



DOI: <https://doi.org/10.15688/jvolsu2.2025.1.6>

UDC 81'42:659.123.4
LBC 81.055.1

Submitted: 12.02.2024
Accepted: 21.10.2024

ALGORITHMIC PROCEDURES OF IDENTIFYING ADVERTISEMENT TEXTS IN MASS MEDIA DISCOURSE

Mikhail V. Kamensky

North-Caucasus Federal University, Stavropol, Russia

Sergey N. Bredikhin

North-Caucasus Federal University, Stavropol, Russia

Abstract. The article presents an algorithm of identifying advertisement blocks in mass media content and determining the type of the given text as either an advertisement or an informative text, which is enabled through automation with the aid of intellectual semantic and syntactic analysis systems. The GATE corpus manager is used as the development environment for the algorithm, and the ANNIE Gazetteer, JAPE Transducer, and Java Regexp Annotator are used as the principal processing resources for the presented algorithm. The use of ANNIE Gazetteer enables the automated identification of the most common lexical units typical of advertisements, as well as various lexical and syntactic markers of the advertisement content. The JAPE Transducer technology enables the development of an algorithm aimed at identifying an array of lexical and syntactic means of psychological influence. Identification of lexical repetitions of proper nouns is performed using a regular expression for the Java Regexp Annotator processing resource. The list of tokens used as advertisement content markers is identified and described. It is noted that lexical and syntactic means of manipulative influence dominate in advertisement texts. Research findings indicate a significant difference in the search results ratio between advertisements and informative texts when advertisements are identified automatically with the aid of formal markers. This proves the effectiveness of natural language processing systems in identifying messages with explicit and implicit advertisement content, determining the discursive type of media texts, and classifying them as either informative texts or advertisements.

Key words: automated analysis system, semantic-and-syntactic analyzer, mass media, corpus analysis, manipulative discourse, advertisement content, automated search algorithms.

Citation. Kamensky M.V., Bredikhin S.N. Algorithmic Procedures of Identifying Advertisement Texts in Mass Media Discourse. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], 2025, vol. 24, no. 1, pp. 64-78. (in Russian). DOI: <https://doi.org/10.15688/jvolsu2.2025.1.6>

УДК 81'42:659.123.4
ББК 81.055.1

Дата поступления статьи: 12.02.2024
Дата принятия статьи: 21.10.2024

АЛГОРИТМИЧЕСКИЕ ПРОЦЕДУРЫ ИДЕНТИФИКАЦИИ РЕКЛАМНЫХ ТЕКСТОВ В ДИСКУРСИВНОМ ПРОСТРАНСТВЕ СРЕДСТВ МАССОВОЙ ИНФОРМАЦИИ

Михаил Васильевич Каменский

Северо-Кавказский федеральный университет, г. Ставрополь, Россия

Сергей Николаевич Бредихин

Северо-Кавказский федеральный университет, г. Ставрополь, Россия

Аннотация. В статье представлен авторский алгоритм идентификации рекламного контента и определения классификационных признаков текстов рекламного / информационного характера в медийном пространстве на основе применения автоматизированных интеллектуальных систем семантико-синтаксическо-

го анализа. Для разработки алгоритмов автоматизированной идентификации рекламного текста в среде GATE применены технологии ANNIE Gazetteer, JAPE Transducer и Java Regexp Annotator. Технология ANNIE Gazetteer позволила разработать файлы для автоматизированной идентификации лексико-синтаксических репрезентантов рекламного контента, а также осуществить идентификацию наиболее частотных лексических единиц. С помощью технологии поиска JAPE Transducer реализована алгоритмическая процедура автоматизированной идентификации лексико-синтаксических средств психологического воздействия, отличающих рекламный текст. Идентификация лексических повторов имен собственных реализована с помощью регулярного выражения для анализатора Java Regexp Annotator. Сформирован перечень токенов, выступающих формальными вербализаторами рекламного контента. Установлено, что в рекламных текстах доминируют лексико-синтаксические приемы манипулятивного воздействия. Продемонстрировано наличие существенного различия в процентном соотношении совпадений поиска ко всему объему текста в рекламных и нерекламных текстах при идентификации рекламного контента на основе формальных маркеров. Доказана эффективность использования автоматизированных систем анализа в идентификации эксплицитных и имплицитных рекламных сообщений в медийных текстах и установления дискурсивной принадлежности текста, опубликованного в СМИ, с целью его классификации как информационного либо рекламного.

Ключевые слова: автоматическая система анализа, семантико-синтаксический анализатор, средства массовой информации, корпусный анализ, манипулятивный дискурс, рекламный контент, алгоритмы автоматического поиска.

Цитирование. Каменский М. В., Бредихин С. Н. Алгоритмические процедуры идентификации рекламных текстов в дискурсивном пространстве средств массовой информации // Вестник Волгоградского государственного университета. Серия 2, Языкознание. – 2025. – Т. 24, № 1. – С. 64–78. – DOI: <https://doi.org/10.15688/jvolsu2.2025.1.6>

Введение

Реклама как имманентный атрибут современного общества, выступающий способом дискурсивного взаимодействия между рекламодателем и потенциальным потребителем продуктов и услуг, обладает существенным прагматическим потенциалом в части оказания информативно-оперативного воздействия на широкую аудиторию потребителей через средства массовой информации. В дискурсивном пространстве СМИ наблюдается определенная гибридизация газетно-публицистического контента, при которой сугубо информационный посыл подается в тесной конвергенции с рекламным сообщением, либо же рекламное сообщение частично «скрывается» за общим информационным наполнением статьи.

В связи со сказанным особую актуальность приобретает проблема идентификации эксплицитных и имплицитных сообщений рекламного характера в медийных текстах, а также установления дискурсивной принадлежности текста, опубликованного в СМИ, с целью его классификации как информационного, либо рекламного.

Поскольку осложненной представляется делимитация именно ситуативной скры-

той рекламной суггестии, а перечень формальных маркеров, способных представлять собой признаковый ряд для алгоритмизации и формирования базовых вводных для автоматического распознавания контента рекламного характера, на сегодняшний день не сформирован, в данной статье предпринята попытка описания и классификации токенов, выступающих формальными вербализаторами рекламного контента, а также демонстрации практической эффективности задаваемых функциональных ограничений. На основе проведенного анализа возможно выработать принципы манипулятивной коммерциализации сообщений по аналогии с выделением «медиазации», как институциональной организации, призванной транслировать образцы и модели восприятия или поведения [Ильинова, Волкова, 2023, с. 20], существенно модифицирующие пространство целеполагания и волюнтативную сферу реципиента. Токен представляет собой единицу анализа, образуемую одночленным формальным вербализатором, который в свою очередь является нечленимым фонографическим элементом с доминирующей валентностью, то есть встречается чаще всего в контексте, содержащем информацию конкретной тематики, в нашем исследовании – это рекламный контент. Воз-

возможности такой идентификации существенно расширяются благодаря динамичному развитию функционала современных корпусных менеджеров, таких как GATE [GATE...] и LancsBox [Brezina, Platt, 2023], способных осуществлять массовую обработку текстовой информации и на основе четких логических критериев соответствия текста определенным шаблонам (например, посредством регулярных выражений или иных метаязыков описания структурно-содержательных компонентов искомым текстовым фрагментам), и с применением систем искусственного интеллекта (например, машинного обучения нейронных сетей).

