

ТИПЫ ДИСКУРСА И ИХ СТРУКТУРНО-СЕМАНТИЧЕСКАЯ РЕПРЕЗЕНТАЦИЯ

УДК 81-13:32.019.51(470)

АНАЛИЗ НОВОСТНЫХ СООБЩЕНИЙ САЙТА МИД РФ МЕТОДОМ КОНТЕНТ-АНАЛИЗА (статья 1)

М.В. Беляков

Московский государственный институт международных отношений (У)
пр-т Вернадского, 76, Москва, Россия, 119454
belmax007@hotmail.com

Современное мировое информационное пространство характеризуется непрекращающимися информационными и гибридными войнами. Извлечение данных из информационного потока превратилось из желаемой в необходимую задачу. В статье рассматриваются формальные методы анализа текстовой информации и проводится анализ новостных сообщений сайта МИД России методом автоматизированного контент-анализа программой QDA Miner. Официальный сайт практически невозможно ориентировать на определенную категорию адресата по возрастному, гендерному, статусному и прочим признакам. Контент-анализ позволяет определить, как содержание сообщения влияет на аудиторию, и количественно описать содержание коммуникации, а также снижает вероятность субъективности при интерпретации текста. Проведенный анализ показал частоту распределения данных по выделенным рубрикам.

Ключевые слова: дипломатический дискурс, сайт, контент-анализ, МИД РФ, поликодовый

ВВЕДЕНИЕ

Современные технологии постоянно создают новые возможности для исследования различных областей науки и разных сфер деятельности человека. Это касается и анализа текстов. Помимо собственно текстологического анализа, дискурс-анализа и других общепринятых методов анализа текста широко распространены и формальные методы анализа текста. Эти методы позволяют решать самые разные задачи, которые ставит перед собой исследователь. Разнообразие методов анализа текста позволяет изучать любые текстовые массивы, содержащие в себе информацию разных тематик, и любые типы текста, включая тексты поликодовые. Это могут быть произведения классической литературы, религиозные тексты, рекламные сообщения, новостные статьи. Новостные статьи вызывают особый

интерес, поскольку международные отношения и внешнеполитическая позиция государства играют в эпоху информационных и гибридных войн особую роль. Точка зрения государства на складывающуюся в мире непростую ситуацию и отражение усилий, предпринимаемых дипломатическими ведомствами на внешней арене, представлены в информационных материалах, опубликованных на официальном сайте Министерства иностранных дел Российской Федерации.

АВТОМАТИЗИРОВАННЫЙ АНАЛИЗ ТЕКСТА

На сегодняшний день существует немало методов, позволяющих осуществлять автоматизированный анализ текста. Систематизация наиболее популярных методов дает следующую классификацию:

◆ По выполняемым функциям:

1. Методы, предполагающие импорт текста и дальнейшую работу с ним.
2. Методы исследования текста, где работа ведется на синтаксическом и грамматическом уровнях, например, поиск слов в тексте.
3. Методы, ориентированные на семантический анализ, где в процессе анализа создаются списки категорий, словари и т.д.
4. Методы, предполагающие экспорт полученных в результате исследования данных.

◆ По объекту анализа:

1. Методы, ориентированные на язык:
 - лингвистические методы;
 - методы работы с данными.
2. Методы, ориентированные на «контент»:
 - качественные методы;
 - количественные методы (категориальные и некатегориальные системы).

Среди современных методов анализа текста и дискурса выделяются:

1. Интент-анализ, который делает возможной реконструкцию интенций, то есть направленность автора на тот или иной объект. Для интент-анализа могут применяться такие прикладные пакеты, как Ethnograph, Leximancer, Minnesota-ContextualContentAnalysis.

2. Контент-анализ дает возможность провести качественный или количественный анализ содержания текстового массива. В результате необходимо получить интерпретацию выявленных закономерностей. Для контент-анализа могут применяться следующие прикладные пакеты: CrawdadDesktop, INTEXT, QDAMiner, Yoshicoder.

3. Фоносемантический анализ позволяет анализировать текст только на фонетическом и семантическом уровнях, безотносительно содержания текста. Возможно использование таких программ, как DIATON и Vaal.

4. Дискурс-анализ представляет собой совокупность техник и методик интерпретации текстов и высказываний. Для проведения дискурс-анализа используется, в частности, программа CATRAS.

