

# ДОКУМЕНТАЛЬНЫЕ ИСТОЧНИКИ ИНФОРМАЦИИ

УДК 81'322.2

Л.Р. Комалова, Е.В. Майорова

## Ошибки и отклонения в процессе реферирования: взгляд лингвиста\*

*Представлены результаты эмпирического и экспериментального исследования вторичных текстов, являющихся кратким изложением текстов-источников. Проанализировано 24 информативных реферата, созданных 15 авторами с опытом и без опыта реферирования, 27 индикативных рефератов, составленных 15 авторами без опыта реферирования, а также 25 рефератов, сгенерированных программой автоматического реферирования. Приведены частотные правки ошибок, выявленных на этапе постредактирования в рукописях информативных рефератов. Показаны степень компрессии текстов-источников, жанровое разнообразие, связность и целостность вторичных текстов, корректность фактологической информации на выходе, а также субъективность и оценочность при разных способах реферирования.*

**Ключевые слова:** реферат, вторичный текст, смысловое сжатие, компрессия текста, постредактирование, когезия, когерентность, автоматическая система реферирования

DOI: 10.36535/0548-0019-2021-06-4

### ВВЕДЕНИЕ

Ежедневно незаметно для себя мы часто прибегаем к реферированию и имеем дело с рефератами: в школе – в процессе выполнения учебного задания; в вузе – при подготовке к экзаменам; на досуге, выбирая книгу для чтения, мы обращаемся к аннотации на обороте титула; в кино читаем синопсис, выбирая фильм для просмотра и т.п. В научной деятельности навык реферирования является обязательным, так как любое научное произведение предполагает раздел «История вопроса».

«Искусство реферирования как извлечения наиболее существенной информации из исходного текста и ее последующей презентации во вторичном (реферативном) тексте стало неотъемлемой частью современной жизни. Тем не менее, реферирование имеет многовековую историю, уходящую корнями во времена Шумерской цивилизации, когда писцы создавали краткие записи на глиняных табличках. Далее, в эллинистический период и в эпоху распада Римской империи на пергаменте и свитках также делались краткие тезисы и описания докумен-

тов. Тогда (предположительно) и возник термин *abstractus*» [1, с. 127] (реферат).

В России первым теоретиком реферирования по праву считается М.В. Ломоносов, сформулировавший основные правила для журналистов-референтов: «Всякий, кто берет на себя труд осведомлять публику о том, что содержится в новых сочинениях, должен, прежде всего, взвесить свои силы. Ведь он затевает трудную и очень сложную работу, при которой приходится докладывать не об обыкновенных вещах и не просто об общих местах, но схватывать то новое и существенное, что заключается в произведениях, создаваемых часто величайшими людьми. Высказывать при этом неточные и безвкусные суждения – значит сделать себя предметом презрения и насмешки; это значит уподобиться карлику, который хотел бы поднять горы» [2, с. 7, цит. по: 1, с. 128].

**Понятие «реферат», виды и функции реферата.** Согласно ГОСТ Р 7.0.99-2018 [3] реферат – это краткое точное изложение содержания первичного документа в текстовой форме, включающее основные фактические сведения и выводы, без дополнительной интерпретации или критических замечаний автора реферата. Выделяют следующие виды рефератов.

\* Статья подготовлена в рамках государственного задания ИНИОН РАН по теме «Лингвокультурные аспекты цивилизационных противоречий».

По форме изложения:

- *информативный* (реферат-конспект) – составляется таким образом, чтобы при его использовании не возникало необходимости обращаться к исходному тексту;

- *индикативный* (реферат-резюме) – представляет максимально краткое изложение основного содержания, что сближает его с аннотацией;

- *информативно-индикативный* (комбинированный) – отражает в сокращенном виде основные положения и аспекты первичного документа.

По количеству реферируемых документов:

- *монографический* – составляется по одному источнику;

- *сводный* (обзорный) – написан на основе двух и более исходных документов.

По составителю:

- *авторский* (авторреферат) – подготовлен автором первичного документа;

- *неавторский* – составлен референтом;

- *машинный* (автоматический) – сделан с помощью компьютерной программы.

Реферат выполняет следующие функции:

- позволяет установить основное содержание документа, определить его релевантность и решить, следует ли обращаться к его полному тексту;

- предоставляет информацию о документе и устраняет необходимость обращаться к полному его тексту в случае, если документ имеет для читателя второстепенный интерес;

- используется в информационных, в том числе автоматизированных, системах.

**Реферирование как компетенция.** Реферирование можно рассматривать как процесс сжатия текстового документа с целью получения вторичного<sup>1</sup> документа – реферата, в котором сохраняется смысл оригинала. В процессе реферирования требуются и осваиваются (развиваются) следующие навыки:

- *смыслового анализа* текста (адекватное извлечение информации, выстраивание логики изложения материала, представленного в источнике);

- *оценки информации с позиций ее значимости* (выделение содержательных ключевых элементов, передающих тематическую основу источника);

- *смыслового свертывания (компрессии)* текста источника (сокращение текста без потери значимой актуальной информации).