Существенный речевоздействующий и манипулятивный потенциал рекламного дискурса актуализирует его исследование в общем контексте научного осмысления дискурсивных приемов манипуляции массовым сознанием. В данной связи и по причине проводимой сегодня в западноевропейском медиапространстве информационной войны против России целесообразной представляется разработка принципов автоматизированной идентификации и автоматизированного анализа формальных языковых маркеров рекламного дискурса с речевоздействующим потенциалом на примере английского языка. В силу того, что предлагаемые алгоритмические процедуры основаны на формальных логических критериях, они могут быть в дальнейшем адаптированы для идентификации рекламного контента в текстах на разноструктурных языках мира, в том числе – на русском.

Целью предпринятого исследования является разработка алгоритмических процедур для автоматизации идентификации рекламного текста на основе ряда формальных признаков языковых средств психологического воздействия, характерных для рекламного текста, на примере английского языка. Полагаем, что это позволит обеспечить возможность оптимизации проведения научных исследований в области автоматизированной классификации текстов и установления их дискурсивной принадлежности, а также в сфере реализации программных средств фильтрации рекламного контента в электронных версиях современных СМИ.

Материал и методы исследования

Отправной точкой настоящего исследования послужила классификация лексических и стилистических средств психологического воздействия в рекламном дискурсе и их формальных признаков в тексте [Каменский и др., 2016, с. 54]. Данная классификация включает установленные в предпринятых нами ранее научных изысканиях частотные экспликативные англоязычного рекламного контента и является открытой. Мы исходим из предположения о том, что если выразительный и манипулятивный потенциал языковых единиц психологического воздействия в рекламном дискурсе соотносится с конкретными лексическими и синтаксическими репрезентантами, то формальные признаки данных репрезентантов могут послужить теоретико-лингвистическим основанием для их автоматизированной идентификации в электронных корпусах текстов, а также для дифференциации текстов с позиции определения их дискурсивной принадлежности к рекламному дискурсу. Кроме того, в качестве исходного принципа автоматизированного анализа сопутствующих вербальной экспликации суггестии в рекламном контенте медиатекста учитывается инициальная интенция на продвижение товара, которая рассчитывается по формулам развернутой формы некооперативного взаимодействия в применении к институциональным типам дискурса с базовым перечнем переменных (о них см.: [Бредихин, 2023, с. 54–55]).

В качестве фактического материала привлечены 176 текстов рекламных кампаний различных категорий товаров и услуг, а также 160 текстов, взятых из электронных версий англоязычных средств массовой информации (газет и журналов), не носящих рекламного характера, однако содержащих импликацию ситуативного рекламного контента. Электронный корпус для автоматизированного анализа разработан на основе публично доступных материалов за период с 2011 по 2024 год. Из отобранных для анализа текстов на основе сформированного списка критериальных признаков были составлены ряды маркеров суггестивного воздействия в их понимании по М.Т. Шакеновой [Шакенова и др., 2020].

Первичный анализ избранного материала продемонстрировал, что наиболее эффективным и частотным приемом информационно-оперативной суггестии в медиатекстах является адъективация. На имплементации эмоционально-оценочных или логико-конкретизирующих прилагательных строится 30 % от общего числа микроконтекстов, при этом подавляющее большинство случаев включает несколько таких единиц. В некоторых исследованиях, посвященных проблемам манипулирования, семантическое содержание маркера представляется вторичным, а на первый план выводятся эмоциональные импринты, сохраняющиеся в реципирующем сознании целевой группы [Котов, Переверзева, 2023, с. 44]. Другие приемы интенсификации манипулятивного контента включают большей частью сходные эмотивные элементы, такие как эпитетация (11,4 %), положительная / лучшая презентация (8 %), образная аффилиация с селебрити (4,5 %), социумно-групповая детерминация (подчеркивание привилегированного статуса обладателя рекламируемого товара) (4,5 %), употребление сленгизмов (4,5 %), метафоризация (2,3 %) и использование образных ассоциаций (1,12 %). В то же время одним из наиболее частотных способов представляется псевдоаргументативное манипулирование на основе введения компонентов когнитивной информации – числовые данные (19,3 %), конкретизация описания бонусных акций (4 %). Встречаются и приемы фонографического маркирования (7,4 %). Вышеописанные способы, типичные для медийных текстов, содержащих рекламный контент, классифицируются нами как лексические. Следует подчеркнуть, что не каждое средство суггестии лексического характера отвечает формальным критериальным признакам, на основе которых допустима их автоматизированная делимитация. Идентификация возможно лишь по соответствию прописанным в алгоритме денотативным компонентам, коннотативы, придающие дополнительные обертоны, интенсифицирующие влияние, зачастую не входят в первичный список семантических маркеров [Ngulube, Elezie, 2023, p. 115]. В перечень формально маркированных и потенциально определяемых семантико-семиотическими анализаторами включаются адъектива-

ция, эпитетация, использование числовых данных, образная аффилиация и фонографическое маркирование.

В плане синтаксического маркирования наибольшее распространение (31,3 % из общего числа микроконтекстов) получает употребление глаголов в императиве, что представляет собой особый прием экспликации манипулирования. Использование повторов (22 %), которые детализируют наиболее значимые признаки рекламируемого товара, занимает второе место по частотности. Подавляющее большинство предложений рекламного компонента медийного текста являются номинативными (20,5 %). Эллиптические конструкции и синтаксический параллелизм демонстрируют одинаковую частотность употребления (по 9,7 %), парцелляция же существенно уступает им (4,5 %) и используется в рамках выделения наиболее значимых характеристик. В рамках обеспечения псевдонезависимости принятия решения о приобретении товара наблюдается такой прием, как диалогизация монологической речи (6,25 %). Наименьшее число речупотреблений приходится на сравнение (2,3 %). В отличие от лексических средств набор формальных маркеров всех вышеперечисленных синтаксических средств составляет достаточно легко и служит основой для создания идентификационных алгоритмических процедур.

В предпринятом исследовании для разработки алгоритмов автоматизированной идентификации рекламного контента в текстовых массивах применен корпусный менеджер GATE. С его помощью реализуются задачи по выделению смыслового содержания текста и его репрезентации в структурированном виде в форме аннотаций, добавляемых к сегментам текста [Cunningham, Humphreys, 1997, p. 65]. Идентификация лексических репрезентантов рекламного содержания реализуется посредством технологии поиска именованных сущностей ANNIE Gazetteer, синтаксических репрезентантов – с применением технологии шаблонного поиска и аннотирования текстовых сегментов JAPE (Java Annotation Patterns Engine), а также аннотирования текста по регулярным выражениям Java Regexp Annotator плагина String Annotation.

