



БИБЛИОТЕКА

СТАТЬИ

Статьи в открытом доступе

ЖУРНАЛЫ

Кибернетика и программирование

Современное образование

Юридические исследования

Философская мысль

Социодинамика

Психолог

Человек и культура

Вопросы безопасности

Genesis: исторические исследования

Международное право

NB: Административное право и практика администрирования

Litera

Полицейская и следственная деятельность

Мировая политика

Финансы и управление

Теоретическая и прикладная экономика

Право и политика

Налоги и налогообложение

Политика и Общество

Философия и культура



Статья опубликована с лицензией [Creative Commons Attribution-NonCommercial 4.0 International License \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/) – Лицензия «С указанием авторства – Некоммерческая».

[Вернуться к содержанию](#)

Litera

Правильная ссылка на статью:

Грушина О.А. — Анализ современных медиатекстов на языке суахили с использованием методов корпусной лингвистики // Litera. – 2020. – № 1. – С. 247 - 256. DOI: 10.25136/2409-8698.2020.1.31993
URL: https://nbpublish.com/library_read_article.php?id=31993

Анализ современных медиатекстов на языке суахили с использованием методов корпусной лингвистики

Грушина Оксана Андреевна

аспирант, кафедра африканистики, Московский государственный университет имени М.В. Ломоносова

119991, Россия, г. Москва, ул. Ленинские Горы, 1

Grushina Oksana Andreevna

Postgraduate students, the department of African Studies, M. V. Lomonosov Moscow State University

119991, Russia, g. Moscow, ul. Leninskie Gory, 1

✉ o.a.grushina@gmail.com

[Другие публикации этого автора](#)



DOI:

10.25136/2409-8698.2020.1.31993

Дата направления статьи в редакцию:

19-01-2020

Дата публикации:

07-03-2020

Аннотация.

В статье рассматривается применение методов корпусной лингвистики для анализа современных медиатекстов на языке суахили. Язык СМИ репрезентативен и продуктивен для анализа и вычленения терминов и выражений, актуальных на данный момент времени и относящихся к различным сферам речепользования. Предметом исследования являются особенности выявления как единичных, так и составных медицинских терминов и выражений в медийных текстах на языке суахили путем анализа данных по частотности употребления n-грамм. Для проведения исследования автором составлен корпус медиатекстов, основой которого стали материалы интернет-издания Mtanzania Танзаниец, опубликованные в 2014–2019 гг. Для сбора материала, составления корпуса и анализа полученных данных были применены методы корпусной лингвистики и компьютерного анализа. Результаты проведенного исследования показали эффективность использования методики вычленения актуальных единичных и составных медицинских терминов путем анализа корпуса медиатекстов раздела Afya na Jamii Здравоохранение и социальная сфера при помощи системы UNIX. Неоспоримым преимуществом анализа корпуса медиатекстов является использование в них «живого» языка с лексическим наполнением, актуальным для современного периода времени. Полученные результаты могут найти в дальнейшем практическое применение: после обработки лексикографами и

- Административное и муниципальное право
- Психология и Психотехника
- Национальная безопасность / nota bene
- Международное право и международные организации / International Law and International Organizations
- Исторический журнал: научные исследования
- Культура и искусство
- Филология: научные исследования
- Педагогика и просвещение
- Полицейская деятельность
- Международные отношения
- Программные системы и вычислительные методы
- Тренды и управление
- SENTENTIA. European Journal of Humanities and Social Sciences
- Финансовое право и управление
- Урбанистика
- RHILHARMONICA. International Music Journal
- Конфликтология / nota bene
- Сельское хозяйство
- Исследования космоса
- Машиностроение
- Экологические исследования
- Геолог
- Биологические исследования
- Химические процессы
- Продовольственные технологии
- Электроника и

специалистами в области медицины выявленные термины и выражения могут использоваться для составления специализированных тематических словарей на языке суахили.

Ключевые слова: суахили, медиатекст, корпус, средства массовой коммуникации, язык масс медиа, словарь, терминология, n-граммы, медицинские термины, корпусная лингвистика

Abstract.