Смысловое свертывание (компрессия) предусматривает устранение избыточной, второстепенной информации в тексте. В целом компрессию можно свести к перефразированию и обобщению информации источника при условии сохранения смысла исходного текста<sup>2</sup>.

<sup>1</sup> «Вторичный текст часто называют вербальным образованием, которое порождается в результате развертывания определенных ментальных структур (ключевых слов, темы и т.д.), являющихся результатом осмысления и понимания первичного текста и отражающих в свернутом виде основное его содержание» [4, цит. по 5, с. 26].

<sup>2</sup> О способах компрессии в научном тексте подробнее см. [7–11].

Смысловое свертывание можно осуществлять на разных языковых уровнях [6]:

- лексическое перефразирование построено на синонимических заменах (*этап* → *период*), местоименных заменах (*тенденции* → *они*), заменах глагольных словосочетаний глаголами (*оказать влияние* → *влиять*);

- грамматическое перефразирование с использованием синонимов (замена сложного предложения простым, придаточного – причастным оборотом);

- лексико-грамматическое перефразирование с использованием конверсии (*обстоятельства не позволили организации... → организация не сумела...*);

- логико-смысловое перефразирование – сопоставление двух понятий, обнаруживающих сходство при определенной степени абстрагирования их смысла, например:

Будущее *безопасного ненасильственного мира* диктует необходимость *совершенствования механизмов не только урегулирования, но и предотвращения споров.*

→

Для укрепления *международной безопасности* необходимо *повысить эффективность миссий по поддержке мира.*

Следует отметить, что связность текста реферата достигается, помимо прочего, и за счет метаязыковых средств языка науки, использование которых задает структуру текста реферата (табл. 1). Подробнее о метаязыковых средствах оформления текста реферата см. [12, с. 53-61].

**Методы реферирования текстов.** Существующие методы реферирования можно разделить на две большие группы: (1) основанные на анкетировании и (2) в основе которых лежит анализ смысловой структуры первоисточника [1, с. 128].

Первая группа методов основана на схожести с анкетированием – ответами на «вопросы», которые помогают найти в тексте первоисточника соответствующую информацию, в неё входят фасетное и поаспектное реферирование. При фасетном реферировании [13] пользуются элементарными логическими сообщениями, создающими табличный (структурированный) реферат. Поаспектное реферирование [14] предполагает семантический анализ текста с помощью разбиения на семантические аспекты. Вторая группа – основана на логико-психологическом анализе [15] реферируемого текста и информационно-классификационном реферировании, в основе которого лежит критерий точности информирования [1, с. 129].

**Методы автоматического реферирования текстов.** «Во второй половине XX века ввиду стремительного роста значения вторичных документов в информационном обслуживании и в связи с появлением реферативных журналов была осознана необходимость создания теоретических основ и методик процесса реферирования с целью его совершенствования и улучшения качества рефератов» [1, с. 128]. Одной из основных задач компьютерной обработки текстов стала формализация методов реферирования<sup>3</sup>.

<sup>3</sup> Сегодня основными издателями реферативных журналов являются ВИНТИ РАН (техническая и естественно-научная информация) и ИНИОН РАН (научная информация по социально-гуманитарному профилю).

## Некоторые языковые средства представления текста реферата

Структурный элемент	Языковые средства
Описание содержания источника	<i>В статье (работе, монографии, сборнике) рассматривается / анализируется / освещается / ставится вопрос о...</i>
Изложение позиции автора источника	<i>Автор считает / полагает / утверждает, что... / выступает в защиту... / убежден в... / подвергает критике... / по мнению автора...</i>
Структура источника	<i>В статье (работе, монографии, сборнике) можно выделить ... частей... / статью (работу, монографию, сборник) можно разделить на несколько частей...</i>
Связки между частями реферата	<i>Вначале речь идет о... / Затем рассказывается... / Далее автор пишет о... / Раскрывается также...</i>
Заключение	<i>В заключение подчеркивается (говорится), что... / В заключение автор приходит к выводу, что... / В ходе анализа было выявлено...</i>
Адресат реферата	<i>Текст адресован... / Материал представляет интерес для... / будет полезен для... / заинтересует...</i>

Исследования по автоматическому реферированию текстов начались почти одновременно с работами по машинному переводу (середина 50-х гг. XX в.). Наибольшую популярность получили метод экстракции (извлечения предложений / sentence extraction / квазиреферирования) и метод абстракции (извлечения содержания / content extraction) [16, с. 60].

При *экстракции* из текста извлекаются наиболее важные информационные блоки (предложения, абзацы). Такой подход характеризуется наличием оценочной функции (*scoring function*) информационного блока (предложения), сложность вычисления которой постоянно росла с развитием методов и с учетом расширения списка количественных и качественных характеристик текста [16, с. 61]. Недостатком данного подхода можно назвать низкую связность текста реферата на выходе.