Анализатор ANNIE Gazetteer работает на основе пользовательских списков слов и выражений, сформированных по тематическому признаку (наиболее употребительные лексические репрезентанты рекламы различного типа). Идентифицированные текстовые совпадения (слова и выражения из тематических списков) отмечаются аннотацией типа Lookup с атрибутами, указанными в словаре.

Технология JAPE представляет собой специализированный метаязык для описания лексико-синтаксических шаблонов с целью обеспечения возможности автоматизированного поиска и аннотирования искомых текстовых сегментов в электронном корпусе текстов [Thakker, 2009, p. 15]. В нашем исследовании для идентификации синтаксических репрезентантов рекламного содержания в тексте применяется анализатор JAPE Transducer. Технология JAPE также позволяет задействовать метаязык регулярных выражений и использовать логические операторы для объединения формальных критериев поиска и идентификации текстовых сегментов по заданному шаблону.

Результаты и обсуждение

Идентификация лексических репрезентантов рекламного контента с применением анализатора ANNIE Gazetteer

Языковой прием активизации когнитивной информации посредством введения числовых данных встречается в 34 из 176 изученных нами рекламных текстов, что составляет 19,3 % от их общего количества. Данный прием наиболее эффективен при конкретизации сроков проведения рекламных акций, распродаж и т. п., а также интенсификации внимания реципиентов на величине процента скидки или указания цены на рекламируемый товар, зачастую формальными маркерами при данных репрезентантах выступают слова, обозначающие дату (день, месяц, неделя, год), а также знаки %, £, €, \$, аббревиатуры USD, EUR и др. Следует указать, что дельмитация вышеописанных символьных маркеров при числовых данных является основой реализации инструментального принципа в

обеспечении маркированного рекламного контента специфическим иллокутивным потенциалом [Каменский, 2022]. Концентрированное употребление таких цифровых показателей в текстовом фрагменте, в том числе в сочетании с лексикой, типичной для рекламных акций (*bonus, buy, sale, savings* и т. п.), в текстах, не относящихся к экономической тематике, может служить одним из показателей его рекламной направленности.

Для автоматизированной идентификации наличия числовых данных манипулятивного рекламного характера в медиатексте использовалась как технология ANNIE Gazetteer, так и технология JAPE. Для ANNIE Gazetteer был создан список потенциальных токенов, сопровождающих данный прием: %, £, €, \$, USD, EUR. Для идентификации таких лексем, как *day, week, month, year, bonus, buy, sale, savings* и т. п., представляется возможным применить и технологию ANNIE Gazetteer, и технологию JAPE Transducer, однако последняя предпочтительнее в силу того, что данные лексемы встречаются в разных грамматических формах. Еще одной причиной выбора двух технологий для автоматизированной идентификации приема использования цифровых показателей является то, что ANNIE Gazetteer идентифицирует отдельные токены, находящиеся в созданных списках. С помощью данной технологии делимитируются токены «%», «€», «\$» и др., если они обособлены от других токенов, то есть идентификация составных конструкций типа «\$14», «20 %» и т. п. будет затруднена. Таким образом, идентификация числовых данных, потенциально выступающих маркерами рекламного характера сообщения, будет двухфазной. На первом этапе с помощью ANNIE Gazetteer идентифицируются указанные выше отдельные токены, перечисленные в файле списка *signs_of_numbers.lst*. Фрагмент данного файла, демонстрирующий принцип его организации, показан ниже (см. рис. 1).

Для аннотирования лексем из списка корпусному менеджеру GATE необходим файл определения, характеризующий целевую аннотацию для создания. Во-первых, токены автоматически идентифицируются по соответствию прописанной аннотации, что позволяет устранить неоднозначность интерпретации и

```

1 %
2 ?
3 $
4 €
5 EUR
6 USD
7 RUR
8 RUB

```

Рис. 1. Фрагмент списка токенов signs_of_numbers.lst

Fig. 1. Fragment of the token list signs_of_numbers.lst

привязать их к тому или иному типу данных. Во-вторых, семантические компоненты, дефинированные в списке ресурсов как целые блоки, применяются для обеспечения семантического полнотекстового поиска в едином информационном потоке [Maynard et al., 2017, p. 77]. Каждая строка, описывающая аннотацию для файла со списком токенов, состоит из пяти частей, разделенных двоеточием:

- 1) название документа со списком слов, которому присваиваются определенные характеристики;
- 2) основная категория, к которой относятся слова (например, advertisement);
- 3) подкатегория, к которой относятся слова (например, lexical means);
- 4) язык, на котором написаны слова в списке (например, English);
- 5) название, которое будет отражено в ANNIE Gazetteer в списке аннотаций Annotation Sets (например, advertisement_markers).

Файл определения signs_of_numbers.def, задающий параметры аннотирования вышеуказанных токенов по списку signs_of_numbers.lst, содержит следующее описание:

```
signs_of_numbers.lst:lexical:cost_markers:
English:Advertisement_Markers
```

В абсолютном соотношении языковой прием положительной / лучшей презентации товара встречается в 14 из 176 микроконтекстов, что в относительном выражении составляет 8 % от общего количества. Данный прием предполагает наличие лексических единиц, описывающих лучшие качества товара, а также (в ограниченном числе употреблений) сравнение с другими товарами, поэтому формальным критерием является наличие таких лек-

сем, как *no better*, *best*, *world's*, *only*, *better*, *than*, *first*, *unique*, *superior* и т. п. Например:

Wheel Deal: 'We don't value them' – the **best** estate cars you can pick up for under 5k INCLUDING a trio of luxurious 'executive' motors (The Sun);

SKIMS LOVE: For 48 hours only, get some of our **most iconic** styles for less than ever before (Skims.com).

Фрагмент перечня формальных маркеров положительной презентации, размещенных по принципу списка по одному на каждой строке в файле списка для анализатора ANNIE Gazetteer, приведен ниже (см. рис. 2).

Файл определения для данного списка лексем назван нами best_product.def, он содержит такое определение аннотации:

```
best_product.lst:lexical:qualifiers:English:
Advertisement_Markers
```

Следующим приемом манипулирования, исследованным с позиции алгоритмизации, было использование лексем, характерных для описания бонусных акций (4 % от общего количества способов манипулирования). Формальным признаком данного приема является наличие таких слов как *win*, *prize*, *bonus*, *free*, *complimentary*, *coupons* и т. п. Данные лексемы обозначены в файле списка bonus.lst, фрагмент которого приведен ниже (см. рис. 3).

Файл определения bonus.def для данного списка содержит следующее определение:

```
prize.lst:lexical:bonus:English:Advertisement_
Markers
```

Последний список, созданный для ANNIE Gazetteer, представляет собой список лексических единиц, наиболее употребительных в рекламных текстах. С помощью программ-

```

1 no better
2 best
3 world's
4 only
5 better
6 than
7 first
8 unique
9 superior
10 good
11 new

```

Рис. 2. Фрагмент списка лексических единиц best_product.lst

Fig. 2. Fragment of the list of lexical units best_product.lst

```

1 win
2 prize
3 bonus
4 free
5 complimentary
6 coupons
7 buy

```

Рис. 3. Фрагмент списка лексических единиц bonus.lst

Fig. 3. Fragment of the list of lexical units bonus.lst

ного продукта Textus Pro (ПО для нахождения частоты и плотности использования ключевых слов) были выделены наиболее частотные ключевые слова в 176 рассмотренных рекламных текстах: *new, taste, free, good, style, quality, enjoy, better, true, best, unique, buy, super, win, prize, visit, cost, way, flavor, significantly, significant*. Например:

Taste the A.I.: **Unique** desserts made with artificial intelligence (Ads of the World).