5. Нарративный анализ соотносит последовательность слов в предложениях, составленных на основе некоторого языкового опыта, с реально существующими в текстах. При нарративном анализе применяют такие программы, как LIWC и PC-ACE.

6. Экспертная атрибутивная оценка текста представляет собой анализ текста с целью выявления некоторой информации о его авторе (где, когда и кем был написан текст, как автор относится к написанному и т.д.). Такими программами являются ЛингвоАнализатор, Атрибутор и др.

7. Морфологический автоматический анализ предполагает исследование всех возможных морфологических интерпретаций каждого слова в тексте. Возможно использование следующих методик: ATLAS. ti, TextArc, Textanz.

8. Семантический автоматический анализ позволяет выявить семантическую структуру предложений и значения реализованной в тексте полисемии. Это можно осуществить при помощи таких программ, как TextAnalyst 2.0, и разнообразных конкордансеров.

9. Синтаксический автоматический анализ сопоставляет линейные последовательности лексем языка с его грамматикой. Применяют методики ProfilerPlus, DictaScore и другие [5. С. 29—38].

УРОВНИ КОНТЕНТ-АНАЛИЗА

Количественный контент-анализ, представляющий особый интерес для данного исследования, — это квантитативный анализ текстов и текстовых массивов с целью последующей содержательной интерпретации выявленных числовых закономерностей. В более узком смысле контент-анализ — методика, позволяющая выявлять частоты появления в тексте определенных характеристик, которые интересуют исследователя и которые позволяют ему делать выводы относительно намерений создателя данного текста или вероятных реакций адресата [4. С. 25]. На сегодняшний день это один из наиболее распространенных количественных методов анализа текста в лингвистике. Изначально под контент-анализом понимался подсчет элементов текста, их дальнейшая систематизация и категоризация с целью последующего выявления основных тенденций (тем, настроений) в тексте. Основная идея контент-анализа состояла в том, чтобы проверить, насколько мнение человека, анализирующего текст на основе своих собственных ощущений относительно прочитанного, соответствует данным об этом тексте, подкрепленным статистически. Пик развития контент-анализа пришелся на первую половину XX в. Связано это было с развитием средств массовой информации. Многие читатели газет замечали, что некоторые темы поднимаются редакторами особенно часто, в то время как другие темы остаются «в тени». Каждое издание, в том числе относящееся и к интернет-ресурсам, определяет спектр этих тем для себя сам. С помощью контент-анализа можно также определить не только кто и что сказал, но еще и с каким эффектом. Эта модель Гарольда Д. Лассуэла [3] легла в основу контент-анализа. Сам он подчеркивал ценность этого вида анализа текстов для политического дискурса.

Контент-анализ можно проводить на разных уровнях текста. Выделяется шесть таких уровней.

1. Синтаксический уровень, где анализируются буквы, слова, слоги или предложения.

2. Семантико-синтаксический уровень, где изучают, как синтаксические структуры могут повлиять на значение.

3. Семантический уровень, где происходит анализ значений слов, словосочетаний и предложений.

4. Прагматико-синтаксический уровень, где анализируются связи синтаксиса и эффекта текста.

5. Семантико-прагматический уровень, где соотносят эффект текста с конкретными его единицами.

6. Прагматический уровень, где ищут структуры, влияющие на восприятие текста, например риторические вопросы [6. С. 93].

Различают традиционный и интерпретативный контент-анализ. Первым их разделил Аарон Ахувиа [1. С. 139—172]. При традиционном контент-анализе текст кодируют несколько кодировщиков по определенным правилам, и затем его интерпретируют. При интерпретативном контент-анализе всю работу выполняет один кодировщик. В этом виде анализа не создается никаких правил кодирования. Предполагается, что кодировщик достаточно компетентен в тематике анализируемых данных и разбирается в тонкостях их значений. В традиционном контент-анализе встает проблема бесконечности контекстов.

Теоретически возможно создать систему правил, по которым одни и те же единицы текста в разных контекстах кодируются по-разному. С другой стороны, количество контекстов может быть не ограничено, и на практике создать подобные правила становится невозможным, поэтому и предлагается интерпретативный контент-анализ. Этот вид контент-анализа имеет свои недостатки. При нем сложно определить степень объективности анализа. Если при традиционном контент-анализе одни и те же тексты кодируются разными кодировщиками, и правильным признается мнение большинства, то при интерпретативном кодировании степень объективности проведенной работы определяет сам автор. Основное отличие интерпретативного контент-анализа от традиционного состоит в способе кодирования и в том, как оценивается качество кодирования. Таким образом, интерпретативный контент-анализ лучше учитывает контекст, так как он не ограничен никакими правилами кодирования.