При *абстракции* на основе текста-источника генерируется текст реферата, содержательно обобщающий первичный документ или документы. Процесс включает три обязательных этапа: (1) анализ исходного текста с генерацией внутреннего представления, (2) семантическое сжатие внутреннего представления и (3) синтез реферата [16].

Специалисты доказали, что автоматическое реферирование лучше всего работает на хорошо структурированных документах, таких как новости, научные и публицистические статьи, доклады [17].

## МЕТОДИКА ИССЛЕДОВАНИЯ

Изучение различных аспектов реферирования проводилось нами в два этапа: на первом – анализировались первичные и скорректированные информативные рефераты, выполненные разными группами испытуемых, для которых создание реферата было

основной задачей их работы с источником (цель – публикация текста реферата в специализированном журнале<sup>4</sup>); на втором – анализировались индикативные рефераты, выполненные испытуемыми в качестве учебно-исследовательского задания, а также рефераты, полученные при помощи программы автоматического реферирования TextAnalyst (<http://www.analyst.ru/index.php?lang=eng&dir=content/products/&id=ta>).

Рабочая гипотеза нашего исследования заключалась в предположении, что первичный<sup>5</sup> текст реферата потребует постредактирования вне зависимости от компетенции автора текста (референта) или от типа и сложности реферируемого текста-источника.

Для большего единообразия в качестве источников реферирования в исследовании использовались хорошо структурированные тексты: на первом этапе – научные произведения (монографии, статьи по лингвистике), на втором – публицистические произведения (статьи и интервью институциональных печатных СМИ). В табл. 2 представлены характеристики текстов-источников, задействованных в нашем исследовании.

На первом этапе исследования авторами рефератов выступили состоявшиеся ученые-лингвисты (доктора и кандидаты филологических наук) с опытом реферативной работы (n = 5), а также начинающие исследователи (студенты выпускных курсов бакалавриата и магистратуры, аспиранты) без опыта реферативной работы (n = 10) – всего 15 человек.

<sup>4</sup> Рефераты прошли экспертизу и опубликованы в журнале «Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6. Языкознание». – URL: <https://www.elibrary.ru/contents.asp?titleid=2783>

<sup>5</sup> На первом этапе исследования первичным текстом реферата выступала рукопись, передаваемая в редакцию журнала, на втором этапе – текст реферата, полученный от испытуемых.

## Характеристики текстов-источников

Характеристика публикации	Тексты-источники	
	первого этапа исследования	второго этапа исследования
Тип	монографии на русском языке (1 ед.), монографии на иностранном языке (2 ед.), статьи на русском языке (25 ед.), статьи на иностранном языке (45 ед.)	аналитические статьи на русском языке (14 ед.), интервью на русском языке (13 ед.)
Формат	печатные издания	электронные версии газет «Аргументы и Факты», «Новая газета»
Тематика	исследования по лингвистике	культура, общество, политика
Объем	от 10800 до 2235600 знаков (средний объем одного источника = 230625 знаков)	от 311 до 1285 слов (средний объем одной публикации = 731 слово)

На втором этапе в эксперименте приняло участие 15 обучающихся по программам высшего и дополнительного образования в Московском государственном лингвистическом университете (без опыта реферативной работы) в возрасте от 18 до 46 лет. Все участники (на первом и втором этапах исследования) были носителями русской лингвокультуры и дали согласие на участие в эксперименте.

На первом этапе в фокусе внимания были ошибки, зафиксированные в процессе постредрактирования в рукописях референтов; на втором – разнообразие жанров, в которых были представлены краткие изложения текстов-источников. В обоих случаях фиксировалась степень компрессии исходного текста-источника.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Согласно данным, полученным на первом этапе исследования (табл. 3), средний показатель компрессии, т. е. смыслового свертывания при реферировании текста-источника (научного текста), составляет 82% (при написании реферата – 76%, при написании сводного реферата – 86%). Состоявшиеся ученые-лингвисты сжимают тексты-источники в среднем на 85%, начинающие – на 78%.

При постредрактировании рукописи (сводного) реферата (67% исследованных рукописей) наблюдается тенденция к увеличению объема текста в среднем на 8,4% (для текстов, подготовленных учеными-лингвистами, – на 6%, начинающими лингвистами – на 12%).

Рассмотрим подробнее правки, которые необходимы в текстах (сводных) рефератов источников на иностранном языке.

Наиболее частой является правка, связанная с поддержанием связности и целостности текста в случаях, когда референт вместо реферирования переводит текст источника (примеры № 1-2 в табл. 4). Начинающие референты порой увлекаются детализацией вместо обобщения (№ 3). Для лингвистических исследований важны языковые примеры (№ 4), перевод иноязычного языкового материала на русский язык (№ 5), указание русскоязычных эквивалентов понятий и терминов (№ 6), а также названия источников на языке оригинала (№ 7) – всё это упрощает понимание выводов, представленных в тексте реферата.