Список данных лексем оформлен в виде файла списка frequent_words.lst. Файл определения frequent_words.def для данного списка содержит следующее определение:

```
frequent_words.lst:lexical:frequent_ad_words:
English:Advertisement_Markers
```

Таким образом, созданы теоретико-методологические основания автоматизированной идентификации лексических единиц, релевантных для идентификации рекламных текстов и рекламного контента в СМИ, на основе потенциального присутствия в них средств суггестивного воздействия с помощью технологии ANNIE Gazetteer корпусного менеджера GATE.

Перейдем к описанию механизма создания программной основы для автоматизированного идентификатора рекламных текстов с помощью технологии шаблонного лексикограмматического анализа и аннотирования JAPE Transducer.

Идентификация синтаксических репрезентантов рекламного контента с применением технологии JAPE Transducer

В отличие от технологии ANNIE Gazetteer, направленной на поиск конкретных лексических единиц, технология JAPE Transducer предназначена для поиска текстовых сегментов по общим формулам, записанным в особом образом организованных JAPE-правилах и отражающих как лексический состав и лексическую сочетаемость слов в искомых текстовых фрагментах, так и синтаксическую структуру данных фрагментов, что расширяет поисковый диапазон алгоритма идентификации рекламного текста [Thakker, 2009].

С помощью технологии поиска JAPE Transducer проведена разработка алгоритмических процедур для автоматизированной идентификации следующих лексико-синтаксических средств психологического воздействия: адыек-

тивация, использование цифровых показателей (в сочетании со списком ANNIE Gazetteer), употребление повелительного наклонения, использование номинативных предложений.

Алгоритмические процедуры записаны с помощью метаязыка описания поисковых шаблонов JARE, основанного на синтаксисе языка программирования Java. Анализатор JARE Transducer, входящий в состав корпусного менеджера GATE, работает на основе специальных файлов с расширением «.jare», являющихся JARE-правилами. Разработанные JARE-правила имеют следующие компоненты:

1) наименование фазы (например: *Phase: IdentifyAdjectives*);

2) типы аннотаций, учитываемые при поиске текста по данному правилу (например: *Input: Token*). В процессе анализа используется аннотация Token, в которой токен принимается в качестве минимальной единицы анализа. При необходимости в перечень учитываемых аннотаций также добавляется стандартная для анализатора ANNIE аннотация Split, отмечающая границы предложений и других структурных блоков текста. При включении дополнительного конкретизатора наблюдается бифуркационная ситуация, строящаяся на рекурсивном сравнении с конечным перечнем отсылок, что не влияет на увеличение операций базового алгоритма [Бредихин, Сидоренко, 2023, с. 51]. Следует отметить, что ограничение операций при создании аннотаций происходит в рамках линейной регрессии, которая демонстрирует демаркацию зависимых и независимых переменных [Ghosh, Senthilrajan, 2023, p. 29234];

3) опции поискового алгоритма (например: *Options: control=brill*). Применяемый в настоящем исследовании алгоритм Brill позволяет создать аннотации для текстовых совпадений по каждому из заданных JARE-правил независимо от того, пересекаются ли данные текстовые совпадения в текстовом массиве или нет;

4) название правила (например: *Rule: Numbers*);

5) формула для осуществления поиска (например: *{Token.category=="JJS"}*), в которой для указания частеречной принадлежности лексем используются общепринятые сокращения Nipple Tagger [Appendix G...];

6) аннотации, присваиваемые найденным с помощью заданного поискового шаблона единицам текста (например: *Item.AdvertisementMarkers = {category = "adjective", degree = "superlative"}*). В данном примере в списке аннотаций Annotation Sets найденные единицы текста будут отражаться в наборе аннотаций «Advertisement_Markers» и иметь аннотацию со свойствами «category: adjective» и «degree: superlative» (превосходная степень сравнения прилагательного).

По результатам анализа фактического материала лексический прием адъективации идентифицирован в 52 из 176 рекламных текстов, что составляет 30 % от рассмотренных рекламных текстов. Поскольку в основе данного приема лежит прилагательное в сравнительной или превосходной степени, то в качестве формального поискового критерия выбрано наличие в микроконтексте по меньшей мере одного прилагательного в сравнительной или превосходной степени, то есть прилагательного с суффиксами *-est / -er*, либо со словами *more, less, most*, опционально предваренными определенным артиклем *the*.

Исходя из вышесказанного, разработанные JARE-правила для идентификации потенциальных контекстов с адъективацией имеют следующий вид (где тег JJR означает прилагательное в сравнительной степени, тег JJS – прилагательное в превосходной степени):

```
Phase: IdentifyAdvertisementMarkers
Input: Token
Options: control=brill
Rule: AdjComparative
(
    {Token.category == "JJR"}
)
:Item -> :Item.Advertisement_Markers =
{category = "adjective", degree = "comparative"}
Rule: AdjSuperlative
(
    {Token.category == "JJS"}
)
:Item -> :Item.Advertisement_Markers =
{category = "adjective", degree = "superlative"}
```

Прием использования числовых данных встретился в 34 из 176 изученных нами рекламных текстах, что составляет 19,3 % от

общего количества рассмотренных рекламных текстов. Данный прием идентифицируется в два этапа, при этом первый этап проводится с применением технологии ANNIE Gazetteer, о чем было указано выше. Технология JAPE Transducer позволяет уточнить поисковый алгоритм в части поиска сочетаний лексических единиц и графических показателей (знаков), при этом для поиска последних представляется возможным использовать регулярное выражение, совпадающее с одним любым символом из заданного класса, либо с одной из заданных аббревиатур. Таким образом, JAPE-правило для идентификации цифровых показателей в рекламном тексте имеет следующий формат:

```
Phase: IdentifyAdvertisementMarkers
Input: Token
Options: control=brill
Rule: AdjComparative
(
    {Token.kind == "number"}
    ({Token.string == ~
    "[%£€$]" || Token.string == "USD" ||
    Token.string == "EUR"})
)
:Item —> :Item.Advertisement_Markers =
{category = "number", degree =
"price_or_percentage"}
```

Приведенное выше JAPE-правило демонстрирует общий принцип идентификации необходимых формальных репрезентантов и может быть дополнено по принципу аналогии другими цифровыми показателями, лексемами и аббревиатурами. Необходимым расширительным компонентом в данном случае будут выступать альтернативы внутрисистемного перераспределения интенционально (заложенного в исходном списке) релевантного набора регулярных выражений по определенным алгоритмическим схематическим моделям [Бредихин, 2014].

