2016. № 3. С. 77–84. DOI: 10.20916/1812-32282016-3-77-84.
16. **Belyaeva L., Chernyavskaya V.** On Methodology of Knowledge Presentation: Terms and translation in seismic discourse // J. of Siberian Federal Univ. Humanities & Social Sciences. 2016. Vol. 9, no. 12. Pp. 2904–2912. DOI: 10.17516/1997-1370-2016-9-12-2904-2912.
17. **Чернявская В.Е.** Методологические возможности дискурсивного анализа в корпусной лингвистике // Вестн. Томского гос. ун-та. Филология. 2017. № 50. С. 135–148. DOI: 10.17223/19986645/50/9.
18. **Muegge U.** Controlled language – does my company need it? URL: www.tekom.de/artikel/artikel_2756.html. 2009.
19. **Lacroix F.** Writing for the 21st Century // Towards a European Competence Framework // tekomp Jahrestagung und tcworld conference in Stuttgart. Zusammenfassungen der Referate – Stuttgart: tcworld GmbH Verantwortlich, 2016. Pp. 102–106.
20. **Albers M.J., Mazur B.** (Eds.). Context and Complexity: information Design in Technical Communication. N.J. Erlbaum, 2003.
21. **Hackos J.T.** Information Development: Managing your documentation projects, portfolio and people. N. Y.: Wiley, 2007.

22. **Rivers N.A.** Future Convergences: Technical Communication Research as Cognitive Science // *Technical Communication Quarterly*. 2011. Vol. 20, no. 4. Pp. 412–442. DOI: 10.1080/10572252.2011.591650.

23. **Rude C.D.** Mapping the Research Questions in Technical communication // *J. of Business and Technical Communication*. 2009. Vol. 23, no. 2. Pp. 174–215. DOI: 10.1177/1050651908329562.

24. **Salton G.** *Automatic Text Processing: the Transformation, Analysis, and Retrieval of Information by Computer*. Boston: Addison-Wesley, 1989.

25. **Spinuzzi C.** *Tracing Genres through Organisations: a Sociocultural approach to Communication Design*. Cambridge: MIT Press, 2003.

26. **Беляева Л.Н.** Лингвистические автоматы в современных информационных технологиях. СПб.: Изд-во РГПУ им. А.И. Герцена, 2001. 130 с.

27. **Scott M., Tribble C.** *Textual Patterns: keyword and corpus analysis in language education*. Amsterdam: Benjamins, 2006.

28. **He Q., Chang K., Lim E., Banerjee A.** Keep It Simple with Time: A Re-examination of Probabilistic Topic Detection Models // *IEEE Transactions on pattern analysis and machine intelligence*. 2009. URL: <http://wwwusers.cs.umn.edu/~banerjee/papers/09/pami-tdt.pdf> (дата обращения: 01.10.2017).

29. **Беляева Л.Н.** Лингвистические технологии в современном сетевом пространстве: language worker в индустрии локализации. СПб.: Книжный дом, 2016. 134 с.

30. **Беляева Л.Н., Блохина В.А., Камшилова О.Н.** Концепция профессиональной подготовки лингвотехнолога в России // *Изв. РГПУ им. А.И. Герцена*. 2017. № 183. С. 62–72.

Беляева Лариса Николаевна

E-mail: laurabel@gmail.com

Чернявская Валерия Евгеньевна

E-mail: tcherniavskaja@rambler.ru

Статья поступила в редакцию 17.03.2019 г.

REFERENCES

[1] **Bazerman Ch.**, *Shaping Written Knowledge: The Genre and Activity of the experimental Article in Science*, Univ. of Wisconsin Press, Madison, 1988.

[2] **Bazerman C., Paradies J.** (Eds.), *Textual Dynamics of the Professions: Historical and contemporary studies of writing in professional communities*, Univ. of Wisconsin Press, Madison, 1991.

[3] **Gross A.G.**, *The Rhetoric of Science*, Cambridge, MA, Harvard Univ. Press, 1990.

[4] **Kreuzbauer G., Gratzl N.**, *Persuasion und Wissenschaft: aktuelle Fragestellungen von Rhetorik und Argumentationstheorie*, LIT, Wien, 2007.

[5] **Tognini-Bonelli E., del Lungo Caniciotti G.** (Eds.), *Strategies in Academic Discourse*, John Benjamins Publ., 2005.

[6] **Chernyavskaya V.E.**, [Foregrounding of new Knowledge in the Text Structure], *Philologicheskie nauki. NDVS*, 5 (2017) 99–107. DOI 10.20339/PhS.5-17.099.

[7] **Chernyavskaya V.E.**, *Nauchnyy diskurs: vydvizhenie nauchnogo resul'tata kak kommunikativnaya i yazykovaya problema [Scientific Discourse: Foregrounding of a new Result]*, URSS, Librocom, Moscow, 2017.

[8] **Chernyavskaya V.**, “Monoculture of the Mind” and Cultural Diversity in Knowledge Dissemination, *Journal of Siberian Federal Univ. Humanities & Social Sciences*, 10 (2) (2017) 219–227. DOI: 10.17516/1997-1370-0024.