Наиболее частотной ошибкой, противоречащей сущности реферирования (краткому изложению

содержания источника без интерпретации), является использование в тексте реферата оценочных высказываний<sup>6</sup>.

В приведенных примерах зачеркиванием отмечены фрагменты текста, которые были сняты при постредрактировании:

- ~~«Важной находкой данного исследования стало то, что И жертвы, и очевидцы кибербуллинга осуждали агрессора за анонимность».~~
- ~~«Заслуживает внимания тот факт, что, как Автором монографии установлено, что...»~~
- ~~«...количество дикторов в каждой группе сбалансировано, и, что очень важно, представлен переходный этап от молодого к старшему возрасту...»~~

В силу того, что реферат – это вторичный документ, одной из функций которого является осведомление об источнике с последующей возможностью обращения к нему, в реферате рекомендуется давать отсылки к страницам первоисточника, чем референты нередко пренебрегают, даже приводя цитату (закрывая текст), например: «материал позволил использовать разные способы анализа ассоциативных и семантических полей (построение моделей концепта, семантического геистальта, ...)» – здесь необходимо указать страницы источника.

В процессе постредрактирования нередко обнаруживаются ошибки, связанные с лексической сочетаемостью, и стилистические ошибки, что свидетельствует об отсутствии у референта навыка литературного редактирования. Далее в примерах зачеркиванием отмечены фрагменты текста, которые были сняты в процессе постредрактирования, подчеркиванием – внесенные исправления:

- ~~«...Появление всемирной сети Интернет да-ло толчок способствовало развитию коммуникации, торговле и науке экономическому и научному прогрессу, в целом, и развитию лингвистического знания, в частности. Лингвистика, как и другие науки, извлекала для себя пользу из новых инструментов и вычислительных мощностей».~~

<sup>6</sup> Подобные ошибки часто встречаются и в рефератах опытных исследователей, что связано с неким автоматизмом, выработанным при рецензировании статей, выпускных квалификационных работ студентов, диссертаций.

- «Современные тенденции в языке говорят Результаты анализа актуальных контекстов употребления слова «animal» свидетельствуют о том, что...»
- «Поэтому В связи с этим было важно определить роль...»
- «...с точки зрения с позиций языковых способов представления психологически актуального содержания...»
- «Оказалось Было обнаружено, что...»
- «За основу сегментации предложения положены взяты законы генеративной грамматики.»

При реферировании иноязычных источников в текстах, созданных начинающими учеными – участниками первого этапа исследования, обнаруживается низкий уровень осведомленности о соотношении терминологии в русском и иностранном языках, например:

- «...обработка естественного языка, всеобъемлющее извлечение информации, интеллектуальный

семантический анализ текста и другие языковые технологии»

- «...корпусов британского английского и американского вариантов английского языка...»
- «авторы отмечают, что все компьютерные науки..., имеют дело с человеческим естественным языком.»

Иногда референты злоупотребляют безличными конструкциями, используя их там, где этого не требуется, например:

- «Автор статьи уделяет большое В статье А. Гонсалес Фернандес в большей мере уделено внимание первому подходу...»
- «Среди преимуществ этого подхода выделяется Автор выделяет следующие преимущества данного подхода...»
- «Кроме того, в этой главе в В научный обиход вводятся и новые имена писателей, и произведений...»

Таблица 3

### Доля сжатия информации в тексте (сводного) реферата

Жанр научного текста-источника <sup>1</sup>	Объем <sup>2</sup> текста-источника (кол-во знаков)	Жанр вторичного текста <sup>3</sup>	Объем текста реферата в рукописи (кол-во знаков)	Компрессия, %	Объем текста реферата после редактирования (кол-во знаков)	Опыт реферирования (есть / нет)
4 С на рус.	95400	СР	11780	88	12164	нет
КМ болг.	2235600	Р	17131	86	16609	есть
С англ.	113400	Р	24192	79	24509	нет
11 С на англ.	198000	СР	14425	93	14426	есть
3 С на рус.	471600	СР	30171	94	41545	есть
М нем.	255600	Р	8250	97	8271	есть
6 С на рус.	235800	Р	14016	94	15032	есть
3 С на англ.	100800	СР	11665	88	11142	нет
3 С на англ.	86400	СР	12710	85	12876	нет
КМ рус.	838800	Р	18083	98	18783	есть
2 С на 2рус., 2 на англ.	127800	СР	14337	89	14568	есть
3 С на англ.	108000	СР	10424	90	10541	есть
С англ.	43200	Р	4245	90	6952	нет
С англ.	25200	Р	14010	44	8576	нет
С англ.	10800	Р	6627	49	6844	есть
5 С на рус.	158400	СР	12116	92	11934	есть
С англ.	32400	Р	6534	80	6376	есть
1 С на 1 рус., 1 на англ.	36000	СР	10239	72	10581	есть
С англ.	14400	Р	7801	46	7155	нет
3 С на англ.	46800	СР	12651	73	12971	есть
9 С на англ.	142200	СР	12010	92	11826	нет
2 С на англ.	27000	СР	4878	82	4749	нет
2 С на англ.	45000	СР	10384	77	10533	нет
4 С на рус.	86400	СР	7402	91	7587	нет

Примечания:

<sup>1</sup> М = монография, КМ = коллективная монография, С = статья.