Следующий прием, для идентификации которого разработано JAPE-правило, – это доминантное употребление глаголов в императиве, частотность которого в суггестивно маркированном контенте достигает 31,3 %.

Формальным показателем, на основе которого автоматически определяется данный

способ интенсификации оперативного компонента в медиатексте, служит не единичное, но концентрированное использование глаголов в повелительном наклонении, преимущественно находящихся на первом месте в предложении, например:

Change your life. **Take** a CELTA course Teaching English as a Foreign Language. **Open** a world of possibility and rewarding international career. **Call** Saxoncourt for more details (London Evening Standard).

Emirates. Luxury holidays. Superb savings. **Enjoy** savings of over 300 on Dubai holidays with Emirates. **Fly** Emirates. **Keep** discovering (London Evening Standard).

Для идентификации данного синтаксического приема применен стандартный тег «VB» для поиска глаголов в повелительном наклонении, при этом позиция глагола в предложении устанавливается на основе стандартных аннотаций типа Split, используемых в анализаторе ANNIE корпусного менеджера GATE в качестве разделяющих маркеров предложений [Appendix G...]. Для корректной идентификации таких маркеров в перечень анализируемых типов аннотаций добавлялись аннотации границ предложения Split, то есть: «Input: Token Split». Результирующее правило JAPE выглядит таким образом:

```
Phase: IdentifyAdvertisementMarkers
Input: Token Split
Options: control=brill
Rule: ImperativeVerb
(
    {Split}
    {Token.category == "VB"}
    ({Token, !Split})+
)
:Item —> :Item.Advertisement_Markers =
{category = "verb", mood = "imperative"}
```

Следующим приемом, рассмотренным нами с позиции алгоритмизации идентификации, было использование предложения номинативного типа, абсолютное количество которых составило 36 из 176 микроконтекстов (20,5 %). Критериальным фактором автоматизированной идентификации в данном случае является отсутствие глагола в любой форме. Следует отметить, что в качестве отдельно-

го номинанта в большинстве случаев выделяется наименование рекламируемого продукта или бренда. Например:

Estee Lauder. Global makeup artist event. Join us at Estee Lauder from 3–9 November for an exclusive makeup event, only at Selfridges London (Stylist);

Zeha Berlin. The berlin shoe with a story. Zeha – the perfect marriage of sports and fashion. Get to know us in one of our unique stores (Berlin & I. City Guide).

Для идентификации данного приема производится поиск произвольной последовательности неглагольных токенов в промежутке между аннотациями границ предложения (Split). Для обучения алгоритма поиску неявных (замаскированных) токенов, необходимо четкое прописывание случаев омографии в качестве инициальной подсказки для добавления операции контекстуального семантического и позиционного сравнения. Такое введение функциональных ограничений предполагает намеренное искажение с помощью ограниченного шаблона с заменой компонента омографом [Jiang et al., 2023]. Формула поиска, исходя из этого, имеет следующий формат:

```
{Split}
(!Split, Token.category !=~ "VB.*")+ .
{Split}
```

В данной поисковой формуле первая и последняя строки (*{Split}*) обозначают границы предложения, вторая строка уточняет, что поиск проходит в границах последовательности слов, объединенных в одно предложение, с отсутствием глагола (*Token.category !=~ "VB.*"*). Знак «+» в конце строки представляет собой стандартный метасимвол JAPE, означающий повторение заданного в строке элемента 1 и более раз. *VB.** в данной формуле является стандартным тегом, содержащим глагол в любой его форме [Appendix]. Результирующее правило JAPE следующее:

```
Phase: IdentifyAdvertisementMarkers
Input: Token Split
Options: control=brill
Rule: NominativeSentences
(
    {Split}
```

```
{!Split, Token.category !=~ "VB.*")+
{Split}
```

)

```
:Item —> :Item.Advertisement_Markers =
{category = "sentence", type = "nominative"}
```

Недостатком такого поискового запроса является то, что по умолчанию он не охватывает первое предложение в рекламном тексте, так как перед первым предложением нет аннотации «Split». Данная проблема может быть решена путем добавления аннотации типа «Split» в начало каждого документа, входящего в пользовательский корпус текстов, то есть внедрения дополнительного оперативного форманта иницирующего заданный поиск [Поздин, Хохлов, 2017].

Формальными признаками приема использования лексических повторов выступают повторение лексических единиц, относящихся к названию торговых марок и рекламируемых товаров, либо к их свойствам и качествам. Например:

Samsung Galaxy Note. The Galaxy Note £10000 prize Draw. Wow! Experience the **Galaxy Note** in store this Thursday and you could win a share of £10000 worth of **Samsung** products. Phone? Tablet? It's **Galaxy Note!** Only at Carphone Warehouse (London Evening Standard).

Идентификация лексических повторов с применением технологии JAPE затруднительна в связи с невозможностью задействовать регулярные выражения в JAPE-правилах для поиска последовательности идентичных токенов. Таким образом, для идентификации приема повторов имен собственных нами использованы регулярные выражения, разработанные нами на платформе Regex 101 [Regular Expressions 101] и далее примененные в алгоритмической процедуре Java Regexp Annotator стандартного плагина String Annotation [String Annotation Plugin]. Получаемые в результате работы данного алгоритма аннотации совместимы с аннотациями анализаторов JAPE и ANNIE Gazetteer. Ниже приводится содержание файла *lexical_repetition.rgx*, задающего шаблон регулярного выражения для поиска лексических повторов с помощью анализатора Java Regexp Annotator:

$(\backslash\text{b}[A-Z][A-Za-z]\backslash\text{w}+\backslash\text{b})[\text{^\.!?}]^+ \backslash 1[\text{^\.!?}]^+[\text{.!?}]^+ 0 \Rightarrow \text{AdvertisementMarkers lexicalunit}=\1

Данное регулярное выражение состоит из следующих блоков:

(...) – объединяет элементы в одну группу (в нашем случае это повторяющаяся лексема);

$\backslash\text{b} \dots \backslash\text{b}$ – делимитирует границы слова;

$[A-Z][A-Za-z]\backslash\text{w}^+$ – задает класс символов, составляющих англоязычную лексему, при этом первая буква лексической единицы должна быть заглавной, поскольку речь идет о повторах имен собственных (преимущественно – названий компаний);

$[\text{^\.!?}]^+$ – указывает, что далее следует любое количество произвольных знаков, не относящихся к терминальным знакам препинания;

$\backslash 1$ – указывает на повтор группы 1, содержащей найденную лексическую единицу;

$[\text{.!?}]^+$ – соответствует любому терминальному знаку препинания или сочетанию таких знаков, завершающих предложение.

Итак, для проведения автоматизированной идентификации лексико-синтаксических средств психологического воздействия, таких как адъективация, использование числовых данных в качестве когнитивного интенсификатора, повелительного наклонения и номинативных предложений наиболее эффективным признается функционал анализаторов JAPE Transducer и Java Regexp Annotator корпусного менеджера GATE. Применение совокупности данных правил формирует алгоритмическую основу автоматической идентификации рекламного текста и рекламного контента в медийном пространстве современных СМИ.