This article examines the application of corpus linguistics techniques for analyzing modern media texts in Swahili language. The language of mass media is representative and productive for the analysis and extraction of terms referred to the various spheres of speech use. The subject of this research is the peculiarities of determination of simple and compound medical terms and phrases in the Swahili language media texts by analyzing data on the frequency of using n-gram. To undertake this study, the author comprised the body of media texts based on the materials of Mtanzania newspaper published in 2014-2019. The methods of corpus linguistics and computer analysis were applied for collection of materials and analysis of the acquired data. The results of the conducted research demonstrated the effectiveness of the use of extraction technique if the relevant simple and compound medical terms by analyzing the body of media texts of the section Afya na Jamii, healthcare and social sphere using the UNIX system. An undeniable advantage of analysis of the body of media texts is the use of a "living" language with lexical saturation relevant for the present time. The obtained results may find future practical application: after processing by the lexicographers and experts in the area of medicine, the discovered terms and phrases can be used for composing specialized thematic dictionaries in Swahili language.

Keywords:

n-grams, terminology, dictionary, mass media language, mass media, corpus, mediatext, Swahili, medical terms, corpus-based linguistics

В условиях глобализации инфосферы и совершенствования информационных технологий, медийная сфера все больше становится неотъемлемой частью жизни современного человека. Масс медиа являются основным источником информации о событиях, происходящих в мире, которая транслируется массовой аудитории в форме дискретных единиц информационных потоков — медиатекстов — посредством различных каналов (печатных изданий, телевидения, радио, интернета). Одним из самых распространенных каналов генерирования и распространения медиатекстов является интернет, благодаря которому многие издания, ранее существовавшие только в печатном формате, обрели онлайн-версии, что помогло им расширить свою аудиторию.

Медийный язык является прекрасным примером современного «живого», а не «искусственного» языка, который регулярно обновляется и понятен массовому читателю. Язык СМИ репрезентативен и продуктивен для анализа и вычленения терминов и выражений, актуальных на данный момент времени и относящихся к различным сферам речепользования.

Данная работа посвящена описанию методов корпусной лингвистики, применяемых для выявления как единичных, так и составных медицинских терминов и выражений в языке суахили путем анализа данных по частотности употребления n-грамм. Выбор медицинской тематики обусловлен несколькими причинами:

- узкая специализация медицинской терминологии и одновременно необходимость того, чтобы она была понятна не только специалистам в области медицины;
- использование медицинских терминов в медиатекстах, рассчитанных на массовую аудиторию, предполагает владение реципиентами данной терминологией, и, соответственно, она относится к «живому» актуальному лексическому пласту;
- медиатексты интернет-версий СМИ обычно категоризированы, так, в нашем случае в издании Mtanzania *Танзаниец* выделен раздел Afya na Jamii *Здравоохранение и социальная сфера*, что позволяет получить релевантные медиатексты для составления корпуса.

Следует отметить, что в лексикографии языка суахили существуют лишь отдельные небольшие словари и терминологические словники по различным отраслям знания, составленные традиционным интуитивно-выборочным способом по причине отсутствия достаточных технических и экономических ресурсов, хотя разработке терминологии на языке суахили уделялось значительное внимание. Это во многом было обусловлено языковой политикой, принятой в Танзании после провозглашения независимости. Так, в 1967 г. был создан Национальный совет языка суахили (Baraza la Kiswahili la Taifa = BAKITA), задачами которого было развитие и поддержка языка, его стандартизация, а также разработка терминологии. Аналогичные задачи по укреплению роли суахили как государственного языка Танзании были возложены на Научно-исследовательский институт языка суахили (Taasisi ya Uchunguzi wa Kiswahili = TUKI), созданный в 1974 г. на базе Межтерриториального языкового комитета языка суахили. Таким образом, задача разработки терминологии велась в стенах государственных учреждений и достигла определенных успехов. До середины 1980-х гг. были выпущены словники, посвященные различным областям знания, как например заболеванием домашнего скота, агрономии, сельскохозяйственной инженерии, растениям и др. BAKITA также оказал содействие выпуску «Англо-суахили технического словаря» (1987), содержащего около 8000 технических терминов.

