

Научная статья

УДК 81"33

DOI: <https://doi.org/10.18721/JHSS.16404>

EDN: <https://elibrary/WYIJRM>



ИМЕНОВАННЫЕ СУЩНОСТИ В НЕМЕЦКОЯЗЫЧНОЙ ПРЕССЕ: КОРПУСНЫЙ И ЭКСПЕРТНЫЙ АНАЛИЗ

М.В. Корышев , М.В. Хохлова  

Санкт-Петербургский государственный университет,
Санкт-Петербург, Российская Федерация

 m.khokhlova@spbu.ru

Аннотация. Анализ имен собственных, упоминаемых в новостных текстах, представляет отдельный исследовательский интерес, поскольку позволяет косвенным образом определить затрагиваемые в изданиях темы. В статье представлены результаты анализа автоматической процедуры по извлечению именованных сущностей на материале немецкоязычной прессы. Исследование было проведено на материале как общегерманских изданий, нацеленных на широкий охват аудитории, так и региональных и локальных газет, ориентированных на более узкую аудиторию федеральных земель Германии. Работа осуществлялась в два этапа: в ходе первого этапа при помощи инструмента Stanza из текстов каждого издания, а также из всей коллекции статей в целом были извлечены сущности, принадлежащие к одной из трех категорий (антропонимы, эргонимы и топонимы), далее для первых 50 частотных единиц были построены семантические сети, отражающие отношения между ними. На следующем этапе работы упомянутые имена собственные были подвергнуты экспертному анализу с последующей кластеризацией, позволившей, во-первых, выделить дополнительные темы, которые не были выявлены на предыдущем шаге при помощи автоматической процедуры, а во-вторых, осуществить глубинный анализ. Результаты показывают превалирование привнесенных в медийное поле тематик, связанных с современной концепцией политического образования, в материалах общегерманской прессы, в то время как локальная тематика по большей части сконцентрирована на местной повестке. Автоматическое выделение именованных сущностей может рассматриваться как необходимый этап для последующего дискурсивного анализа, при этом полученный материал нуждается в дополнительной экспертной оценке.

Ключевые слова: немецкий язык, именованные сущности, кластеризация, корпус текстов, общегерманская пресса, региональная пресса, локальная пресса.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 24-28-00937, <https://rscf.ru/project/24-28-00937/>

Для цитирования: Корышев М.В., Хохлова М.В. Именованные сущности в немецкоязычной прессе: корпусный и экспертный анализ // Terra Linguistica. 2025. Т. 16. № 4. С. 59–73. DOI: 10.18721/JHSS.16404



NAMED ENTITIES IN THE GERMAN-LANGUAGE PRESS: CORPUS AND EXPERT ANALYSIS

M.V. Koryshev , M.V. Khokhlova 

St. Petersburg State University, St. Petersburg, Russian Federation

✉ m.khokhlova@spbu.ru

Abstract. The analysis of proper names mentioned in news texts is of particular research interest, as it allows for indirect identification of the topics covered in the publications. This article presents the results of an analysis of an automatic procedure for extracting named entities using material from the German-language press. The study was conducted on both national German publications aimed at a broad audience and regional and local newspapers aimed at a narrower audience in the federal states of Germany. The work was conducted in two stages: during the first stage, entities belonging to one of three categories (anthroponyms, ergonyms, and toponyms) were extracted from the texts of each publication, as well as from the entire article collection, using the Stanza tool. Semantic networks reflecting the relationships between these entities were then constructed for the first 50 frequent lexemes. In the next stage of the work, the aforementioned proper names were subjected to expert analysis and subsequent clustering, which allowed, firstly, the identification of additional themes not identified in the previous step using the automated procedure, and secondly, the implementation of an in-depth analysis. The results show the prevalence of themes introduced into the media field related to the modern concept of political education in national press materials, while local themes were largely concentrated on the local agenda. Automatic identification of named entities can be considered a necessary step for subsequent discourse analysis, although the resulting material requires additional expert evaluation.

Keywords: German language, named entities, clusterization, text corpora, national German text-based media, regional press, local press.

Acknowledgements: The research was carried out at the expense of the Russian Science Foundation grant No. 24-28-00937, <https://rscf.ru/project/24-28-00937/>

Citation: Koryshev M.V., Khokhlova M.V., Named Entities in the German-Language Press: Corpus and Expert Analysis, *Terra Linguistica*, 16 (4) (2025) 59–73. DOI: 10.18721/JHSS.16404

Введение

Автоматическое распознавание именованных сущностей (англ. named entity recognition) относится к одной из актуальных задач более широкого направления в компьютерной лингвистике — извлечения информации [1]. Под именованными сущностями понимаются имена, названия мест и организаций, которые упоминаются в текстах. Полученные данные могут быть структурированы и внесены в базу данных с целью дальнейшего применения в разнообразных приложениях. Соответствующая информация традиционно востребована в поисковых и в вопросно-ответных системах (например, [2, 3]) или при суммаризации материала (например, [4, 5]), однако заслуживает отдельного изучения при анализе средств массовой информации, поскольку связана с затрагиваемыми в них темами и может указывать на интересы читателей (или редакторов). Т.В. Шмелева [6] подчеркивает важность изучения ключевых слов, среди которых выделяет категорию «субъекты», при описании жизни общества, а также отношения граждан к текущей политической ситуации. Л. Цонева [7] рассматривает ключевые имена в медиадискурсе на материале российских СМИ, уделяя внимание политическим деятелям и предлагая для их обозначения термин ключевые имена, отмечая тем самым их значение. Таким образом именованные сущности могут служить важной приметой времени, создавая его портрет [8, 9].



Исследования ономастического материала проводились в разных работах¹ [10; 11], однако представляют отдельный интерес при исследовании медиадискурса ввиду их субъектной отнесенности. В нашей работе проведен анализ, направляемый корпусом, предложенный Дж. Синклером [12] и развиваемый в настоящее время в традициях дискурсивного подхода (см., например, [13]).