Верификация разработанных алгоритмических процедур проведена на основе пользо-

вательского корпуса, включающего 176 рекламных текстов и 160 нерекламных текстов различных функциональных стилей. Результаты сопоставительного анализа показали следующее:

– в 176 рекламных текстах лексико-синтаксические средства психологического воздействия в среднем составляют 24,6 % от общего объема рекламного текста;

– в 160 медийных текстах, не содержащих эксплицитный рекламный контекст, лексико-синтаксические средства психологического воздействия в среднем составляют 8,9 % от общего объема текста.

Результаты количественного сравнительного анализа фактического материала представлены в таблице.

Существенное различие в процентном соотношении количества совпадений на основании автоматизированного анализа к общему числу токенов в рекламных текстах и медийных текстах другой жанровой принадлежности позволяет говорить о валидности созданных алгоритмических процедур в качестве базиса для автоматизации идентификации рекламных текстов и рекламного контента в СМИ, то есть разработанный перечень формальных маркеров экспликации / импликации суггестии в совокупности является необходимым и достаточным основанием для автоматического распознавания манипулятивных публикаций. Полученные результаты, таким образом, согласуются с другими научными изысканиями в области распознавания манипулятивного текста (см.: [Кенжебалина и др., 2020, с. 123]) и дополняют их в части описания принципов формализации суггестивного воздействия в тексте.

При использовании разработанного программного модуля для автоматизированной

Количественное соотношение формальных маркеров потенциально рекламного контента в рекламных и нерекламных медийных текстах

Quantitative ratio of formal markers of potentially advertising content in advertising and non-advertising media texts

Тип текста	Название документа	Общее количество токенов	Количество совпадений	Процентное соотношение
Рекламный	advertisement	57 010	14 052	24,6
Медийный нерекламный	media	61 020	5 480	8,9

идентификации текстов на предмет принадлежности к рекламе следует обращать внимание на количество идентифицируемых числовых показателей, которые необходимо сопоставляются с контекстно-детерминированными шаблонами, содержащими как вербальные, так и формальные маркеры в неоднородных семантических сетях целого ряда сходных высказываний (подробно об этом см.: [Салимовский и др., 2019, с. 85]). Следует учитывать тот факт, что среди обрабатываемых с помощью данного программного модуля документов могут находиться документы с различными статистическими сведениями, включающими большое количество цифровых данных, что окажет существенное влияние на количество совпадений поиска. Во избежание ошибочной трактовки подобных текстов как рекламных необходимо приведение процентного соотношения количества цифровых показателей в рекламных текстах к общему числу токенов. Исходя из анализа 176 рекламных и 160 нерекламных текстов, был получен следующий результат: в рекламном тексте количество цифровых показателей, как правило, не превышает 5 % от общего количества токенов. Таким образом, если в тексте, классифицируемом по количеству лексико-синтаксических приемов как рекламном, количество цифровых данных существенно превышает данный порог, то существует значительная вероятность того, что он является бухгалтерским документом или текстом с иной дискурсивной принадлежностью, например, текстом экономической тематики.

Количество повторов в рекламных текстах составило 34 % от общего объема текста, в то время как в нерекламных текстах – 54 %. Исходя из этого, данный прием не может рассматриваться в качестве ключевого критерия, достаточного для идентификации рекламного текста, однако применим в качестве сопутствующего критерия в сочетании с другими описанными в настоящем исследовании критериями.

Результаты проведенной верификации позволили сделать следующие выводы: предложенные алгоритмы автоматизированного анализа текстов отвечают поставленным целям и задачам, реализуя алгоритм поиска формальных маркеров рекламного контента в ка-

честве подключаемого модуля GATE, состоящего из JAPE-правил, Gazetteer-списков и регулярного выражения для плагина Java Regex Annotator; апробация на материале рекламных и нерекламных текстов позволила говорить о высокой степени валидности предложенного алгоритма и, как следствие, о применимости проекта для идентификации рекламных текстов.

Выводы

Рекламные тексты, имеющие в своей основе ряд специфических языковых средств психологического воздействия, являющихся имманентным атрибутом рекламного дискурса, подлежат автоматизированной идентификации и классификации на основе контекстуально идентифицируемых формальных маркеров, манифестирующихся на лексическом, грамматическом, стилистическом и графическом уровнях. Формальные критерии идентификации рекламных текстов открывают возможность автоматизации идентификации рекламного контента в медийном пространстве современных электронных СМИ.

Проведя сопоставительный анализ рекламных текстов и нерекламных медийных текстов с применением разработанного автоматизированного анализатора, мы выяснили, что в рекламных текстах доминируют лексико-синтаксические приемы манипулятивного воздействия. Результаты исследования демонстрируют наличие существенного различия в процентном соотношении совпадений поиска ко всему объему текста в рекламных и нерекламных текстах при идентификации рекламного контента на основе формальных маркеров, позволяющих интеллектуальным системам, основанным на семантико-синтаксических анализаторах, четко делимитировать высказывания манипулятивного характера в привязке к объектному позиционированию. Это доказывает эффективность использования автоматизированных систем анализа в идентификации эксплицитных и имплицитных сообщений рекламного характера в медийных текстах, а также установления дискурсивной принадлежности текста, опубликованного в СМИ, с целью его классификации как информационного либо рекламного.

Полученные результаты и выводы имеют практическую применимость в области классификации текстов по их дискурсивной принадлежности к рекламному или нерекламному типу дискурса, в сфере идентификации в текстах различной дискурсивной принадлежности формальных маркеров скрытой суггестии и скрытой рекламы, а также автоматизации идентификации в электронных корпусах текстов дискурсивных средств речевого воздействия, обладающих потенциалом эксплицитной и имплицитной манипуляции массовым сознанием.