К середине 1980-х гг. деятельность по разработке терминологии была приостановлена по причине ослабления поддержки со стороны государства. В течение 20 лет не было издано ни одного терминологического словника. Работа возобновилась лишь в 2003 г., когда были составлены терминологические списки слов, относящихся к тематической категории «СПИД и другие заболевания», а в 2004 г. — терминологические списки по литературе, психиатрии и другим областям знаний [1]. Активную работу по составлению и выпуску терминологических словников вел также TUKI, который выпустил ряд работ: «Терминология автомобильных и тракторных механизмов» [2], «Англо-суахили словарь по экономике и торговле» [3], «Англо-суахили словарь юридической терминологии» [4], «Медицинский словарь» [5], «Словарь по истории» [6] и другие.

электротехника

Энергетический вестник

Транспортный вестник

Вопросы
здравоохранения

Арктика и Антарктика

Ресурсы Мирового
океана

Историческая
информатика

Архитектура и дизайн

КНИГИ

Все книги

К сожалению, работа по созданию терминологии, которая велась в стенах государственных специализированных учреждений, охватывала лишь некоторые области жизнедеятельности общества и по скорости отставала от стремительных изменений в области науки, техники, экономики, политической, социальной и культурной областей. Данные учреждения не обладали достаточными ресурсами для своевременного генерирования всего объема актуальной терминологии, отражающей современные реалии. Все это привело к созданию неофициальной терминологии, которая рождалась в среде тех, кого непосредственно затрагивали изменения и нововведения. Объем терминологии, появившейся неформально, спонтанно за последние годы, неуклонно растет. Данные термины требуют проверки и стандартизации со стороны специалистов [1]. Однако вплоть до настоящего времени специализированные учреждения не вовлечены активно в данный процесс, что приводит к некоему антагонизму официальной и неофициальной терминологии. Эксперты в области лингвистики рассматривают термины, созданные неофициально, как идущие вразрез с общепринятыми правилами и не соответствующие принципам стандартизации. В свою очередь, те, кого непосредственно касаются изменения и нововведения, так называемые «практики», вынуждены восполнять пробелы и генерировать термины, отражающие современные реалии, при этом они не используют официальные термины, которые создаются с опозданием или кажутся им странными для номинации понятий в соответствующих областях [2]. Таким образом, в настоящее время отсутствует формализованный эффективный процесс создания стандартизированной актуальной терминологии, своевременно фиксирующей современные реалии. Данные процессы привели к тому, что в суахили для номинации одного и того же понятия может существовать несколько наименований, которые отличаются степенью распространенности. Исследование же лексики современных медиатекстов позволяет понять, какие термины и выражения той или иной семантической группы наиболее знакомы массовой аудитории, используются не в узких кругах специалистов, а среди широких масс.

В последнее время совершенствование компьютерных технологий послужило катализатором «корпусной революции», которая, несмотря на различные трудности, не обошла стороной ряд африканских языков, включая и суахили [8]. Одним из крупнейших корпусов на языке суахили является общий корпус текстов, разработанный в Университете Хельсинки под руководством А. Хурсканена, содержащий около 25 миллионов слов. За его основу были взяты различные литературные произведения на суахили, а также материалы прессы. Исследователи из Кении и ЮАР разработали параллельный корпус суахили и английского языков для машинного перевода. Корпус SAVA включает в себя разнообразные тексты: из Библии, Корана, субтитры к фильмам, инвестиционные отчеты, и другие. Общее количество токенов (словоформ) на английском языке 1463 миллиона, на суахили — 1201 миллион [9].

Танзанийские исследователи из Университета в Дар-эс-Саламе разработали тематический корпус языка суахили по медицине (С. Севанги и другие) содержащий около 2 миллионов слов. Для его создания были использованы тексты из книг, специализированных журналов, брошюр. Данный корпус, наряду с общим корпусом, состоящим из 500 тысяч слов, составленным на основе новостных текстов различных интернет-изданий, использовался в работе С. Севанги и П. Даниэльссон для выявления определений медицинских терминов [10].