Являясь широким семантическим классом, имена собственные включают целый ряд разнородных объектов. Исследователи сходятся во мнении, что к ним можно причислить, например, антропонимы, эргонимы, топонимы, зоонимы или космонимы² [11], однако в рамках обработки текстов на естественном языке ставится более узкая задача. Традиционно используется классификация по трем категориям, подразумевающая разделение на антропонимы, эргонимы и топонимы, тем не менее дополнительно в ходе проведения процедуры алгоритмы могут маркировать и иные типы информации, например темпоральные выражения (даты) или числовые комплексы (суммы, индексы, номера телефонов). Решение задачи по автоматическому выделению именованных сущностей сводится к следующим процедурам: во-первых, определение, является ли найденная единица таковой, и, во-вторых, в случае положительного ответа — установление для нее соответствующей метки.

В рамках нашего исследования извлечение лексических единиц, обозначающих имена собственные, было произведено на материале немецкоязычных новостных текстов с целью определения круга потенциальных тем, которые в них затрагиваются эксплицитно или имплицитно, а также векторов интересов аудитории. При междисциплинарном подходе, реализуемом в настоящей работе, собственно лингвистической исследовательской задачей является анализ места именованных сущностей в дискурсивных практиках разноуровневых печатных средств массовой информации.

Поскольку в ФРГ важную роль играют федеральные земли, обладающие автономией по обширному кругу вопросов, тогда как за федеральным центром закреплены вопросы внешней политики обороны и т.п., что обусловлено исторически (единая Германия сложилась в 1871 году именно как федеральный союз во главе с императором немцев, а не Германии, и эта империя была в значительной мере децентрализована в культурном и экономическом отношении), то медийный ландшафт этой страны отражает особенности ее государственного устройства: как и в любой стране, имеются средства массовой информации общегосударственного уровня, однако сильными игроками на этом поле являются и региональные издания, освещающие как общегерманскую, так и региональную повестку с позиции того или иного культурно-экономического ландшафта, и локальные издания, являющиеся рупором умонастроений и интересов жителей отдельного города или муниципального образования. Таким образом речь идет о трех категориях прессы, которые нацелены как, с одной стороны, на разную тематику (или на ее отражение на страницах печати в разном процентном соотношении), так и, с другой стороны, на разную аудиторию.

Если сравнить ландшафт печатных СМИ в Германии и России, то необходимо отметить две особенности: 1) децентрализованность — если в Российской Федерации общегосударственные СМИ связаны с Москвой как столицей страны, то германские СМИ этого уровня выпускаются в самых разных регионах этой страны; 2) большая значимость региональных и локальных СМИ в медийном ландшафте Германии как отражение исторически обусловленного своеобразия страны, о котором говорилось выше; 3) большая приверженность к чтению печатных изданий германской аудитории в сравнении с Россией (чтение печатной газеты как культурный ритуал).

¹ Супрун В.И. Ономастическое поле русского языка и его художественно-эстетический потенциал: дисс. ... д-ра филол. наук: специальность 10.02.01. Волгоград, 2000; Фоякова О.И. Имя собственное в художественном тексте: учебное пособие Л.: ЛГУ, 1990. 103 с.

² Мадиева Г.Б., Супрун В.И. Теория и практика ономастики: учебное пособие. Алматы: Қазақ университеті; Волгоград: Перемена, 2015. 152 с.



С учетом особенностей германского медийного ландшафта будет справедливым утверждение: изучение процессов становления общественного мнения как комплексной сущности, отражающей как умонастроения населения, так и установки, формируемые в рамках политического образования государственными и квазигосударственными акторами («СМИ как голос населения» и «СМИ как школа политического образования»), подводит нас к анализу материалов трех вышеупомянутых групп печатных СМИ, что позволит выявить общее ядро, в котором будут совмещены умонастроения германского общества, созвучные идеям политического образования, а также периферийные зоны, которые дадут представление о умонастроениях населения как на общегерманском, так и на регионально-локальном уровнях.

Материал и методология исследования

В качестве источника общегерманского печатного СМИ нами было рассмотрено такое крупное издание, как “Die Zeit”³. Общий объем собранных текстов составил более 30 тыс. статей (5,5 млн словоупотреблений), период охвата материала – с 1997 по 2024 год по разным тематическим разделам, доступным на сайте. Региональная и локальная пресса представлена девятью изданиями, отражающими разнообразие германского культурно-экономического ландшафта: север Германии – “Lübecker Nachrichten”⁴, обширный нижнесаксонский регион (север центральной части страны) – “Niedersächsische Tageszeitung”⁵ и таким локальным изданием, как “Neue Osnabrücker Zeitung”⁶, западные регионы страны – “Westdeutsche Allgemeine Zeitung”⁷, рейнский регион – “Rheinische Post”⁸, а также “Kölner Stadt-Anzeiger”⁹ как локальное издание, восток Германии – “Thüringer Allgemeine”¹⁰, юго-западные регионы страны – “Stuttgarter Allgemeine”¹¹ и, наконец, юго-восточный культурный ландшафт (Бавария) – “Augsburger Allgemeine”¹². Нами были собраны тексты общим объемом более 11 тыс. статей (4,5 млн словоупотреблений).

Для немецкого языка существует целый ряд систем, в том числе мультязычных, которые позволяют автоматически извлекать именованные сущности (например, SpaCy¹³ или Stanza¹⁴). Для дальнейшей работы нами был выбран фреймворк Stanza, который основан на нейросетевых алгоритмах, что дает возможность их эффективно обучать на пользовательских аннотированных данных, и показывает высокие результаты на немецкоязычных датасетах German CoNLL-2003 и GermEval2014.

Применение методов автоматического извлечения именованных сущностей было этапом анализа, направляемого корпусом. Выделенные единицы были отнесены к одной из трех категорий: LOC – топоним; ORG – название организации; PER – персоналия. Маркером-показателем значимости той или иной единицы выступила частотность. Далее была произведена экспертная кластеризация для первых 50 позиций в списках для каждого из описанных выше источников. Метод экспертного анализа, примененный в настоящей работе, заключается в определении позиций экспертов с последующим выявлением на этом основании информации, подвергаемой в свою очередь интерпретационному анализу. Экспертами выступали исследователи-германисты политологического и филологического профилей. Дополнительно было проведено сравнение Топ-250 лексем по всем изданиям.