<sup>2</sup> Объем рассчитывался исходя из правила «одна страница печатного текста = 1800 знаков с пробелами».

<sup>3</sup> Р = реферат, СР = сводный (обзорный) реферат.

## Примеры корректуры текста (сводного) реферата

№ примера	Первоначальная рукопись реферата	Скорректированный текст
1	<i>Группа исследователей под руководством Р. Юнг провели эксперимент, в котором они поставили задачу...</i>	<i>В реферируемой работе описывается эксперимент, задачей которого является...</i>
2	<i>Для проведения этого эксперимента была выбрана площадка для вопросно-ответного общения. Данный сайт был выбран по причине того, что более половины его пользователей на момент проведения эксперимента были несовершеннолетними.</i>	<i>Объектом исследования стали дискурсивные практики пользователей сайта ask.fm, более половины пользователей которого на момент проведения эксперимента были несовершеннолетними.</i>
3	<i>Исследователи постарались выявить закономерности, связывающие агрессивные комментарии и ответы на них. Таким образом, им удалось обнаружить, что...</i>	<i>(Первое предложение удалено)*. При анализе языкового материала удалось обнаружить, что...</i>
4	<i>В русском и английском вариантах идея скорости передается с помощью лексического повтора (приведите языковой материал, если он есть в статье, и переведите его на русский язык)*. В немецком языке, для которого такая конструкция не характерна, используется причастный оборот (приведите языковой материал, если он есть в статье, и переведите его на русский язык)*.</i>	<i>В русском и английском вариантах идея скорости передается с помощью лексического повтора: «Папа пишет быстро-быстро» и «Daddy writes very, very quickly» (2, с. 70). В немецком языке, для которого такая конструкция не характерна, используется причастный оборот «Papa schreibt schnell, in fliegender Hast» (2, с. 70).</i>
5	<i>«Вещество / субстанция» – самая продуктивная в английском политическом дискурсе метафорическая модель (пример: «solid support of Republicans» (3, с. 52)) (дайте перевод на русский язык)*...</i>	<i>... (пример: «solid support of Republicans» / «Твердая поддержка республиканцев» (3, с. 52))<sup>1</sup>. В сноске: <sup>1</sup> Здесь и далее пер. с англ. наш. – Прим. реф.</i>
6	<i>Термин «шифтеры валентности» (дайте английский эквивалент)* используется для...</i>	<i>Термин «шифтеры валентности» (valence shifters) используется для...</i>
7	<i>«Корпус высказываний на испанском языке, предназначенных для анализа тональности на материале рекламных брендов в Интернете» (дайте название корпуса на языке источника)*</i>	<i>«Корпус высказываний на испанском языке, предназначенных для анализа тональности на материале рекламных брендов в Интернете» (Spanish Corpus for Sentiment Analysis Towards Brands)</i>

\* В скобках обычным шрифтом указаны замечания редактора.

Та же ситуация наблюдается с перегруженностью текста реферата вводными конструкциями, например:

- ~~«Касается первого вопроса: все Анализируемые в статье словари...»~~
- ~~«Таким образом, можно сделать вывод о том, что...»~~
- ~~«Таким образом, в очередной раз автором подтверждается предположение о...»~~
- ~~«Основными выводами в данной работе предстают В ходе исследования авторы пришли к следующим выводам. Во первых, проведенное исследование в рамках единого проекта 102 образцов американской диалогической речи доказало, что Оптимизация просодических средств сохранения ритма проходит...»~~

Обратная ошибка – это опущение значимых конкретизаторов, например (дополнения отмечены подчеркиванием):

- ~~«Возросший интерес к исследованию категорий субъективности, оценки и мнения может быть рассмотрен как аффективное направление в области философии...».~~
- ~~«В исследуемой проблематике особое значение имеют описания образов в языке, ...».~~
- ~~«Определение по тексту намерения тоже является одной из задач классификации».~~
- ~~«Лингвистическая тематика, нашедшая отражение в докладах, представленных на конференции...».~~