Разнообразие корпусов обусловлено разнородностью задач, стоящих перед лингвистами. Для целей нашей работы — выявления актуальных медицинских терминов и выражений — существующие корпуса неприменимы, так как они содержат в том числе и устаревшие медиатексты, не категоризированные по определенным тематическим группам, то есть не содержат исключительно современные медиатексты медицинской тематики. Особое внимание в нашей работе уделяется выявлению современных терминов и выражений, которые встречаются в текстах, рассчитанных на массовую аудиторию, то есть актуальны и понятны широкой аудитории. Именно поэтому для составления нашего корпуса были использованы материалы исключительно современных медиатекстов на языке суахили.

При составлении корпуса мы руководствовались рядом общепризнанных принципов, таких как объем, сбалансированность, репрезентативность, релевантность для целей исследования, систематизация с точки зрения структуры и содержания. Материалами для данного исследования послужили медиатексты интернет-издания Mtanzania *Танзаниец*, опубликованные в 2014–2019 гг. Выбор данного издания обусловлен его ориентированностью на читателей среднего класса, то есть изложение должно быть доступно и понятно массовому читателю, а значит используемый язык достаточно стандартный, не специализированный. В рамках издания медиатексты логично систематизированы: сгруппированы по тематическим разделам, что позволяет легко вычлнить материалы, необходимые для составления корпуса. Для целей нашей работы были использованы тексты раздела *Afuа* на Jamii *Здравоохранение и социальная сфера*. Немаловажным представляется тот факт, что разделы интернет-издания Mtanzania обновляются на регулярной основе, что предоставляет нам возможность получить достаточно материала для анализа.

На основании данных медиатекстов в электронном формате нами был сформирован тематический корпус. Общее количество медиатекстов, использованных для составления корпуса, составляет 584 единицы, общее количество слов (словоформ, токенов) — 285 477 (из них уникальных типов токенов — 29 981). Для сбора данных нами были написаны скрипты, позволившие собрать ссылки на новостные статьи раздела *Afuа* на Jamii, а затем скачать тексты этих статей, сформировав таким образом общий тематический файл. Далее каждый файл был преобразован так, чтобы на одной строке содержалось одно предложение. В результате мы получили корпус в формате, доступном для дальнейшего анализа. Из полученного корпуса были сформированы следующие файлы: списки частотности употребления слов; список частотности биграмм, триграмм. В сформированных файлах вручную анализируются конкордансы на предмет коллокаций — словосочетаний, представляющих собой синтаксически и семантически целостную единицу. Под конкордансами понимается набор контекстов, в которых встречается слово или словосочетание. Для верификации результатов была использована программа AntConc, которая подтвердила полученные результаты. Однако небольшие расхождения показали, что программа AntConc менее точна при анализе корпуса текстов, так, например, в программе не учитывается различие между знаками дефис и тире.

Последовательная проверка и анализ списка частотности употребления слов позволила выявить термины, слова и выражения медицинской и социальной тематики. В Таблице 1 представлены 20 наиболее часто встречающихся слов данной тематики.

Таблица 1.

Частота	Слово	Перевод
933	afua	здоровье

634	dawa	лекарство /лекарства
543	damu	кровь
492	wagonjwa	больные
491	saratani	рак
490	ugonjwa	болезнь
464	magonjwa	болезни
428	mwili	тело
425	hospitali	больница
416	matibabu	лечение
410	upasuaji	операция
380	moyo	сердце
318	uzazi	роды, рождение
279	tiba	лечение, лекарство, медицин
236	mgonjwa	больной
235	daktari	доктор, врач
201	maambukizi	инфекция
189	mimba	беременность, зачатие, плод
183	madaktari	доктора, врачи
153	maumivu	боль

Данные по частоте встречающихся в корпусе биграмм, триграмм приведены в Таблицах 2 и 3 соответственно.