³ <https://www.zeit.de>

⁴ <https://www.ln-online.de/>

⁵ <https://www.haz.de/>

⁶ <https://www.noz.de/>

⁷ <https://www.waz.de/>

⁸ <https://rp-online.de/>

⁹ <https://www.ksta.de/>

¹⁰ <https://www.thueringer-allgemeine.de/>

¹¹ <https://www.stuttgarter-zeitung.de/>

¹² <https://www.augsburger-allgemeine.de/>

¹³ <https://spacy.io/>

¹⁴ <https://stanfordnlp.github.io/stanza/>



Результаты

Всего для изданий были сформированы пять потенциальных кластеров, которые соответствуют тематической составляющей СМИ (локальная, общегерманская и общеевропейская тематики, международная повестка дня, финансы). Выделение финансовой тематики в отдельный кластер продиктовано данными материала: соответствующие лексемы нередко открывают частотные списки, полученные при анализе отдельных печатных изданий. Противопоставление европейской и международной проблематики связано с тем, что ФРГ, входя в ЕС, является частью европейского политического организма. В результате чего между этой страной и другими странами — членами ЕС выстраиваются иные взаимоотношения, чем с государственными акторами, членами ЕС не являющимися. Таким образом, под общеевропейской тематикой подразумевается круг понятий, соотносимых с ЕС и входящими в него странами.

Автоматический анализ именованных сущностей

Количество сформированных кластеров, а также распределение входящих в них лексем зависит от типа новостного источника. Прежде всего обратимся к анализу распределения всех выделенных именованных сущностей по вышеупомянутым категориям (рис. 1)¹⁵.

Большинство выделенных единиц (45,70%) относится к названиям территорий и населенных пунктов: среди первых наиболее частотными оказываются сама Германия (*Deutschland*) и государства, являющиеся важными игроками на мировой арене (*USA*, *Russland*), в то время как вторые представлены упоминанием столицы крупной земли Баден-Вюртемберга (*Stuttgart*), столицы государства (*Berlin*) и названия одного из центров Рурского региона (*Bochum*). Около трети сущностей (30,34%) именует персон, среди которых наиболее часто упоминаются иностранные лидеры (*Trump*, *Putin*, *Erdoğan*), что подчеркивает важность внешнеполитической повестки в рассматриваемых новостях, в то время как оставшиеся единицы (23,96%) обозначают организации: наиболее часто встречаются упоминания политических партий (таких как *SPD* и *CDU*) и военно-политического союза (*NATO*).

Далее для типов PER и ORG были построены семантические сети связей по типу графа, в которых вершинами являются сущности, а ребрами — связи между ними (упоминания в одном контексте). Размер узла зависит от частоты сущности в источнике, в то время как толщина ребра — от частоты совместной встречаемости единиц (чем она выше, тем сильнее связь и, следовательно, больше толщина ребра). Для более удобной визуализации было решено отфильтровать данные по количеству узлов и весу ребер: оставлены 50 наиболее частотных сущностей (узлов) и ребра с весом 2 и более (то есть по крайней мере две цитаты, в которых встречаются лексемы-узлы).

Рассмотрим полученный граф на примере издания “*Kölner Stadt-Anzeiger*” (рис. 2), в котором сущности, относящиеся к разным категориям, обозначены разными цветами (красным цветом — антропонимы, в то время как синим — эргонимы).

Приведенный семантический граф позволяет выявить два четко сформированных кластера, в которые включены лексемы, тематически связанные с политикой и футболом. Также можно отметить отдельные узлы, не связанные с другими ребрами, но при этом достаточно частотные и соответствующие первым 50 единицам в ранжированном списке (например, *BMW*, *ARD*, *Hamas*).

В первом кластере присутствуют упоминания как немецких, так и иностранных политиков (*Merz*, *Putin*, *Rutte*, *Trump*, *Selensky*), партий (*CDU*, *CSU*), военно-политического блока (*NATO*), а также новостного издания (*Bild*), освещающего знаковые события. Так, в выпусках газеты за август 2025 г. основное внимание было уделено встрече двух мировых лидеров на Аляске, что объясняет сильную связь между соответствующими узлами в графе¹⁶:

¹⁵ Обозначения на рис. 1: AA — “Augsburger Allgemeine”; KSA — “Kölner Stadt-Anzeiger”; LN — “Lübecker Nachrichten”; NOZ — “Neue Osnabrücker Zeitung”; NT — “Niedersächsische Tageszeitung”; RP — “Rheinische Post”; SA — “Stuttgarter Allgemeine”; TA — “Thüringer Allgemeine”; WAZ — “Westdeutsche Allgemeine Zeitung”; Zeit — “Die Zeit”.

¹⁶ Приведенные здесь и далее цитаты не отражают позицию авторов и использованы в иллюстративных целях.

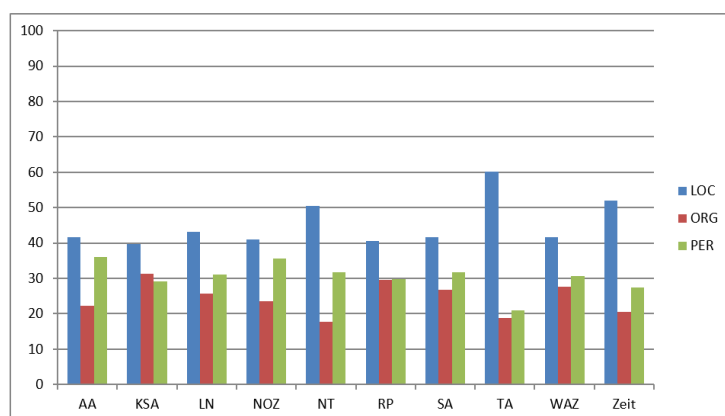


Fig. 1. Distribution of named entities by category in the news

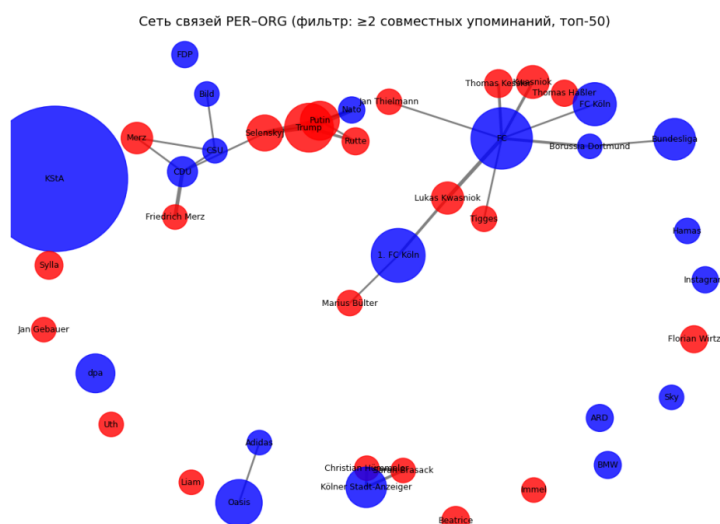


Рис. 2. Семантический граф для издания “Kölner Stadt-Anzeiger”
Fig. 2. Semantic graph for the edition of “Kölner Stadt-Anzeiger”

Am kommenden Freitag will sich **Trump** persönlich mit **Putin** treffen – nicht auf neutralem Boden, wie zuvor spekuliert, sondern im US-Bundesstaat Alaska.