Сам метод и цели метода менялись на протяжении времени. Тем не менее, некоторые из них остаются актуальными по сей день. Во-первых, контент-анализ проводят, чтобы определить, как содержание сообщения влияет на аудиторию. Во-вторых, его используют для того, чтобы количественно описать содержание коммуникации. В то же время следует заметить, что контент-анализ, и количественный, и качественный, снижает вероятность субъективности при интерпретации текста.

ПРОЦЕДУРА КОНТЕНТ-АНАЛИЗА

Процедура проведения контент-анализа включает в себя несколько стадий: определяется цель исследования, его задачи и тема исследования. На этой основе формируется выборка. Следует понимать, что не всегда удастся собрать и проанализировать весь материал по той или иной проблеме. Поэтому исследователи прибегают к частичной выборке или выборке, формируемой по вероятностному принципу. Процесс отбора выполняется в несколько этапов — выбирается отправитель; происходит отбор непосредственно документов для анализа; происходит отбор подмножества документов. [2. С. 601]. Процесс категоризации и кодирования занимает ключевое место при проведении контент-анализа. Категоризация — это процесс и результат распределения некоторого числа элементов по категориям. В свою очередь, категория — это множество слов, объединенных по некоторому признаку.

При проведении контент-анализа необходимо корректно составить систему категорий, для чего применяются программы кодирования и декодирования текста, такие как QDAMiner, WordStat, NVivo и другие. Каждую категорию рекомендуется иллюстрировать примерами, чтобы в дальнейшем кодировщику было проще определить, подходит ли та или иная единица под данную категорию. В случае если во время кодирования обнаруживаются новые категории, следует перекодировать весь материал, опираясь уже на новый список категорий.

Существуют перечни типов категорий, на которые исследователь может опираться при выделении категорий. Один из таких перечней был составлен Оле Холсти [2]. Он предлагал выделять тему, которой посвящен текст; ценности, т.е. установки и желания, выраженные в тексте; средства, используемые для достижения цели, и т.п. Выделенные категории должны покрывать полностью весь текст. Интерпретация полученных результатов должна соответствовать тем задачам, которые исследователь ставил изначально.

Интерес к исследованию сложного взаимодействия элементов в рамках мультисемиотических структур, таких как поликодовые тексты в самом широком понимании этого слова, в последнее время подпитывается интенсивным развитием интернет-культуры. За последние тридцать лет активного развития интернет-среды сложилась не существовавшая до этого времени система опосредованной коммуникации, предлагающая собственные правила в рамках существующей мультикультурной среды. В этой среде присутствует своя система ценностей, знаковых систем, символов и смыслов, эстетических норм, уже сложившихся традиций. Одной из множества составляющих интернет-среды, где реализуются эти понятия, является веб-сайт, т.е. место в Интернете, имеющее свой адрес (URL) и владельца, состоящее из веб-страниц, воспринимающихся как одно целое. В настоящее время наличие собственного веб-сайта считается необходимой составляющей существования любой организации, в том числе и государственной, такой как, например, МИД РФ.

Теперь уже трудно представить себе сайт, оформленный в виде *plain text*, т.е. представленный только в виде гомогенного вербального компонента. Все современные сайты, включая и официальные, стремятся к негомогенности, поликодовости, мультисемиотичности, мультимодальности.

ПОЛИКОДОВОСТЬ ТЕКСТА

Поликодовый текст — это сочетание элементов естественного языка с кодами других семиотических систем в рамках одного текста. Фактура таких текстов состоит из двух негомогенных частей: вербальной — языковой/речевой и невербальной, принадлежащей к иным знаковым системам, отличным от естественного языка.