Таблица 2

Частота	Биграмма		Перевод
	1	2	
71	daktari	bingwa	врач-специалист
43	kutumia	dawa	принимать лекарства
42	kupata	saratani	заболеть раком
35	mganga	mkuu	главный врач
34	kufanyiwa	upasuaji	быть прооперированным
30	tezi	dume	мужская железа /простата
27	maumivu	makali	острая боль

26	kupata	magonjwa	заболеть
26	kufanya	urasuaji	делать операцию
24	kupoteza	maisha	потерять жизнь
23	kupata	maambukizi	получить инфекцию, заразиться
22	kupatiwa	matibabu	получать лечение
22	kupata	matibabu	получать лечение
21	kupata	ugonjwa	заболеть
20	uzito	mkubwa	большой вес
20	kupata	mtoto	родить ребенка
20	kupata	mimba	забеременеть
20	kudhibiti	UKIMWI	защищать от СПИДа
19	kupata	ujauzito	забеременеть
18	kupima	afya	проверять здоровье

Как правило, биграмы состоят из:

а) транзитивного глагола, управляющего прямым дополнением: kuimarisha afya *укреплять здоровье* (частотность 18), kuchangia damu *сдавать кровь* (частотность 16);

б) пассивной формы глагола, указывающей, что агент подвергается некой процедуре, номинируемой дополнением: kupatiwa matibabu *букв.быть тем, кому предоставляют лечение > получать лечение* (частотность 22), kufanyiwa uchunguzi *быть обследованным* (частотность 11);

в) именной группы, образованной путем примыкания: tezi dume *мужская железа/простата* (частотность 30), uric acid *мочевая кислота* (частотность 14);

г) именной группы, состоящей из существительного и согласованного с ним прилагательного: taumivu makali *острая боль* (частотность 27), waganga wakuu *главные врачи* (частотность 9).

Таблица 3

Частота	Триграммы			Перевод
18	maambukizi	ya	VVU	заражение ВИЧ
15	mapigo	ya	mooyo	удары сердца
15	magonjwa	ya	saratani	онкологические заболевания
15	huduma	ya	afya	здравоохранение
15	dhidi	ya	ugonjwa	против болезни
14	wakati	wa	ujauzito	во время беременности
13	ugonjwa	wa	malaria	заболевание малярией
13	tiba	ya	mifupa	ортопедическое лечение
13	taasisi	ya	tiba	институт медицины,

				медицинское учреждение
13	maambukizi	ya	ugonjwa	заражение заболеванием
12	ugonjwa	wa	fizi	пародонтоз /заболевание десен
12	mbegu	za	kiume	сперма
12	elimu	ya	afya	медицинское образование
11	maumivu	ya	mgongo	боль в спине
11	maduka	ya	dawa	аптеки
11	kiwango	cha	maambukizi	уровень заражения
10	uzito	wa	mwili	вес тела
10	ugonjwa	wa	UKIMWI	болезнь СПИД
10	ugonjwa	wa	saratani	болезнь рак
10	ugonjwa	wa	ebola	болезнь эбола

Триграммы представляют собой именную группу «существительное + пассивный формант, согласованный по классу вершинного существительного + второе существительное».

Последовательная проверка полученных результатов показала, что в анализируемом корпусе медиатекстов присутствует значительное число единичных и составных терминов, при этом отсутствуют определения в явной форме. Таким образом, можно сделать вывод о том, что авторы медиатекстов рассчитывают на владение реципиентами терминологией в достаточной степени для адекватного восприятия передаваемой информации, и дополнительные разъяснения не требуются. При этом значение того или иного термина может быть получено посредством анализа контекста и непосредственного окружения (коллокаций). Так, например, даже в тех случаях, когда реципиент не владеет терминами, обозначающими названия болезней, он может идентифицировать лексическую семантику данного термина благодаря наличию слов *ugonjwa / magonjwa* *болезнь, заболевание / заболевания* : *magonjwa ya saratani онкологические заболевания* (частотность 15), *ugonjwa wa UKIMWI* *болезнь СПИД* (частотность 10).