**Доля сжатия информации во вторичном тексте при ручном  
и автоматическом реферировании**

Тексты СМИ <sup>1</sup>	Кол-во слов в исходном тексте	Кол-во слов во вторичном тексте (TextAnalyst)	Компрессия, %	Кол-во слов во вторичном тексте (испытываемые)	Компрессия, %
АС1	452	74	84	135	70
АС2	548	39	93	88	84
АС3	702	143	80	109	84
АС4	808	99	88	134	83
АС5	472	59	88	121	74
АС6	847	170	80	126	85
АС7	522	71	86	79	85
АС8	537	41	92	120	78
АС9	1285	172	87	127	90
АС10	493	152	69	51	90
АС11	852	139	84	58	93
АС12	471	25	95	140	70
АС13	311	72	77	56	82
АС14	803	100	88	39	95
И1	890	18	98	53	94
И2	1043	– <sup>2</sup>	–	28	97
И3	984	34	97	30	97
И4	567	–	–	54	90
И5	759	61	92	130	83
И6	885	49	94	66	93
И7	905	39	96	49	95
И8	649	28	96	81	88
И9	810	112	86	15	98
И10	600	26	96	59	90
И11	829	22	97	79	90
И12	1105	88	92	95	91
И13	603	20	97	133	78

<sup>1</sup> АС = аналитическая статья, И = интервью.

<sup>2</sup> Прочерк означает, что программа не сгенерировала вторичный текст.

Таблица 6

**Оценка целостности вторичных текстов, %**

Создатель вторичного текста	Аналитическая статья		Интервью	
	Обладают целостностью	Не обладают целостностью	Обладают целостностью	Не обладают целостностью
TextAnalyst	64	36	55	45
Испытуемые	92	8	95	5

Таблица 7

**Оценка связности вторичных текстов, %**

Создатель вторичного текста	Аналитическая статья		Интервью	
	Обладают связностью	Не обладают связностью	Обладают связностью	Не обладают связностью
TextAnalyst	36	64	55	45
Испытуемые	96	4	90	10

Общая для всех референтов трудность связана с оформлением библиографических описаний источников по ГОСТ Р 7.0.100-2018 [18].

**На втором этапе исследования** были получены следующие результаты сжатия (компрессии) текста-источника (табл. 5) при ручном и автоматическом (с помощью программы TextAnalyst) реферировании.

Средние показатели компрессии в неавторском (выполнен испытуемым) и машинном (сгенерирован программой автоматического реферирования TextAnalyst) вторичных текстах сопоставимы: программа автоматического реферирования сжимает текст источника в среднем на 89% (тексты аналитических статей – на 97%, тексты интервью – на 45%), испытуемые при реферировании – в среднем на 87% (аналитические статьи – на 83%, интервью – на 91%).

Существенная разница в доле сжатия информации текста-источника наблюдается при обработке программой TextAnalyst текстов интервью. Это можно объяснить однообразием структуры текста-источника (вопросно-ответные реплики) по сравнению со структурным разнообразием аналитической статьи (введение, анализ данных, разбивка на подтемы, критика, обобщение и т.д.), что воспринимается системой автоматического реферирования как значимый фактор при анализе текста-источника.

Для определения соответствия содержания текстов рефератов содержанию текстов-источников в ходе нашего исследования необходимо было ответить на следующие вопросы:

Какому жанру соответствует вторичный текст (реферат ли это)?

Обладает ли полученный текст целостностью и связностью?

Отражает ли полученный текст фактологическую информацию исходного текста?

Присутствует ли в изложении субъективность и оценочность?

**Жанр вторичного текста.** Несмотря на то, что исследовательское задание формулировалось как задача представить краткое изложение текста-источника, т. е. написать реферат, в процессе создания вторичного документа на основе исходного были получены тексты разных жанров – реферат, эссе, статья, интервью.

Применительно к текстам-источникам в виде аналитических статей формально программа автоматического реферирования справилась с задачей лучше, чем испытуемые референты (57% текстов в жанре реферата сгенерировано системой и 42% создано испытуемыми). Следует отметить, что при кратком изложении содержания источника испытуемые вносят собственные суждения, что соответствует жанру эссе (34%).

Автоматическая система испытывает существенные затруднения при реферировании интервью, о чем свидетельствует большая доля случаев, в которых программа не справилась с задачей (15%), а также доля текстов, жанр которых не удалось определить (31%), т. е. они состояли из элементов различных жанров. Для программы TextAnalyst затруднения вызвало и переструктурирование текста из жанра интервью в жанр реферата (31% вторичных текстов сохранил структуру жанра интервью).

Было выявлено, что и испытуемые – составители рефератов не справились с задачей, сжимая интервью в эссе (62%) с элементами собственного рассуждения, оценочности и интерпретации.

**Целостность и связность вторичных текстов.** Под целостностью текста мы понимаем его когерентность, т. е. логико-семантическую, грамматическую и стилистическую соотнесенность и взаимозависимость элементов текста (слов, предложений и т.д.) в совокупности с его содержательной связностью [19, 20], а под связностью – когезию, т. е. лексическую и грамматическую связность, которая объединяет отдельные предложения в текст [21]. В табл. 6 и 7 приведены результаты оценки целостности и связности вторичных текстов, созданных испытуемыми референтами и программой TextAnalyst.