Исследователи выделяют три разновидности нелинейного текста: монокодовый, дикодовый и поликодовый текст:

<i>Монокодовый текст</i>	<i>Поликодовый текст</i>
Каналы воздействия на адресата	
Воздействует на один канал восприятия (аудиальный — если текст устный, или визуальный — если текст письменный)	Воздействует на несколько каналов восприятия информации (аудиальный + визуальный)
Способы кодирования информации	
Использует только один семиотический код — вербальный, иконический или звуковой	Использует несколько семиотических кодов одновременно: вербальный, иконический, звуковой

Поликодовые тексты Интернета могут быть как с нулевой, так и ненулевой гетерогенностью: статичными (информационный текстовый блок) или динамичными (баннерная реклама); двумерными; устными (с озвученной анимацией или элементами звукового видео) и письменными; включающими графические знаки нескольких естественных языков, а также использующими разнообразные варианты шрифтов, цвета и графики.

Считается, что на уровне глубинной семантики не существует принципиальной разницы между значением вербальных и невербальных знаков, тем не менее специальные исследования показывают, что вербально и невербально передаваемая информация воспринимается по-разному. При этом вербально представленная информация влияет на сознание рациональным путем, а использование иных, паралингвистических средств автоматически переводит восприятие на подсознательный уровень. Изображение, в отличие от слова, вербализующего позицию автора, принимается в качестве объективной картинки и не соотносится в сознании адресата с установкой адресанта, а кажется более демократичным.

Такие конструктивные признаки, как цельность и связность в поликодовом тексте, являются результатом взаимодействия вербального, иконического и в ряде случаев аудио- и визуального компонентов. Изображение и слово при этом не являются простой суммой знаков, их значения интегрируются и образуют сложное смысловое единство. При этом уровень интеграции может быть различным: полным или частичным. В текстах с полной креолизацией, где под креолизацией мы понимаем процесс дополнения вербальных интернет-текстов рисунками, звуком, анимацией, гипертекстовыми ссылками и пр., наблюдается логическое соединение всех компонентов. Если подобного не происходит, наступает коммуникативная неудача.

ОФИЦИАЛЬНЫЙ САЙТ КАК ПОЛИКОДОВЫЙ ТЕКСТ

Примером поликодового текста в электронных СМИ можно считать официальный сайт любой организации, включая организации столь высокого уровня ответственности, как МИД России. Выбранный для анализа сайт является официальным сайтом Министерства иностранных дел Российской Федерации. Сайт содержит в себе информацию о работе министерства и его текущей деятельности, отчеты о прошедших дипломатических встречах, переговорах, брифингах. На нем также публикуются комментарии представителей министерства по актуальным проблемам, обзоры обращений граждан и др. По структуре этот сайт является поликодовым.

Официальные (государственные) сайты обладают рядом особенностей, отличающих их от коммерческих сайтов, сайтов-визиток и пр. на уровне целей сайта, подачи информации, восприятия информации и пр. Официальный сайт практически невозможно ориентировать на определенную категорию адресата по возрастному, гендерному, статусному и прочим признакам. Такие сайты рассчитаны, как правило, на широкий круг адресатов. В то же время адресат, как правило, представляет собой конкретную личность, т.е. индивидуален.

При этом адресант — некая обобщенно-личная государственная структура, задачей которой является формирование образа легитимного, объективного, взвешенного проведения государственной политики, в рассматриваемом нами примере — внешней. Для официальных российских сайтов, таких как kremlin.ru, mvd.ru, nalog.ru, mid.ru, характерен способ умеренной креолизации сайта.

Как упоминалось выше, анализ текстов сайта МИД РФ проводился методом количественного контент-анализа. Материалом для исследования послужили новостные сообщения сайта. В выборку вошли все статьи, опубликованные на сайте в разделе «Новости» с 1 февраля по 28 февраля 2015 г., т.е. за один календарный месяц. Всего анализу подверглось 190 статей. Для проведения контент-анализа использовалась программа QDA Miner — компьютерная программа, используемая как инструмент проведения качественных исследований. Программа позволяет кодировать и анализировать текстовые документы, а также визуальную информацию. QDA Miner используют для анализа интервью, речей политиков, официальных документов, газет и журналов, текстов книг, картин, фотографий и так далее.

В пакет также входят дополнительные приложения:

— WordStat позволяет производить контент-анализ и глубинный анализ текста. С помощью него можно производить анализ слов и фраз, содержащихся в различных документах или сегментах текста. WordStat проводит описательный анализ и устанавливает взаимосвязи между словами;

— Simstat — это модуль, который позволяет анализировать количественные данные, проводить буквенно-цифровые вычисления, перекодирования переменных.