СПИСОК ЛИТЕРАТУРЫ

- Бредихин С. Н., 2014. Схемопостроение в рамках метаединиц герменевтического процесса понимания и интерпретации // *Современные проблемы науки и образования*. № 4. URL: <https://science-education.ru/ru/article/view?id=13920&ysclid=lsblym0dt3574940601>
- Бредихин С. Н., 2023. Теория некооперативных игр в порождении и интерпретации высказываний институционального дискурса // *Профессиональная коммуникация: актуальные вопросы лингвистики и методики*. № 16. С. 50–58.
- Бредихин С. Н., Сидоренко С. Г., 2023. Рекурсивное сравнение как основа поиска трансляционных эквивалентов в агглютинативных и флективных языках // *Вестник Волгоградского государственного университета. Серия 2, Языкознание*. Т. 22, № 3. С. 43–56. DOI: <https://doi.org/10.15688/jvolsu2.2023.3.4>
- Ильинова Е. Ю., Волкова О. С., 2023. Динамика медиатизации транскультурного концепта «Индивидуальная мобильность»: корпусно-ориентированное исследование // *Вестник Волгоградского государственного университета. Серия 2, Языкознание*. Т. 22, № 5. С. 19–39. DOI: <https://doi.org/10.15688/jvolsu2.2023.5.2>
- Каменский М. В., 2022. Дискурсивные маркеры в трансдисциплинарном освещении: от философии языка Л. Витгенштейна к когнитивной лингвистике и искусственному интеллекту. Ставрополь : СКФУ. 133 с.
- Каменский М. В., Ломтева Т. Н., Кабылкина Н. С., Бутенко С. А., Триско И. Ю., Шереметова Д. А., Федина Е. Р., 2016. Формализация исследовательских процедур анализа семантики языковых единиц. Ставрополь : СКФУ. 170 с.
- Кенжебалина Г. Н., Шаикова Г. К., Шакенова М. Т., Акоева И. Г., 2020. Распознавание манипулятивного текста: структура и доминантные языковые параметры // *Научный диалог*. № 7. С. 105–125. DOI: [10.24224/2227-1295-2020-7-105-125](https://doi.org/10.24224/2227-1295-2020-7-105-125)
- Котов А. А., Переверзева С. И., 2024. Автоматический анализ фреймов для оценки воздействия текста // *Вестник Волгоградского государственного университета. Серия 2, Языкознание*. Т. 23, № 3. С. 41–54. DOI: <https://doi.org/10.15688/jvolsu2.2024.3.4>
- Поздин В. Н., Хохлов М. Г., 2017. Расчет формант по участку речевого сигнала // *Advanced Science*. № 1. URL: [http://advanced-science.ru/assets/mgr/docs/1\(2017\)/%D0%A2%D0%B5%D1%85%D0%BD%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B5/pozdin-hohlov-k-pechati.pdf](http://advanced-science.ru/assets/mgr/docs/1(2017)/%D0%A2%D0%B5%D1%85%D0%BD%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B5/pozdin-hohlov-k-pechati.pdf)
- Салимовский В. А., Девяткин Д. А., Каджая Л. А., Мишланов В. А., 2019. Автоматическое распознавание ментальных действий, реализуемых в научных эмпирических текстах // *Научно-технические ведомости СПбГПУ. Гуманитарные и общественные науки*. Т. 10, № 3. С. 74–88. DOI: [10.18721/JHSS.10307](https://doi.org/10.18721/JHSS.10307)
- Шакенова М. Т., Ташимханова Д. С., Оспанова У. А., Булдыбаев Т. К., 2020. Лексико-грамматические маркеры манипулятивности массмедийного дискурса (на материале корпуса русскоязычных текстов СМИ в Республике Казахстан) // *Научный диалог*. № 2. С. 141–160. DOI: [10.24224/2227-1295-2020-2-141-160](https://doi.org/10.24224/2227-1295-2020-2-141-160)
- Appendix G. Part-of-Speech Tags used in the Happle Tagger // *GATE – General Architecture for Text Engineering*. URL: <https://gate.ac.uk/sale/tao/splitap7.html>
- Brezina V., Platt W., 2023. #LancsBox X [software]. Lancaster University. URL: <https://lancsbox.lancs.ac.uk>
- Cunningham H., Humphreys K., 1997. *Software Infrastructure for Natural Language Processing*. Sheffield : University of Sheffield. 272 p.
- GATE. General Architecture for Text Engineering. URL: <https://gate.ac.uk>
- Ghosh A., Senthilrajan A., 2023. Comparison of Machine Learning Techniques for Spam Detection // *Multimed Tools Application*. № 82. P. 29227–29254. DOI: [10.1007/s11042-023-14689-3](https://doi.org/10.1007/s11042-023-14689-3)
- Jiang Y., Yu X., Wang Y., Xu X., Song X., Maynard D., 2023. Similarity-Aware Multimodal Prompt Learning for Fake News Detection // *Information Sciences*. Vol. 647. DOI: [10.1016/j.ins.2023.119446](https://doi.org/10.1016/j.ins.2023.119446)
- Maynard D., Roberts I., Greenwood M. A., Rout D., Bontcheva K., 2017. A Framework for Real-Time Semantic Social Media Analysis // *Journal of Web Semantics*. Vol. 44. P. 75–88. DOI: [10.1016/j.websem.2017.05.002](https://doi.org/10.1016/j.websem.2017.05.002)
- Ngulube I. E., Elezie K., 2023. Language Style in Advertising: A Study of the Print Media // *International Journal*

of Literature, Language and Linguistics. Vol. 6, iss.3. P. 97–128. DOI: 10.52589/IJLLL_9DX2II7X
Regular Expressions 101. URL: <https://regex101.com>
String Annotation Plugin. URL: <https://gatenlp.github.io/gateplugin-StringAnnotation/>
Thakker D., 2009. GATE JAPE Grammar Tutorial. Version 1.0. Nottingham : Nottingham Trent University. 201 p.

ИСТОЧНИКИ

Ads of the World. November 1, 2022. URL: <https://www.adsoftheworld.com>
Berlin & I. City Guide. 2013. № 5. URL: <https://www.berlin-and-i.de>
London Evening Standard. November 1, 2011. URL: <https://www.standard.co.uk/>
Skims. August 14, 2024. URL: <https://skims.com>
Stylist. November 2, 2011. Iss. 100. P. 10. URL: <https://www.stylist.co.uk/>
The Sun. August 29, 2024. URL: <https://www.thesun.co.uk/>

REFERENCES

Bredikhin S.N., 2014. Skhemopostroyeniye v ramkakh metaedinitov germenevticheskogo protsessa ponimaniya i interpretatsii [Scheme Derivation within Metaunits of Hermeneutic Processes of Understanding and Interpretation]. *Sovremennyye problemy nauki i obrazovaniya* [Modern Problems of Science and Education], no. 4. URL: <https://science-education.ru/ru/article/view?id=13920&ysclid=lsblym0dt3574940601>
Bredikhin S.N., 2023. Teoriya nekooperativnykh igr v porozhdenii i interpretatsii vyskazyvaniy institutsionalnogo diskursa [Theory of Noncooperative Games in Generation and Interpretation of Utterances in Institutional Discourse]. *Professionalnaya kommunikatsiya: aktualnye voprosy lingvistiki i metodiki* [Professional Communication: Topical Issues of Linguistics and Methodology], no. 16, pp. 50-58.
Bredikhin S.N., Sidorenko S.G., 2023. Rekursivnoe sravneniye kak osnova poiska translyatsionnykh ekvivalentov v agglutinativnykh i flektivnykh yazykakh [Recursive Comparison as a Basis to Translation Equivalents Search in Agglutinative and Inflectional Languages]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], vol. 22,

no. 3, pp. 43-56. DOI: <https://doi.org/10.15688/jvolsu2.2023.3.4>

Ilyinova E.Yu., Volkova O.S., 2023. Dinamika mediatizatsii transkulturnogo kontsepta «Individualnaya mobilnost»: korpusno-orientirovannoe issledovanie [Transcultural Concept “Individual Mobility”: Mediatization Dynamics Through Corpus-Based Study]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], vol. 22, no. 5, pp. 19-39. DOI: <https://doi.org/10.15688/jvolsu2.2023.5.2>

Kamenskiy M.V., 2022. *Diskursnyye markery v transdistsiplinarnom osveshchenii: ot filosofii yazyka L. Vitgenshteyna k kognitivnoy lingvistike i iskusstvennomu intellektu* [Discourse Markers in Transdisciplinary Coverage: From L. Wittgenstein’s Philosophy of Language to Cognitive Linguistics and Artificial Intelligence]. Stavropol, SKFU. 133 p.