Из табл. 6 видно, что большинство текстов, написанных испытуемыми, обладает целостностью (92% – для аналитических статей, 95% – для интервью). Тексты, сгенерированные программой TextAnalyst, в целом являются когерентными (64% – для аналитических статей, 55% – для интервью), хотя количество вторичных текстов, не обладающих целостностью, в случае автоматического реферирования значительно выше (45%).

В отношении связности вторичных текстов наблюдается похожая картина: почти все тексты, созданные испытуемыми, характеризуются связностью (96% – для аналитических статей, 90% – для интервью); больше половины текстов, сгенерированных автоматически, не обладает связностью (64% – для аналитических статей, 45% – для интервью).

**Фактологическая информация во вторичных текстах.** В целом, и при неавторском (испытуемые референты), и при машинном реферировании (TextAnalyst) аналитических статей фактологическая информация была передана верно (57%). При этом испытуемые сделали больше фактологических ошибок (29%), чем программа TextAnalyst, которая допустила больше неточностей (43%).

Автоматическая система справляется гораздо хуже с задачей реферирования интервью (лишь 9% текстов, в которых фактологическая информация передана верно) по сравнению с результатами реферирования испытуемыми (73% текстов без ошибок и неточностей).

**Субъективность и оценочность при реферировании текста.** Под оценочным и/или субъективным высказыванием мы понимаем выражение, устанавливающее абсолютную или сравнительную ценность какого-либо объекта, дающее ему оценку [22]. К субъективным мы также относим внутреннюю речь и несобственно прямую речь автора.

Результаты анализа вторичных текстов свидетельствуют о том, что и при автоматическом реферировании в текстах обнаруживаются признаки субъективности и оценочности (28%). Во вторичных текстах, подготовленных испытуемыми референтами, субъективность проявляется через выражение их собственного отношения к тексту (например, с помощью вводных слов *вероятно, возможно, кажется*), в то время как во вторичных текстах, сгенерированных программой TextAnalyst, субъективные реплики переносятся из текста-источника.

## ЗАКЛЮЧЕНИЕ

При сопоставлении данных первого и второго этапов исследования (результатов ручного и автоматического реферирования) можно сделать вывод о том, что и при ручном, и при машинном реферировании доля компрессии информации колеблется в пределах 82–89%. Различия наблюдаются при учете таких факторов, как компетенция и опыт референта, тип реферируемого источника, прагматическая задача реферирования. Полученные данные подтверждают рабочую гипотезу исследования о том, что вне зависимости от указанных факторов, первоначальные рукописи рефератов нуждаются в постредактировании. Рефераты начинающих исследователей, хотя и отличаются меньшей долей сжатия информации, требуют большего редактирования. При этом сокращение объема текста касается избыточных вводных конструкций, несущественной информации, а расширение – уточняющей информации, обусловленной требованиями к формату вторичного текста (например, ориентация реферата на русскоговорящего читателя, что требует перевода на русский язык материала, приводимого в реферате на иностранном языке в качестве иллюстраций).

При кратком изложении испытуемыми текст-источника в учебно-исследовательских условиях наблюдается смещение фокуса их внимания с формы вторичного текста, что проявляется в жанровом разнообразии вторичных документов (эссе, статья, гибридный жанр), отличных от жанра реферата, иногда совпадающих с жанром текста-источника (интервью – интервью, статья – статья). Та же тенденция наблюдается во вторичных текстах, созданных программой автоматического реферирования TextAnalyst, что, вероятно, связано с техническими ограничениями семантического анализатора этой программы.

Таким образом, проведенное исследование подтвердило утверждение о зависимости качества автоматического реферирования от типа исходного текста. Так, при автоматическом реферировании текстов интервью наблюдалось наибольшее число неудач, неточностей и ошибок, особенно отчетливо это обнаруживалось в передаче фактологической информации и целостности вторичного текста.