При работе с программой можно использовать следующие инструменты:

— “Text Retrieval” проводит поиск текстовых единиц в документе;

— “Section Retrieval” позволяет найти сегменты документа, ограниченного определенными характеристиками. Функция используется для автоматического приписывания кодов;

— “Keyword Retrieval” позволяет осуществлять информационный поиск по ключевым словам;

— “Coding Frequencies” позволяет составить список кодов и описать их статистически “Coding Retrieval” составляет список из всех текстовых сегментов и связывает их с заданными кодами;

— “Codes Co-occurrences” определяет, насколько близки те или иные коды в тексте;

“Codes Sequences” выявляет повторяющиеся последовательности кодов.

Работа велась с файлами формата txt, которые импортировались в программу для дальнейшего анализа. За единицу анализа было принято одно сообщение.

В ходе анализа было выявлено, что по формату подачи информации сообщения разделяются на 10 типов:

- о прошедших переговорах,
- интервью,
- комментарий,
- поздравление,
- ответ,
- вступительное слово,
- заявление,
- сообщение для СМИ,
- брифинг,
- выступление.

Наиболее популярным форматом в исследуемом массиве оказались сообщения о прошедших переговорах. Сообщения этого типа встретились в статьях 85 раз и составили 44,7% от всего массива. Вторым по популярности стал формат комментария. Сообщения этого типа встретились в статьях 34 раза и по встречаемости составили 17,9% от всего массива. Сообщения типа выступления встретились в массиве 13 раз и составили 6,8% по встречаемости в тексте. Сообщения типа «Сообщение для СМИ» встретились в массиве 9 раз и составили 4,7% по встречаемости в тексте. Сообщения типа «заявление» встретились в массиве 8 раз и составили 4,2% по встречаемости в тексте. Сообщения типа «интервью» встретились в массиве 7 раз и составили 3,7% по встречаемости в тексте. Сообщения типа «ответ» встретились в массиве также 7 раз и составили 3,7% по встречаемости в тексте. Сообщения типа «брифинг» встретились в массиве 4 раза и составили 2,1% по встречаемости в тексте. Сообщения типа «вступительное слово» встретились в массиве 3 раза и составили 1,6% по встречаемости в тексте. Сообщения типа «поздравление» встретились в массиве 2 раза и составили 1% по встречаемости в тексте (рис. 1).

При обработке отобранного материала сообщения были сгруппированы по неделе, чтобы проследить, как менялась тема текстов каждую неделю в течение месяца.