В следующую пятницу **Трам** планирует лично встретиться с **Путиным** — не на нейтральной территории, как предполагалось ранее, а в американском штате Аляска.

Trump machte dies später jedoch nicht zur Voraussetzung, um sich mit **Putin** zu treffen.

Однако позднее **Трам** не стал делать это обязательным условием для встречи с **Путин**ым.

*Am Freitag treffen US-Präsident **Donald Trump** und der russische Präsident **Wladimir Putin** in Alaska aufeinander.*

В пятницу президент США Дональд Трамп и президент России Владимир Путин встретятся на Аляске.



Вместе с тем отмеченная на графе связь иностранных политиков с партиями ФРГ объясняется на самом деле упоминанием первых в одном контексте с немецкими политическими деятелями, для которых в новостях указывается партийная принадлежность, так что эта связь оказывается в силу ее опосредованного характера мнимой в содержательном плане:

Macron hatte sich zuvor mit Selenskyj, Bundeskanzler Friedrich Merz (CDU) und dem britischen Premierminister Keir Starmer ausgetauscht.

Ранее **Макрон** обменивался мнениями с Зеленским, канцлером Фридрихом Мерцем (ХДС) и премьер-министром Великобритании Киrom Стармером.

Второй кластер представлен лексемами, отражающими спортивную тематику: упоминаются футбольные клубы и лига (*FC, 1. FC Köln, FC Köln, Borussia Dortmund, Bundesliga*), а также спортсмены (*Jan Thielmann, Tigges, Thomas Kessler*). Размеры узлов и сила связей между ними позволяют сделать предположение, что освещению футбольных событий уделяется большее внимание по сравнению с первой темой в данном издании и что футбольная тематика действительно занимает читателей, в то время как вовлечение читателей в политический контекст — дань привычному видению газетного издания как органа политического образования. Это мы можем наблюдать по тем пропорциям читательской вовлеченности, где субъективно значимое увлечение — футбол — занимает более видное место, чем, казалось бы, общезначимая политическая тематика.

Ниже дадим краткое описание автоматически выделенных кластеров в некоторых изданиях (полный перечень тем и их упоминание в источниках приведены в табл.).

Так, несмотря на имеющиеся узлы, соответствующие сущностям, которые обозначают политических деятелей, в семантической сети для издания “Augsburger Allgemeine” был сформирован только один кластер, в который вошли лексемы, имеющие отношение к внутренней политике ФРГ. К ним относится упоминание федерального канцлера (*Friedrich Merz*) и партий (*CDU, SPD, Union*).

Таблица. Автоматически выделенные кластеры в новостях
Table. Automatically selected clusters in the news

	Германская внешняя политика	Германская внутренняя политика	Спорт	Локальные новости (муниципальное образование)	Экономика
“Augsburger Allgemeine”		+			
“Kölner Stadt-Anzeiger”	+		+		
“Lübecker Nachrichten”		+		+	
“Neue Osnabrücker Zeitung”	+		+	+	+
“Niedersächsische Tageszeitung”				+	+
“Rheinische Post”	+	+			
“Stuttgarter Allgemeine”	+	+	+		+
“Thüringer Allgemeine”				+	+
“Westdeutsche Allgemeine Zeitung”	+	+	+		
“Die Zeit”	+	+			
Объединенные источники	+	+	+		+



В семантической сети, построенной для издания “Neue Osnabrücker Zeitung”, четко выделены кластеры, посвященные внешнеполитической жизни ФРГ, экономическим вопросам, футболу и событиям местного масштаба. В первой группе больше внимания (по сравнению с остальными изданиями) уделяется военной составляющей: в текстах встречаются упоминания Североатлантического альянса (*NATO/Nato*) и вооруженных сил ФРГ (*Bundeswehr*). Отметим, что обращение к указанной теме связано с иными странами (например, в контексте дружеского взаимодействия с Израилем и сдерживания потенциальных угроз со стороны России и Китая).

*Sie übersehe auch, „wie wichtig die sicherheitspolitische Zusammenarbeit mit Israel für Deutschland ist, um die **Bundeswehr** und die **Nato** zu stärken“.*

*Она также упускает из виду, «насколько важно для Германии сотрудничество с Израилем в области политики безопасности для укрепления **бундесвера** и **НАТО**».*

*Hunderte Male im Jahr müssen **Nato**-Piloten aufsteigen, um **russische** Militärflugzeuge in den eigenen Luftraum zurückzuführen.*

*Сотни раз в год пилотам **НАТО** приходится вылетать, чтобы вернуть **российские** военные самолеты в свое воздушное пространство.*

*Je mehr Deutschland und die **Nato** sich auf einen bewaffneten Konflikt mit **Russland** vorbereiten, desto mehr wird aus dem wunderschönen Ostseeraum wieder ein strategischer Knackpunkt jedweder Kriegsüberlegungen.*

*Чем больше Германия и **НАТО** готовятся к вооруженному конфликту с **Россией**, тем больше прекрасный регион Балтийского моря снова становится стратегическим камнем преткновения в любых военных планах.*

*Ein Szenario: Bei einem Konflikt drängen die gut ausgestatteten **russischen** Seestreitkräfte samt Luftunterstützung die **Nato** kurzerhand aus der Ostsee.*

*Один из сценариев: в случае конфликта хорошо оснащенные **российские** военно-морские силы, включая поддержку с воздуха, просто вытеснят **НАТО** из Балтийского моря.*