## СПИСОК ЛИТЕРАТУРЫ

1. Нестерова Н.М., Герте Н.А. Реферирование как способ извлечения и представления основного содержания текста // Вестник Пермского университета. Российская и зарубежная филология. – 2013. – Вып. 4(24). – С. 127-132.
2. Гречихин А.А., Здоров И.Г., Соловьев В.И. Жанры информационной литературы: Обзор. Реферат. – Москва: Книга, 1983. – 320 с.
3. ГОСТ Р 7.0.99-2018 (ИСО 214:1976) Система стандартов по информации, библиотечному и издательскому делу. Реферат и аннотация. Общие требования, 2019. – URL: <http://docs.cntd.ru/document/1200160041> (дата обращения 16.01.2021).
4. Новиков А.И., Сунцова Н. Концептуальная модель порождения вторичного текста // Обработка текста и когнитивные технологии. – 1999. – № 3. – С. 158-166.
5. Левкова Л.Ю. Обучение созданию вторичных текстов // Среднее профессиональное образование. – 2010. – № 7. – С. 24-27.
6. Маслова Н.М. Основы работы над текстом. – Москва: МГЛУ, 1999.
7. Зимовец Н.В., Лелюк Г.К. Синтаксические способы компрессии // Современные тенденции развития науки и технологий. – 2017. – № 1-6. – С. 45-47.
8. Ичкинеева, Л.А. Тема и микротема как инструменты компрессии информации в тексте // Социо- и психолингвистические исследования. – 2017. – № 5. – С. 146-148.
9. Кудинова Т.А. Способы компрессии словообразовательной структуры терминологической словосочетаний (на материале английского подязыка биотехнологий) // Вестник Омского государственного педагогического университета. Гуманитарные исследования. – 2018. – № 1(18). – С. 63-66.
10. Плохая Е.Е. Способы языковой компрессии в энциклопедическом тексте // Профессиональная коммуникация: Актуальные вопросы лингвистики и методики. – 2018. – № 11. – С. 71-76.
11. Бредихин С.Н., Серебрякова С.В., Лиховид А.А. Способы компрессии когнитивной информации в научно-популярном тексте // Актуальные проблемы филологии и педагогической лингвистики. – 2019. – № 3. – С. 139-145.
12. Реферирование / сост. Н.К. Студенникова. – Самара: Самар. гос. техн. ун-т, 2010. – 63 с.
13. Жолкова А.И. К вопросу о реферировании научно-технической периодики в реферативном журнале ВИНТИ // Научно-техническая информация. Сер. 1. – 1970. – № 5. – С. 24-25.
14. Соловьев В.И. Поаспектный метод реферирования // Научно-техническая информация. Сер. 2. – 1971. – № 2. – С. 14-17.
15. Вейзе А.А. Реферирование текста. – Минск: Белорусский государственный университет, 1978. – 128 с.
16. Тарасов С.Д. Современные методы автоматического реферирования // Научно-технические ведомости СПбГПУ. – 2010. – № 6. – С. 59-74.
17. Dixit R.S., Apte S.S. Improvement of text summarization using fuzzy logic based method // IOSR Journal of Computer Engineering (IOSRJC). – 2012. – Vol. 5(6). – P. 5-10.
18. ГОСТ Р 7.0.100-2018 «Библиографическая запись. Библиографическое описание», 2019. – URL: [https://www.rsl.ru/photo/!ORS/5-PROFESSIONALAM/7\\_sibid/%D0%93%D0%9E%D0%A1%D0%A2\\_%D0%A0\\_7\\_0\\_100\\_2018\\_1204.pdf](https://www.rsl.ru/photo/!ORS/5-PROFESSIONALAM/7_sibid/%D0%93%D0%9E%D0%A1%D0%A2_%D0%A0_7_0_100_2018_1204.pdf) (дата обращения 16.01.2021).
19. Бобрикова Е.Н. Средства связности текста в литературе «потока сознания» (на материале романа Джеймса Джойса «Улисс»). – Ростов-на-Дону, 2008. – URL: <https://www.elibrary.ru/item.asp?id=30333861> (дата обращения 16.01.2021).
20. Стариннова Т.Б., Круглякова Г.В. Проявление категорий целостности и связности в публицистических текстах // Вестник ВУиТ. – 2016. – № 1. – URL: <https://cyberleninka.ru/article/n/proyavlenie>

kategoriy-tselostnosti-i-svyaznosti-v-publitsisticheskikh-tekstakh (дата обращения 16.01.2021).

21. Величко М.А. Когезия и когерентность: особенности разграничения и определения понятий // Вестник Адыгейского государственного университета. Сер. 2: Филология и искусствоведение. – 2016. – № 2(177). – URL: <https://cyberleninka.ru/article/n/kogeziya-i-kogerentnost-osobnosti-razgranicheniya-i-opredeleniya-ponyatiy> (дата обращения 16.01.2021).
22. Оценочное высказывание // Изба-читальня. – URL: <https://www.chitalnya.ru/tolkslovar/o/o9281.php> (дата обращения 16.01.2021).

*Материал поступил в редакцию 28.01.21.*

#### **Сведения об авторах**

**КОМАЛОВА Лилия Ряшитовна** – доктор филологических наук, ведущий научный сотрудник отдела языкознания Института научной информации по общественным наукам РАН; профессор кафедры прикладной и экспериментальной лингвистики Московского государственного лингвистического университета.  
e-mail: komalova@inion.ru

**МАЙОРОВА Екатерина Владимировна** – младший научный сотрудник отдела языкознания Института научной информации по общественным наукам РАН.  
e-mail: mayorovacatherine@gmail.com

**ДЛЯ ЗАМЕТОК**

**ДЛЯ ЗАМЕТОК**