Наши результаты значительно отличаются от данных, полученных С. Севанги и П. Даниэльссон при анализе термина UKIMWI (СПИД). Они выявили определения термина UKIMWI (СПИД) в явной форме, что объясняется использованием для составления корпуса специализированной, в том числе и обучающей литературы:

UKIMWI ni hali anayoipata mtu wakati ambapo mwili wake hauwezi kujitengenezea kinga ya magonjwa. *СПИД — это состояние, в которое приходит человек, когда его организм не вырабатывает иммунитет от заболеваний.*

UKIMWI ni ugonjwa anaoupata mtu baada ya kinga yake ya mwili kuharibiwa.

СПИД – это болезнь, которую приобретает человек после того, как разрушается иммунная система организма.

Таким образом, применение методики вычленения актуальных единичных и составных медицинских терминов путем анализа корпуса медиатекстов раздела Afya на Jamii *Здравоохранение и социальная сфера* при помощи системы UNIX показало свою эффективность. Полученные результаты являются примером предварительных данных, которые нуждаются в дальнейшей обработке лексикографами и специалистами в области медицины. Неоспоримым преимуществом анализа корпуса медиатекстов является использование в них «живого» языка с лексическим наполнением, актуальным для определенного периода времени (в нашем случае период 2014–2019 гг.).

Полученные результаты могут найти в дальнейшем практическое применение: после лексикографической обработки выявленные термины и выражения могут использоваться для составления специализированных тематических словарей на языке суахили.

Библиография

1. Legere K. Formal and informal development of the Swahili language: Focus on Tanzania. // Selected Proceedings of the 36th Annual Conference on African Linguistics, 2006, pp. 176–184.
2. Mwansoko H J M Istilahi za ufundi wa magari na matrekta, Dar es Salaam: Institute of Kiswahili Research in University of Dar es Salaam, 1997.
3. Tumbo–Masabo Z.N., Chuwa A.R. Kamusi ya biashara na uchumi: Kiingereza–Kiswahili, Dar es Salaam: Institute of Kiswahili Research in University of Dar es Salaam, 1997.
4. Mlacha S.A.K. Kamusi ya sheria : Kiingereza–Kiswahili, Dar es Salaam: Institute of Kiswahili Research in University of Dar es Salaam, 1999.

5. Mwita A. M. A., Mwansoko H. J. M. Kamusi ya Tiba, Dar es Salaam: Institute of Kiswahili Research in University of Dar es Salaam, 2003.
6. Mwansoko H.J.M., Tumbo–Masabo Z.N. Sewangi. S.S. Kamusi ya Historia. Dar es Salaam: Institute of Kiswahili Research in University of Dar es Salaam, 2004.
7. Sewangi S.S. Tapping the neglected resource in kiswahili terminology: automatic compilation of the domain-specific terms from corpus // Nordic Journal of African Studies, Vol.9, No.2, 2000, pp. 60–84
8. Выдрин В.Ф. Электронные корпуса африканских языков: завтра или послезавтра? // Петербургская африканистика. Памяти А.А. Жукова. СПб.: Издательство Санкт–Петербургского университета, 2008. стр. 279, с. 29–39
9. Aaron M.O. et al. Swahili text and speech corpus: a review // Asian Journal of Computer Science And Information Technology, Vol. 2. No. 11, 2012, pp. 286–290.
10. Sewangi S.S., Danielsson P. Deriving contextual defining information for technical terms from specialized corpus — the case of Kiswahili health care terminology // Kiswahili. Journal of the Institute of Kiswahili Research, Vol. 72, 2009, pp. 37–51.

References (transliterated)