*Im Bericht des Geoinformationszentrums der **Bundeswehr** heißt es: „Trotz bestehender Kooperationsbemühungen, etwa bei Such- und Rettungsoperationen, wachsen die Spannungen weiter an, auch bedingt durch **Chinas** wachsenden Einfluss und seine Ambitionen in der Arktis“.*

*В докладе Геоинформационного центра **Вооруженных сил Германии** говорится: «Несмотря на существующие усилия по сотрудничеству, например, в поисково-спасательных операциях, напряженность продолжает расти, что отчасти связано с растущим влиянием и амбициями **Китая** в Арктике».*

Экономические новости представлены упоминанием судебного разбирательства между двумя крупными производителями напитков *Paulaner* и *Berentzen*:

***Paulaner** hat sich im Rechtsstreit um das Design einer Cola-Mix-Flasche gegen **Berentzen** durchgesetzt.*

*Компания **Paulaner** одержала победу над компанией **Berentzen** в судебном споре по поводу дизайна бутылки для напитка *Coca-Cola*.*

*Das Haselünner Unternehmen **Berentzen** hat angekündigt, Berufung gegen das Urteil des Münchener Landgerichtes einzulegen.*

*Базирующаяся в Хазелюнде компания **Berentzen** объявила, что подаст апелляцию на решение Мюнхенского земельного суда.*



Традиционно для ФРГ уделяется внимание вопросам «зеленой» энергии, которые обнаруживаются в найденных контекстах в силу частотности упоминания бизнесмена (*Christoph Pieper*):

Christoph Pieper erläuterte, man bemühe sich um weitere Kunden für die Wasserstofftechnik.

Кристоф Пипер пояснил, что предпринимаются усилия по привлечению дополнительных клиентов для водородных технологий.

Pieper beklagte aber, dass erhebliche formale, rechtliche und bürokratische Hindernisse bestünden, die den Hochlauf der Wasserstofftechnik behinderten.

Однако Пипер посетовал на существование серьезных формальных, юридических и бюрократических препятствий, которые мешают развитию водородных технологий.

Und da sandte Christoph Pieper von Cec Haren bei einer Rundfahrt zum Speicherfeld in Wesuwe, zur H2-Tankstelle im Industriegebiet und dann zum Bürgerwindpark Fehndorf/Lindloh durchaus gemischte Signale.

В этот раз Кристоф Пипер из СЕС Харен подал неоднозначную оценку во время осмотра хранилища в Везуве, заправочной станции H2 в промышленной зоне, а затем ветровой электростанции в районе Фендорф/Линдло.

Примечательно, что обсуждаемые местные новости носят умиротворяющий характер: присутствует рассказ о животных, которые нуждаются в доме (кролики *Bandit* и *Blanca*). Можно сказать, что это является показателем «немецкости» жизни — того, что уделяется особое внимание тематике, которую можно было бы определить как обывательскую.

Als Blanca und Bandit ins Tierheim Melle gebracht wurden, kamen sie direkt in eine Pflegestelle denn sie mussten alle 3 Stunden mit Aufzuchtmilch versorgt werden.

Когда Бланку и Бандита привезли в приют для животных Мелле, их сразу же поместили в приемную семью, поскольку им нужно было давать молоко каждые три часа.

Также на страницах приводится рассказ о котах *Angel* и *Aurelio*, которые нуждаются в новом доме:

Zwei besondere Katzen: Angel & Aurelio brauchen ein Zuhause.

Два особенных кота: Ангел и Аурелио нуждаются в доме.

Ein Zuhause für zwei: Was Angel und Aurelio glücklich macht.

Дом для двоих: что делает Ангела и Аурелио счастливыми.

Gesucht wird ein gemeinsames Zuhause für Angel und Aurelio.

Мы ищем общий дом для Ангела и Аурелио.

В следующей цитате речь идет о бульдоге по кличке *Shamo*:

Shamo lebte vorher unerlaubt in Meller Notunterkunft, bis es dort brannte. Seit dem vergangenen Herbst lebt der Hund im Tierheim Melle.

Шамо ранее незаконно проживал в приюте для животных в Мелле, пока там не случился пожар. С прошлой осени собака живет в приюте для животных Мелле.



В издании “Lübecker Nachrichten” неожиданно большое внимание уделяется развлекательно-му материалу и связанным с ним местным новостям: так, освещается проведение крупнейшего в мире фестиваля тяжелой музыки *Wacken Open Air*, который традиционно проходит в Ваккене (город в земле Шлезвиг-Гольштейн). Вместе с тем в новостных текстах заслуживают упоминания крупные западные развлекательные корпорации, продукция которых транслируется в ФРГ (*Netflix*, *Amazon Prime Video*), и их продвижение на немецкоязычном рынке:

Mit eigenen Produktionen, Originals genannt, möchte der Ende 2019 gestartete Streamingsservice Apple TV+ den Platzhirschen Netflix und Amazon Prime Video, aber auch neueren Anbietern wie Disney+ Konkurrenz machen.

Стриминговый сервис *Apple TV+*, запущенный в конце 2019 года, предлагает собственный проект под названием *Originals* и стремится конкурировать с лидерами рынка *Netflix* и *Amazon Prime Video*, а также с новыми провайдерами, такими как *Disney+*.

Portale wie Netflix und Spotify boten nicht nur für leistbare zehn Euro im Monat einen riesigen Katalog an Filmen, Serien oder Musik.

Такие порталы, как *Netflix* и *Spotify*, предлагают огромный каталог фильмов, сериалов и музыки за доступные десять евро в месяц.

Несмотря на локальный характер издания “Stuttgarter Allgemeine”, тематически оно приближается к региональным источникам: в нем выделяются уже упомянутые выше значимые темы, однако они характеризуются бóльшим весом входящих в них единиц (то есть соответствующих лексем в новостях). Земля Баден-Вюртемберг является одним из крупнейших экономически развитых регионов ФРГ, поэтому неудивительным оказывается высокая частотность в рассмотренных контекстах названий автомобильных компаний (*Mercedes*, *Porsche*, *BMW*, *VW*) и в целом темы, связанной с автомобилестроением:

Natürlich gehört es zu den Aufgaben von Konzernchefs, Zuversicht zu verbreiten — doch es ist mehr als nur Zweckoptimismus, wenn Porsche-Chef Oliver Blume darauf hinweist, wie ertragsstark Porsche nach wie vor dasteht, wenn man auch nur die Kosten für den Personalabbau und für die US-Zölle einmal außer Acht lässt.