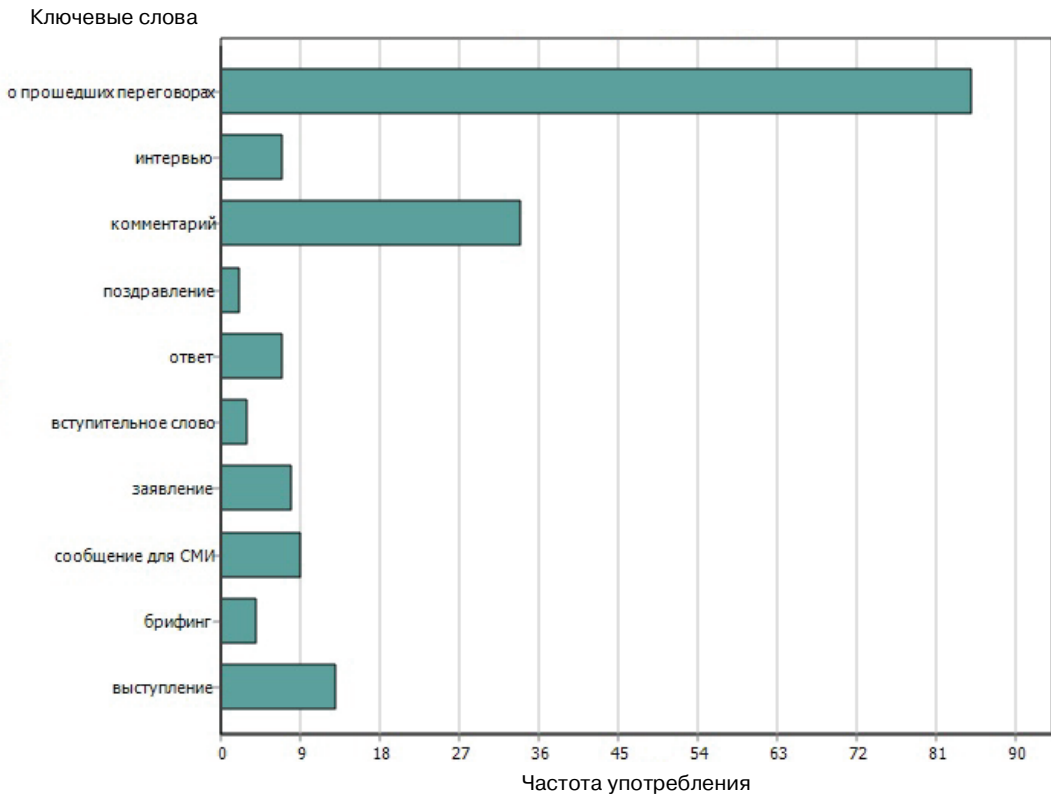


Рис. 1. Распределение ключевых слов по частоте употребления в разных типах сообщений

ЗАКЛЮЧЕНИЕ

В современных условиях официальный сайт, в нашем случае это сайт внешнеполитического ведомства, нельзя оценивать лишь как информационное отражение деятельности государства, его следует воспринимать как мощный инструмент формирования мнения не только отдельных личностей, но и всего общества в целом, что отражается в частоте вербальных показателей, характеризующих затрагиваемые темы. Этому также активно способствует поликодовая структура официальных сайтов. Получить резюмирующие результаты, с минимальной субъективностью показывающие реальное отражение существующей в мире внешнеполитической ситуации и реакции государства на нее, позволяют компьютерные системы извлечения данных из текста и контент-анализа, такие как Crawdad Desktop, INTEXT, QDA Miner, Yoshicoder.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- [1] Ahuvia A. Traditional, Interpretive, and Reception Based Content Analyses: Improving the Ability of Content Analysis to Address Issues of Pragmatic and Theoretical Concern // *Social Indicators Research* 54(2). P. 139—172. URL: <http://hdl.handle.net/2027.42/43696>.
- [2] Holsti O. (1969). *Content analysis for the social sciences and humanities*. Reading, mass etc.
- [3] Lasswell H. (1967). *The Uses of Content Analysis Data in Studying Social Change* // *Science and Culture*. Vol. 33. No 4.

- [4] *Максименко О.И.* (2013). *Формализованная лингвистика* [*Maksimenko O.I. Formalized linguistics*]. Москва: изд-во МГОУ.
- [5] *Митина О.В., Евдокименко А.С.* (2014). *Методы анализа текста: методологические основания и программная реализация* [*Mitina O.V., Evdokimenko A.S. Methods of text analysis: methodological fundamentals and programme realization*] // Вестник ЮУГУ. Серия Психология. № 40 (216). С. 29—38.
- [6] *Тичер С., Веттер Е., Водак Р., Мейер М.* (2009). *Методы анализа текста и дискурса* [*Titscher S., Vetter E., Wodak R., Meyer M. Methods of text and discourse analysis*] / пер. с англ. / под ред. А.А. Киселевой. Харьков: изд-во Гуманитарный центр.

THE ANALYSIS OF NEWS MESSAGES OF THE SITE OF THE RUSSIAN FEDERATION MINISTRY OF FOREIGN AFFAIRS APPLYING CONTENT-ANALYSIS

(Article 1)

M.V. Belyakov

Moscow state institute of international relations
(MGIMO-University), RF Ministry of Foreign Affairs
Vernadskogo ave., 76, Moscow, Russia, 119454
belmax007@hotmail.com

Modern global information space is characterized by incessant information and hybrid wars. Data mining from an information stream has turned from desirable in a necessary task. The article is devoted to the consideration of formal methods of the text information analysis and the quantitative analysis of information of a website of the Ministry of Foreign Affairs of Russia carried out by QDA Miner computer program. The official website cannot be focused on the certain category of the addressee on age, gender, status and the other attributes. The content-analysis allows determining how the content of messages influences an audience and quantitatively describes the contents of the communications, reduces probability of subjectivity of the text interpretation. The carried out analysis has shown the frequency of distribution of the data of the allocated rubrics.

Key words: diplomatic discourse, site, content-analysis, MID RF, polycode