1. Legere K. Formal and informal development of the Swahili language: Focus on Tanzania. // Selected Proceedings of the 36th Annual Conference on African Linguistics, 2006, pp. 176–184.
2. Mwansoko H J M Istitlahi za ufundi wa magari na matrekta, Dar es Salaam: Institute of Kiswahili Research in University of Dar es Salaam, 1997.
3. Tumbo–Masabo Z.N., Chuwa A.R. Kamusi ya biashara na uchumi: Kiingereza–Kiswahili, Dar es Salaam: Institute of Kiswahili Research in University of Dar es Salaam, 1997.
4. Mlacha S.A.K. Kamusi ya sheria : Kiingereza–Kiswahili, Dar es Salaam: Institute of Kiswahili Research in University of Dar es Salaam, 1999.
5. Mwita A. M. A., Mwansoko H. J. M. Kamusi ya Tiba, Dar es Salaam: Institute of Kiswahili Research in University of Dar es Salaam, 2003.
6. Mwansoko H.J.M., Tumbo–Masabo Z.N. Sewangi. S.S. Kamusi ya Historia. Dar es Salaam: Institute of Kiswahili Research in University of Dar es Salaam, 2004.
7. Sewangi S.S. Tapping the neglected resource in kiswahili terminology: automatic compilation of the domain-specific terms from corpus // Nordic Journal of African Studies, Vol.9, No.2, 2000, pp. 60–84
8. Vydrin V.F. Elektronnye korpusa afrikanskikh yazykov: zavtra ili poslezavtra? // Peterburgskaya afrikanistika. Pamyati A.A. Zhukova. SPb.: Izdatel'stvo Sankt–Peterburgskogo universiteta, 2008. str. 279, s. 29–39
9. Aaron M.O. et al. Swahili text and speech corpus: a review // Asian Journal of Computer Science And Information Technology, Vol. 2. No. 11, 2012, pp. 286–290.
10. Sewangi S.S., Danielsson P. Deriving contextual defining information for technical terms from specialized corpus — the case of Kiswahili health care terminology // Kiswahili. Journal of the Institute of Kiswahili Research, Vol. 72, 2009, pp. 37–51.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Современный информационный мир тяготеет к тому, чтобы ряд технологий были усовершенствованы настолько, насколько это может на данный момент позволить технический прогресс. Медийная сфера, текстовый контент, рецепция коннотативных импульсов также приобретают новые формы, трансформируются для более удобного и быстрого разрешения. Так или иначе, массовая аудитория требует дискретного потока информации. Рецензируемая статья, на мой взгляд, затрагивает достаточно серьезный вопрос оценки как медиаиндустрии, так и способов ее совершенствования. Актуальность проблематики не вызывает сомнений, ибо последние два-три-четыре года научный мир сориентирован на смежный вектор вопросов и это не только гуманитарная (лингвистическая) стезя, но и социальная, философская, техническая, математическая. Таким образом, предмет исследования – медийный язык, регулирование и анализ сферы речепользования, оценка современных медиатекстов на языке суахили – выбран правильно, а расшифровка, концептуальное описание указанных категорий есть верификация мейнстрим граней развития современного общества. Методы корпусной лингвистики для проведения исследования вполне подходят, их актуальность не вызывает нареканий. Автор в начале текста комментирует данный выбор, делается это профессионально и грамотно. Выбор медицинской тематики обосновывается также, в большей степени это связано с тем, что «использование медицинских терминов в медиатекстах, рассчитанных на массовую аудиторию, предполагает владение реципиентами данной терминологией, и, соответственно, она относится к «живому» актуальному лексическому пласту». Работа весьма грамотно построена, в ней нет каких-либо лишних частей, избыточных фактов, избыточной информации. Язык суахили не имеет столь широкого распространения, следовательно, комментирование лексического корпуса и пропаганда данной формы есть попытка объемно и целостно изучить пути его современного состояния. Например, тезис, что «исследование лексики современных медиатекстов позволяет понять, какие термины и выражения той или иной семантической группы наиболее знакомы массовой аудитории, используются не в узких кругах специалистов, а среди широких масс» подтверждает это. Автор включает в текст работы довольно большое количество фактического материала, статистических данных, верифицированной информации. На мой взгляд, это делает работу объективной, строгой, научно-правильной. Следование внутренней логики поддерживает нарочитыми «повторами», введением «промежуточных итогов», усложнением индивидуальных «позиций», «точки зрения». Безусловно, впечатляет объем изученного/исследованного материала: «на основании данных медиатекстов в электронном формате нами был сформирован тематический корпус. Общее количество медиатекстов, использованных для составления корпуса,