Конечно, в задачи руководителей корпораций входит распространение уверенности, но это больше, чем просто целесообразный оптимизм, когда генеральный директор *Porsche* Оливер Блум указывает, насколько прибыльной продолжает быть компания *Porsche*, даже если игнорировать расходы на сокращение персонала и пошлины США.

Porsche wie auch Mercedes haben überdies längst ihre allzu optimistische Einschätzung des Markts der vollelektrischen Autos korrigiert und pragmatisch umgesteuert.

Более того, и *Porsche*, и *Mercedes* давно скорректировали свою чрезмерно оптимистичную оценку рынка полностью электрических автомобилей и прагматично изменили курс.

Автоматическая обработка всех данных позволила объединить сущности, встречающиеся в разных источниках, и построить семантическую сеть, в которую входят четыре крупных кластера. Первые два связаны с внешней и внутренней политикой ФРГ: упоминаются антропонимы (*Trump*, *Putin*, *Erdogan*, *Biden*, *Selenskyj*, *Obama*, *Assad* и *Merkel*) и эргонимы (*Bundeswehr*, *NATO*, *UN*, *AfD*, *FDP*, *CDU*, *Linke*, *Grüne*, *Union* и *Hamas*). Упоминание крупнейших автоконцернов ФРГ (*BMW*, *Mercedes*, *VW*) составляет третью тему, в то время спортивные новости позволили выделить последнюю (*VfB*, *VfL*, *VfL Bochum*, *MSV Duisburg*, *Bundesliga*, *Borussia Dortmund*).



Вместе с тем из 250 наиболее частотных единиц общими для рассмотренных новостных источников оказались следующие: *Berlin*, *Deutschland*, *deutsch*, *Europa*, *CDU*, *USA*, *Frankreich*. Вполне ожидаемым представляется наличие в этом списке упоминания самой страны (*Deutschland*), соответствующего прилагательного (*deutsch*) и столицы (*Berlin*), а также одной из крупнейших политических партий ФРГ (*CDU*), к которой принадлежат два федеральных канцлера, занимавшие этот пост в течение двух последних десятилетий (Ангела Меркель и Фридрих Мерц). Бесспорно, важными являются для политической жизни государства его внешние связи со странами-партнерами: в этой связи наиболее значимыми выступают США (*USA*) как приоритетный трансатлантический партнер и Европа в целом (*Europa*), где Франции (*Frankreich*) уделяется особое внимание ввиду ее историко-географической близости к Германии.

Экспертный анализ Топ-50 именованных сущностей

Общегерманская пресса представлена текстовым материалом крупного ежедневника “*Die Zeit*”. Прежде всего следует отметить, что в этом издании среди первых 50 лексем никак не представлена локальная тематика. Разве что значительной частотностью обладает *Berlin* (986 вхождений), однако Берлин здесь — не локальный актер, а столица германского государства. Характерно, что общеевропейская тематика ожидаемо включает лексемы, именующие ЕС и континент. При этом лексема *EU* (1930 вхождений) встречается на треть чаще, чем *Europa* (1237 вхождений), что подводит нас к выводу, что принадлежность к Европе определяется через принадлежность к ЕС, а не наоборот. Кроме того, примечательно, что внутриевропейские связи наиболее часто описываются в контексте германо-польских и германо-французских отношений, что видно по частотности использования лексем *Frankreich* (477 вхождений) и *Polen* (314 вхождений), тогда как названия таких крупных немецкоязычных соседей, как *Österreich*, *Schweiz*, в число 50 частотных лексем не входят, что тоже объяснимо, если учитывать особое положение этих двух государств: Австрия не входит в НАТО и не отличается сильным еврооптимизмом, а Швейцария сохраняет нейтралитет. Среди лексем, отнесенных к общегерманской тематике, внимания заслуживают наименования партий: помимо двух крупнейших, упоминаются критикуемая истеблишментом *AfD* (394 вхождения), либерально-буржуазная *FDP* (9364 вхождения) и леволиберальная партия [*Bündnis 90/*]/*Die Grünen* (315 вхождений). Отметим, что частотность использования лексем не соответствует реальному балансу политических сил, что позволяет выдвинуть гипотезу о главенстве принципа «школы политического образования» с целью трансляции либеральной повестки в общественные массы. Общемировая тематика явно акцентирует Россию и США как крупнейших игроков на мировой арене с вниманием к таким акторам, как Китай и Турция, при отсутствии интереса к Индии, а также пристальное внимание к внешнеполитической проблематике, связанной с Украиной, и к роли Североатлантического альянса в текущей политической повестке. Финансово-экономическая тематика представлена наименованиями таких валют, как евро и доллар, при отсутствии наименований иных валют среди частотных лексем, что свидетельствует при всем внимании к экономике об однополярном трансатлантическом характере этого экономического мышления.

Такое региональное издание, как “*Rheinische Post*”, характеризуется менее выраженным, чем в случае с “*Die Zeit*”, интересом к внешнеполитической повестке: 21 лексическая единица в “*Rheinische Post*” с общим числом 510 вхождений против 27 единиц в “*Die Zeit*” с общим числом 18063 вхождения. Внешнеполитический акцент в этом издании делается на связях с США, однако при интересе к Украине Россия отходит на задний план, впереди оказывается интерес к Китаю и Бразилии, а также объяснимое актуальной повесткой внимание к Израилю и Сектору Газа. Следует отметить, что внешнеполитические акценты на этом региональном уровне оказываются смазанными, что выражается в присутствии среди первых 50 частотных лексем совершенно случайных единиц (претендент на победу в колумбийской президентской гонке Густаво Петро, король-эмерит Испании Хуан Карлос, упоминаемый в связи с публикацией