Kamenskiy M.V., Lomteva T.N., Kabyalkina N.S., Butenko S.A., Trisko I.Yu., Sheremetova D.A., Fedina E.R., 2016. *Formalizatsiya issledovatel'skikh protsedur analiza semantiki yazykovykh edinitov* [Research Procedures Formalization for Analyzing the Semantics of Linguistic Units]. Stavropol, SKFU. 170 p.

Kenzhebalina G.N., Shaikova G.K., Shakenova M.T., Akoyeva I.G., 2020. Raspoznavaniye manipulativnogo teksta: struktura i dominantnyye jazykovyye parametry [Manipulative Text Recognition: Structure and Dominant Linguistic Parameters]. *Nauchnyy dialog*, no. 7, pp. 105-125. DOI: <https://doi.org/10.24224/2227-1295-2020-7-105-125>

Kotov A.A., Pereverzeva S.I., 2024. Avtomaticheskiiy analiz freymov dlya otsenki vozdeystviya teksta [Automatic Frame Analysis for the Evaluation of Text Influence]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], vol. 23, no. 3, pp. 41-54. DOI: <https://doi.org/10.15688/jvolsu2.2024.3.4>

Pozdin V.N., Khokhlov M.G., 2017. Raschet formant po uchastku rechevogo signala [Formants Calculation Based on a Section of the Speech Signal]. *Advanced Science*, no. 1. URL: [http://advanced-science.ru/assets/mgr/docs/1\(2017\)/%D0%A2%D0%B5%D1%85%D0%BD%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B5/pozdin-hokhlov-k-pechati.pdf](http://advanced-science.ru/assets/mgr/docs/1(2017)/%D0%A2%D0%B5%D1%85%D0%BD%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B5/pozdin-hokhlov-k-pechati.pdf)

Salimovskiy V.A., Devyatkin D.A., Kadzhaya L.A., Mishlanov V.A., 2019. Avtomaticheskoe raspoznavaniye mentalnykh deystviy, realizuemykh

- v nauchnykh empiricheskikh tekstakh [Automatic Identification of Mental Actions in Scientific Empirical Texts]. *Nauchno-tekhnicheskie vedomosti SPbGPU. Gumanitarnye i obshchestvennye nauki* [St. Petersburg State Polytechnic University Journal. Humanities and Social Sciences], vol. 10, no. 3, pp. 74-88. DOI: 10.18721/JHSS.10307
- Shakenova M.T., Tashimkhanova D.S., Ospanova U.A., Buldybayev T.K., 2020. Leksiko-grammaticheskie markery manipulativnosti massmediynogo diskursa (na materiale korpusa russkoyazychnykh tekstov SMI v Respublike Kazakhstan) [Lexical-Grammatical Markers of the Manipulativeness of Mass Media Discourse (Based on the Material of the Body of Russian-Language Media Texts in Republic of Kazakhstan)]. *Nauchnyi dialog*, no. 2, pp. 141-160. DOI: 10.24224/2227-1295-2020-2-141-160
- Appendix G. Part-of-Speech Tags used in the Hepple Tagger. *GATE – General Architecture for Text Engineering*. URL: <https://gate.ac.uk/sale/tao/splitap7.html>
- Brezina V., Platt W., 2023. *#LancsBox X Software*. Lancaster University. URL: <https://lancsbox.lancs.ac.uk>
- Cunningham H., Humphreys K., 1997. *Software Infrastructure for Natural Language Processing*. Sheffield, University of Sheffield. 272 p.
- GATE. General Architecture for Text Engineering*. URL: <https://gate.ac.uk>
- Ghosh A., Senthilrajan A., 2023. Comparison of Machine Learning Techniques for Spam Detection. *Multimed Tools Application*, no. 82, pp. 29227-29254. DOI: <https://doi.org/10.1007/s11042-023-14689-3>
- Jiang Y., Yu X., Wang Y., Xu X., Song X., Maynard D., 2023. Similarity-Aware Multimodal Prompt Learning for Fake News Detection. *Information Sciences*, vol. 647. DOI: 10.1016/j.ins.2023.119446
- Maynard D., Roberts I., Greenwood M.A., Rout D., Bontcheva K., 2017. A Framework for Real-Time Semantic Social Media Analysis. *Journal of Web Semantics*, vol. 44, pp. 75-88. DOI: 10.1016/j.websem.2017.05.002
- Ngulube I.E., Elezie K., 2023. Language Style in Advertising: A Study of the Print Media. *International Journal of Literature, Language and Linguistics*, vol. 6, iss. 3, pp. 97-128. DOI: https://doi.org/10.52589/IJLL_9DX2II7X
- Regular Expressions 101*. URL: <https://regex101.com>
- String Annotation Plugin*. URL: <https://gatenlp.github.io/gateplugin-StringAnnotation/>
- Thakker D., 2009. *GATE JAPE Grammar Tutorial. Version 1.0*. Nottingham, Nottingham Trent University. 201 p.

SOURCES

- Ads of the World*. November 1, 2022. URL: <https://www.adsoftheworld.com>
- Berlin & I. City Guide*, 2013, no. 5. URL: <https://www.berlin-and-i.de>
- London Evening Standard*. November 1, 2011. URL: <https://www.standard.co.uk/>
- Skims*. August 14, 2024. URL: <https://skims.com>
- Stylist*, 2011, November 2, iss. 100, p. 10. URL: <https://www.stylist.co.uk/>
- The Sun*. August 29, 2024. sURL: <https://www.thesun.co.uk/>

Information About the Authors

Mikhail V. Kamensky, Doctor of Sciences (Philology), Professor, Department of Linguistics, Linguodidactics and Intercultural Communication, North-Caucasus Federal University, Stavropol, Russia, stavdev@mail.ru, <https://orcid.org/0000-0001-8358-9516>

Sergey S. Bredikhin, Doctor of Sciences (Philology), Professor, Department of Theory and Practice of Translation, North-Caucasus Federal University, Stavropol, Russia, bredichinsergey@yandex.ru, <https://orcid.org/0000-0002-2191-4982>

Информация об авторах

Михаил Васильевич Каменский, доктор филологических наук, профессор кафедры лингвистики, лингводидактики и межкультурной коммуникации, Северо-Кавказский федеральный университет, г. Ставрополь, Россия, stavdev@mail.ru, <https://orcid.org/0000-0001-8358-9516>

Сергей Николаевич Бредихин, доктор филологических наук, профессор кафедры теории и практики перевода, Северо-Кавказский федеральный университет, г. Ставрополь, Россия, bredichinsergey@yandex.ru, <https://orcid.org/0000-0002-2191-4982>