[9] **Noguchi J.**, *The Science Review Article. An opportune Genre in the Construction of Science*, Peter Lang, Bern, 2006.

[10] **Porter A.L., Kongthon A., Lu J.C.**, Research profiling: improving the literature review, *Scientometrics*, 53 (3) (2002) 351–370.

[11] **Van Leeuwen T.N., Visser M.S., Moed H.F., Nederhof T.J., van Raan A.F.J.**, The Holy Grail of science policy: exploring and combining bibliometric tools in search of scientific excellence, *Scientometrics*, 57 (32) (2003) 257–280.

[12] **Swales J.M., Feak Ch.B.**, *Abstracts and the Writing of Abstracts*, Univ. of Michigan Press (Michigan Series in English for Academic & Professional Purposes), 2009.

[13] **Gollner J.**, *Information 4.0 for Industry 4.0, Towards a European Competence Framework // tekomp Jahrestagung und tcworld conference in Stuttgart. Zusammenfassungen der Referate*, Stuttgart: tcworld GmbH Verantwortlich, 2016, pp. 93–94.

[14] **Gallon R.** *Information 4.0, the Next Steps, Towards a European Competence Framework // tekomp Jahrestagung und tcworld conference in Stuttgart. Zusammenfassungen der Referate*, Stuttgart: tcworld GmbH Verantwortlich, 2016, pp. 95–97.

[15] **Belyaeva L.N., Chernyavskaya V.E.**, [Evidence-based linguistics: methods in cognitive paradigm], *Vo-*

prosy kognitivnoy lingvistiki, 3 (2016) 77–84. DOI: 10.20916/1812-32282016-3-77-84.

[16] **Belyaeva L., Chernyavskaya V.** On Methodology of Knowledge Presentation: Terms and translation in seismic discourse, *Journal of Siberian Federal Univ. Humanities & Social Sciences*, 9 (12) (2016) 2904–2912. DOI: 10.17516/1997-1370-2016-9-12-2904-2912.

[17] **Chernyavskaya V.**, [Towards methodological application of Discourse Analysis in Corpus-driven Linguistics], *Tomsk State Univ. Journal of Philology*, 50 (2017) 135–148. DOI: 10.17223/19986645/50/9.

[18] **Muegge U.**, Controlled language – does my company need it? Available at: www.tekom.de/artikel/artikel_2756.html. 2009.

[19] **Lacroix F.**, Writing for the 21st Century, Towards a European Competence Framework // *tekom-Jahrestagung und tcworld conference in Stuttgart. Zusammenfassungen der Referate*, Stuttgart: tcworld GmbH Verantwortlich, 2016, pp. 102–106.

[20] **Albers M.J., Mazur B.** (Eds.), *Context and Complexity: information Design in Technical Communication*, N.J. Erlbaum, 2003.

[21] **Hackos J.T.**, *Information Development: Managing your documentation projects, portfolio and people*, Wiley, New York, 2007.

[22] **Rivers N.A.**, Future Convergences: Technical Communication Research as Cognitive Science, *Technical Communication Quarterly*, 20 (4) (2011) 412–442. DOI: 10.1080/10572252.2011.591650.

[23] **Rude C.D.**, Mapping the Research Questions in Technical communication, *Journal of Business and*

Technical Communication, 23 (2) (2009) 174–215. DOI: 10.1177/1050651908329562.

[24] **Salton G.**, *Automatic Text Processing: the Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley, Boston, 1989.

[25] **Spinuzzi C.**, *Tracing Genres through Organisations: a Sociocultural approach to Communication Design*, MIT Press, Cambridge, 2003.

[26] **Belyaeva L.N.**, *Lingvisticheskie avtomaty v sovremennykh informatsionnykh tekhnologiyakh [Linguistic machines in the modern information technologies]*, RGPU im. A.I. Gertsena Publ., St. Petersburg, 2001.

[27] **Scott M., Tribble C.**, *Textual Patterns: keyword and corpus analysis in language education*, Benjamins, Amsterdam, 2006.

[28] **He Q., Chang K., Lim E., Banerjee A.**, Keep It Simple with Time: A Re-examination of Probabilistic Topic Detection Models, *IEEE Transactions on pattern analysis and machine intelligence*. 2009. Available at: <http://wwwusers.cs.umn.edu/~banerjee/papers/09/pami-tdt.pdf> (accessed 01.10.2017).

[29] **Belyaeva L.N.**, *Lingvisticheskie tekhnologii v sovremenom setevom prostranstve: language worker v industrii lokalizatsii [Linguistic Technologies in the modern Net-Space: language worker in the localization industry]*, Kniznyy dom, St. Petersburg, 2016.

[30] **Belyaeva L.N., Blokhina V.A., Kamshilova O.N.**, [Concept of language working training in Russia], *Izvestiya RGPU im. A.I. Gertsena*, 183 (2017) 62–72.

Belyaeva Larisa N.

E-mail: laurabel@gmail.com

Chernyavskaya Valeriya E.

E-mail: tcherniavskaia@rambler.ru

Received 17.03.2019.