его мемуаров). Такая калейдоскопичность наводит на мысль о второстепенности внешнеполитической повестки для региональной прессы при сохранении трансатлантического вектора, что соответствует и идеям политического образования, и реально существующим настроениям в обществе. Европейская повестка представлена слабо — все тот же ЕС, превалирующий над Европой (48 против 14 вхождений). Это свидетельствует об укорененности мысли о том, что границы Европы определяются все же границами ЕС — на региональном уровне становится ясным, что Европейский Союз не только становится реальным актором на несколько абстрактном внешнеполитическом уровне, но и определяет через законодательно-бюрократическую машину отдельного государства повседневность конкретного региона этой страны. Если анализировать лексемы, относящиеся к общегерманской тематике, то характерной особенностью регионального уровня является появление среди частотных единиц упоминаний крупных игроков на информационном поле (*ARD, ZDF*), политиков, не являющихся первыми лицами государства, но получившими в какой-то момент всплеск массмедийного внимания (например, министр жилья, городского развития и строительства Германии Верена Хуберц), а также Всеобщего немецкого автомобильного клуба (*ADAC*), что тоже интересно в связи с особым местом, которое автомобиль занимает в немецкой культуре. Таким образом, и на примере общенациональной лексики отмечается тяготение к мозаичности, совмещающей перспективы и такого конструкта, как «простой немец», реально существующего в совокупности населения этого региона, и того видения, которое задается идеями политического образования. Вероятно, этим и объясняется отсутствие среди частотных лексем наименования такой партии, как *AfD*: с одной стороны, ее позиции в прирейнском регионе относительно слабы, с другой — система политического образования не привносит в общественную повестку в этих условиях критики этой партии. Собственно локальная тематика в “*Rheinische Post*” представлена слабо, говорить можно о региональной тематике: среди первых 50 частотных лексем встречаются только наименование земли Северный Рейн-Вестфалия и имя собственное [Tina] Elsner (представитель полицейского ведомства, курирующий вопросы профилактики правонарушений преимущественно имущественного характера), имеющие по десять вхождений. Таким образом, региональная повестка сконцентрирована как на конкретной географической локации, что предсказуемо, так и на вопросах безопасности, которые получают в этом случае конкретно-личное звучание, связанное с защитой собственного имущества от посягательств. Говоря в завершение о кластере финансово-экономической проблематики, нельзя не отметить, что абстрактные наименования валют получают в “*Rheinische Post*” региональных конкурентов в виде компаний, активных в регионе (*WERDING, LEG, TKMS*). Таким образом, тенденция к калейдоскопичности/мозаичности, отмеченная выше, прослеживается и в этом кластере.

Особое место в палитре печатных средств массовой информации занимает “*Westdeutsche Allgemeine Zeitung*”. Внушительный тираж этого издания (более 275 000 экземпляров на фоне “*Rheinische Post*”, более 189 000 экземпляров) делает его одним из основных игроков на рынке региональной прессы, однако применительно к целям настоящего исследования в фокусе внимания оказывается высокий удельный вес регионально-локальной тематики: из 50 наиболее частотных именованных сущностей на долю лексем, относящихся к региональной тематике, приходится более 80%, тогда как лексемы общегерманской тематики составляют 8%, а на лексемы, соотносимые с внешнеполитической проблематикой, — 2%. Если принять во внимание приведенное выше сопоставление тиражей “*Westdeutsche Allgemeine Zeitung*” и “*Rheinische Post*”, то становится понятным, что такая тематическая структура “*Westdeutsche Allgemeine Zeitung*” является одним из факторов популярности этого издания, что свидетельствует о расположенности германского читателя к тому, чтобы уделять преимущественное внимание локальной тематике, оставляя за скобками не касающиеся его непосредственно события большой политики (ср. сопоставление Обломова и Штольца у Гончарова, а также идею о локальном



и муниципальном уровне участия населения в политической жизни страны как цели германского механизма политического просвещения).

Анализ прочих семи изданий, относящихся к группе локальных печатных СМИ, показывает, что они характеризуются преобладанием локальной тематики: на долю таких лексем среди первых 50 частотных единиц приходится в среднем 57%. Лексемы-репрезентанты этой группы расходятся по трем основным группам: топонимы, антропонимы и наименования событий (фестивали, концерты и т.п.). Внешнеполитическая тематика связана, прежде всего, с США и Украиной, Россия упоминается реже, ближневосточная тематика освещена только в “Augsburger Allgemeine”, тогда как “Thüringer Allgemeine” не имеет среди первых 50 наиболее частотных лексем единиц, номинирующих внешнеполитическую повестку дня. Финансово-экономическая тематика выражена преимущественно наименованием европейской валюты (*Euro*) и названиями фирм, отсутствие же наименования денежной единицы США можно рассматривать как свидетельство абсолютно локального горизонта экономических интересов. Таким образом, идея «гражданина мира», активно продвигаемая системой германского политического образования, остается не воспринятой на локальном уровне, где сильны, как и прежде, мелкобуржуазные тенденции.

Заключение

Анализ показал превалирование привнесенных в медийное поле тематик, связанных с современной концепцией политического образования, где целью ставится формирование «гражданина мира», вовлеченного в соответствующие бытийно-содержательные контексты, в материалах общегерманской прессы, с последующей трансформацией этой линии воздействия в концепцию формирования активно участвующего в общественной жизни гражданина на регионально-муниципальном уровне, что прослеживается на уровне региональной и локальной прессы. Другой тематический вектор связан с узко локальной тематикой, преобладающей на уровне локальной прессы, ярко высвечивающей обывательско-мелкобуржуазные устремления, также характерные для германского общества. Сказанное выше позволяет предположить, что частный обывательский мотив является превалирующим постольку, поскольку он коррелирует с прочно укоренившимися в немецком обществе ценностями, связанными с собственным «Я», своей автономией и ответственностью перед собой, «заботой о себе» и готовностью включаться в преобразовательные процессы на конкретно достигаемом локальном уровне, тогда как сравнительно новая модель «гражданина мира» оказывается привнесенной, не вполне интериоризованной германским социумом. Будущее покажет, насколько и с какой скоростью эта внешняя тематика, освещаемая общегерманскими печатными изданиями, будет осваиваться на частно-локальном уровне. Наблюдение за этими процессами и их осмысление — задача дальнейших лонгитюдных исследований.

В лингвистическом плане именованные сущности выступают ведущими компонентами, характеризующими дискурсивные практики, культивируемые издательскими домами и редакционными коллегиями. Поскольку именованные сущности имеют определенный денотат и поддаются автоматическому извлечению и анализу, то для исследования дискурсивных практик и установок с применением больших текстовых данных именованные сущности и являются тем материалом, который дает основу для получения доказательно верного результата.