составляет 584 единицы, общее количество слов (словоформ, токенов) — 285 477 (из них уникальных типов токенов — 29 981)». Примечательно для работы факт включения табличных «оттисков», они подтверждают системность обработки информации, при этом потенциальный читатель сможет цельно представить себе корпус обработанного контента. Научная новизна исследования заключается в перегруппировке медицинских понятий и терминов, а также в определении их актуальности на данный момент. Это можно было бы прописать в работе более конкретно и точно, не только в финальной части. Текст однороден, стилистически выверен, терминологически универсален. Содержательная часть статьи информативно, доказательна, пересмотра выводов/аргументов не требуется. Небезынтересен и факт, что автор апеллирует к уже «имеющемуся опыту» систематизации понятий, хотя согласия/примирения в данном случае нет: «наши результаты значительно отличаются от данных, полученных С. Севанги и П. Даниэльссон при анализе термина UKIMWI (СПИД). Они выявили определения термина UKIMWI (СПИД) в явной форме, что объясняется использованием для составления корпуса специализированной, в том числе и обучающей литературы...». Завершает работы вывод, свидетельствующий о качественной расшивке заявленной темы. Логично подводится итог и высказывается, что «применение методики вычленения актуальных единичных и составных медицинских терминов путем анализа корпуса медиатекстов раздела Afya na Jamii Здравоохранение и социальная сфера при помощи системы UNIX показало свою эффективность. Полученные результаты являются примером предварительных данных, которые нуждаются в дальнейшей обработке лексикографами и специалистами в области медицины. Неоспоримым преимуществом анализа корпуса медиатекстов является использование в них «живого» языка с лексическим наполнением, актуальным для определенного периода времени (в нашем случае период 2014–2019 гг.)». Как видно и текста, материалы должны стать импульсом для дальнейшего изучения «современного состояния языка суахили в медиапространстве», а также найти практическое применение, в частности при составлении тематических специализированных словарей и справочников. Библиографический список достаточен, формальный ценз при составлении выдержан. Работа имеет завершённый вид, специальной правки текста не требуется. Статья «Анализ современных медиатекстов на языке суахили с использованием методов корпусной лингвистики» может быть рекомендована к открытой публикации в журнале «Litera».

Издательство

О нас

Наши реквизиты

Наши партнеры

Контактная информация

Политика

конфиденциальности



© 1998 – 2019 Nota Bene. Publishing Technologies. NB-Media Ltd.

Авторам

Договоры

Политика авторских прав и лицензий

Авторская зона

Услуги

Институт

рецензирования

Разное

Как написать статью?

Журналы в перечне ВАК

Мобильное приложение

Вопрос - ответ

Наши сайты

Aurora group s.r.o.



Журнал позволяет авторам сохранять без каких-либо ограничений авторские права. Все авторы автоматически приобретают авторские права на свои произведения в момент их создания. Авторские права защищены действующим законодательством Российской Федерации.

Тип лицензии, поддерживаемый журналами издательства: Attribution- NonCommercial 4.0 International (CC BY-NC 4.0) – Лицензия «С указанием авторства – Некоммерческая».

Журнал имеет свободный доступ, это означает, что статьи можно читать, загружать, копировать, распространять, печатать и ссылаться на их полные тексты с указанием авторства в некоммерческих целях в соответствии с условиями Лицензии «С указанием авторства – Некоммерческая» (Creative Commons Attribution-NonCommercial 4.0 International License).

Вы вправе свободно:

Делиться (обмениваться) — копировать и распространять материал на любом носителе и в любом формате.

Адаптировать (создавать производные материалы) — делать ремиксы, видоизменять, и создавать новое, опираясь на этот материал. Лицензиар не вправе аннулировать эти свободы пока вы выполняете условия лицензии.

Указанные выше права реализуются при соблюдении следующих условий:

Атрибуция — вы должны обеспечить соответствующее указание авторства, предоставить ссылку на лицензию и указать, были ли внесены изменения в произведение. Вы можете делать это любым разумным способом, но не таким, который подразумевал бы, что лицензиар одобряет вас или ваш способ использования произведения.

Некоммерческий — вы не можете использовать материал в коммерческих целях.

Никаких дополнительных ограничений — вы не вправе применять юридические ограничения или технологические меры, создающие другим юридические препятствия в реализации тех действий, что разрешены лицензией.