Результаты работы продемонстрировали, что автоматическое выделение именованных сущностей может рассматриваться как необходимый этап для последующего дискурсивного анализа, поскольку полученный материал нуждается в дополнительной экспертной оценке. Дальнейшая работа будет направлена на установление тональности применительно к выделенным сущностям на материале разных источников.



СПИСОК ИСТОЧНИКОВ

1. Филиппова Е.А. Извлечение информации // Прикладная и компьютерная лингвистика / под ред. И.С. Николаева, О.В. Митрениной, Т.М. Ландо. М.: ЛЕНАНД, 2016. С. 211–232.
2. Moreda P., Llorens H., Saquete E., Palomar M. Combining semantic information in question answering systems // Information Processing & Management. 2011. Vol. 47, Iss. 6. P. 870–885. DOI: 10.1016/j.ipm.2010.03.008
3. Кораблинов В.О. Подготовка набора данных для вопросно-ответного поиска по базе знаний. Первый этап: сопоставление сущностей // Компьютерная лингвистика и вычислительные онтологии. 2020. № 4. С. 98–108. DOI: 10.17586/2541-9781-2020-4-98-108
4. Berezin S., Batura T. Named Entity Inclusion in Abstractive Text Summarization // Proceedings of the Third Workshop on Scholarly Document Processing, Gyeongju: Association for Computational Linguistics, 2022. P. 158–162.
5. Marek P., Müller Š., Konrád J., Lorenc P., Pichl J., Šedivý J. Text Summarization of Czech News Articles Using Named Entities // The Prague Bulletin of Mathematical Linguistics. 2021. No. 116. P. 5–26. DOI: 10.14712/00326585.012
6. Шмелева Т.В. Ключевые слова текущего момента // Collegium. 1993. № 1. С. 33–38.
7. Цонева Л. Имена и люди. Ключевые имена в медиадискурсе. Велико-Тырново, 2017. 196 с.
8. Baturina L.A., Lepikhov N.V., Panova E.P., Popova A.V., Karpova G.G., Gumerova L.Z. The newspaper heading and proper name are parts of medial space // Linguistics and Culture Review. 2021. Vol. 5, No. S4, 2021. P. 583–590. DOI: 10.21744/lingcure.v5nS4.1675
9. Шмелева Т.В. Кризис как ключевое слово текущего момента // Политическая лингвистика. 2009. Вып. 2 (28). С. 63–68.
10. Борисов И.В. Антропонимы, как картина личности. М.: Научное издание, 2016. 100 с.
11. Суперанская А.В. Общая теория имени собственного. М.: Книжный дом «Либроком», 2009. 366 с.
12. Sinclair J.M. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991. 179 p.
13. Чернявская В.Е. Деонтическое значение ключевого слова в дискурсе: westliche Werte (западные ценности) в немецком общественном пространстве как предмет анализа, направляемого корпусом // Terra Linguistica. 2025. Т. 16, № 1. С. 82–98. DOI: 10.18721/JHSS.16106

REFERENCES

- [1] Filippova Ye.A., Izvlecheniye informatsii [Information extraction], Prikladnaya i kompyuternaya lingvistika [Applied and computational linguistics], LENAND, Moscow, 2016, pp. 211–232.
- [2] Moreda P., Llorens H., Saquete E., Palomar M., Combining semantic information in question answering systems, Information Processing & Management, 47 (6) (2011) 870–885. DOI: 10.1016/j.ipm.2010.03.008
- [3] Korablinov V.O., Dataset Creation for Question Answering Over Knowledge Bases. First Stage: Entity Linking, Kompyuternaya lingvistika i vychislitelnyye ontologii [Computational Linguistics and Computational Ontologies], 4 (2020) 98–108. DOI: 10.17586/0000-0000-2020-4-98-108
- [4] Berezin S., Batura T., Named Entity Inclusion in Abstractive Text Summarization, Proceedings of the Third Workshop on Scholarly Document Processing, Gyeongju, Association for Computational Linguistics, 2022, pp. 158–162.
- [5] Marek P., Müller Š., Konrád J., Lorenc P., Pichl J., Šedivý J., Text Summarization of Czech News Articles Using Named Entities, The Prague Bulletin of Mathematical Linguistics, 116 (2021) 5–26. DOI: 10.14712/00326585.012
- [6] Shmeleva T.V., Klyuchevyye slova tekushchego momenta [Key words of the moment], Collegium, 1 (1993) 33–38.
- [7] Tsoneva L., Imena i lyudi. Klyuchevyye imena v mediadiskurse [Names and People: Key Names in Media Discourse], Veliko-Tyрноvo, 2017.
- [8] Baturina L.A., Lepikhov N.V., Panova E.P., Popova A.V., Karpova G.G., Gumerova L.Z., The newspaper heading and proper name are parts of medial space, Linguistics and Culture Review, 5 (S4) (2021) 583–590. DOI: 10.21744/lingcure.v5nS4.1675



- [9] **Shmeleva T.V.**, Krizis (Crisis) as the Key Word of the Present Moment, *Political Linguistics*, 2 (28) (2009) 63–67.
- [10] **Borisov I.V.**, Antroponimy, kak kartina lichnosti [Anthroponyms as a picture of personality], Nauchnoye izdaniye, Moscow, 2016.
- [11] **Superanskaya A.V.**, Obshchaya teoriya imeni sobstvennogo [General Theory of Proper Names], Knizhnyy dom “Librokom”, Moscow, 2009.
- [12] **Sinclair J.M.**, Corpus, Concordance, Collocation, Oxford University Press, Oxford, 1991.
- [13] **Chernyavskaya V.E.**, Deontic meaning of a discursive key word: corpus-assisted analysis of westliche Werte (western values), *Terra Linguistica*, 16 (1) (2025) 82–98. DOI: 10.18721/JHSS.16106

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Корышев Михаил Витальевич

Mikhail V. Koryshev

E-mail: m.koryshev@spbu.ru

ORCID: <https://orcid.org/0000-0001-8946-4431>

Хохлова Мария Владимировна

Maria V. Khokhlova

E-mail: m.khokhlova@spbu.ru

ORCID: <https://orcid.org/0000-0001-9085-0284>

Поступила: 13.09.2025; Одобрена: 11.11.2025; Принята: 17.11.2025.

Submitted: 13.09.2025; Approved: 11.11.2025; Accepted: 17.11.2025.