

Хвостик А.Д.

Тирасполь, Приднестровье, Молдова

Приднестровский государственный университет им. Т.Г. Шевченко,

Воронежский государственный университет

tptcfrench2017@mail.ru

Артемова О.Г.

доктор филологических наук, доцент

Воронеж, Россия

Воронежский государственный технический университет,

Воронежский государственный университет

МАРКЕМОЛОГИЧЕСКИЙ ПОДХОД К ФОРМАЛИЗОВАННОМУ СОДЕРЖАТЕЛЬНОМУ АНАЛИЗУ ХУДОЖЕСТВЕННОГО ТЕКСТА

Аннотация: *Цель статьи – описать формализованный содержательный анализ текста в русле маркемологического подхода, предлагающего количественные методы выборки потенциально маркированных лексем. Рассмотрены основополагающие понятия маркемологии, продемонстрированы на примере художественного текста этапы создания маркемного списка произведения, выявлены его доминанта и вице-доминанта.*

Ключевые слова: *маркемологический подход, формализованный содержательный анализ, маркема, маркемология, система фильтров*

Khvostik A.D.

Tiraspol, Transnistria, Moldova

T.G. Shevchenko Pridnestrovian State University,

Voronezh State University

tptcfrench2017@mail.ru

Artemova O.G.,

Doctor habil. in Philology, Associate Professor

Voronezh, Russia

Voronezh State Technical University,

Voronezh State University

THE MARKEMOLOGICAL APPROACH TO FORMALIZED CONTENT ANALYSIS OF LITERARY TEXTS

Abstract: *The purpose of this article is to describe a formalized content analysis of texts within the framework of the markemological approach, which employs quantitative methods for selecting potentially marked lexemes. The foundational concepts of markemology are discussed, and the stages of creating a markeme list for a literary text are demonstrated. The dominant and subdominant elements of the analyzed text are identified.*

Keywords: *markemological approach, formalized content analysis, markeme, markemology, filtering system.*

Внедрение компьютерно-математических методов в прикладную сферу лингвистических исследований не первое десятилетие волнует умы как отечественных, так и зарубежных ученых. В частности, маркемологический подход, рассматриваемый в данной статье, основан на содержательном и формализованном исследовании текста или совокупности текстов при использовании автоматизированного, компьютеризированного способа расчета определенной индивидуальной величины.

Одним из родоначальников квантитативного метода в лингвистике можно считать американского профессора Джозефину Майлз (1911-1985), первую женщину, принятую на работу на кафедру английского языка Калифорнийского университета в Беркли. Благодаря ее систематическому анализу и эмпирическим наблюдениям ей удалось продемонстрировать ценность статистических методов для изучения различных аспектов литературных произведений, среди которых выделяется частота встречаемости слова в тексте. Для своего исследования, которое впоследствии было изложено в работе *“The Continuity of Poetic Language: Studies in English Poetry from the 1540’s to the 1940’s”*, Дж. Майлз обратилась к произведениям английской поэзии 40-х годов XVI-XX веков [Miles 1951]. Основываясь на частотности употребления лексемы и на количестве авторов, в текстах которых она встречалась, ученая сумела выявить лексическое разнообразие языка того или иного поэта, основную и нерегулярную, базовую и индивидуальную лексику.

Маркемологический подход, стоящий во главе угла при исследовании массивов текстов с точки зрения маркемологии, продвинулся несколько дальше. Маркемология, будучи сравнительно молодым направлением лингвистики, предложенным д.ф.н., профессором А.А. Кретовым, предлагает формализованный содержательный, а следовательно, максимально объективный подход к анализу текста (преимущественно – художественного). Сам метод, активно развиваясь в научно-методическом центре компьютерной лингвистики Воронежского государственного университета, основывается на параметрическом анализе лексики и формулы веса слова [Титов 2002, 11]. Профессор, д.ф.н. В.Т. Титов высказал предположение о выявлении математической зависимости между двумя неизменными в контексте произведения или их совокупности параметрами слова. Первый параметр (Q-вес, Ч-вес, вес по частоте) обозначает частоту лексемы, количество ее употребления в тексте. Второй параметр (F-вес, Д-вес, вес по длине) обозначает длину лексемы, т.е. количество букв в данном слове [Кретов 2010, 139].

Маркемологический подход предполагает извлечение потенциальных маркем при использовании компьютерных технологий. Сами так называемые маркемы, наиболее важные языковые единицы, выделяются из частотного списка (или частотного словаря) одного или целой последовательности текстов автора или авторов. В своей монографии «Языковые ключи к английской литературе от Шекспира до Фаулза» д. ф. н. О.Г. Артемова предлагает называть маркемой «нарицательное, стилистически нейтральное имя существительное в единственном числе, именительном падеже, не являющееся обращением, названием месяцев, дней недели, литературных жанров, названием артефактов (кроме символов), словом-классификатором, не представляющее специфику жанра или направления и не выполняющее обстоятельственную функцию» [Артемова 2020, 33]. Важно отметить, что выбор именно имени существительного, а не какой-либо другой языковой категории, в качестве потенциальной маркемы обусловлен наименьшей маркированностью и

ориентацией на внеязыковую действительность. Более того, необходимо обратить внимание на то, что категория падежа важна во флективных языках, к коим относится русский язык. Что же касается английского языка, то данная характеристика оказывается нерелевантной, уступая место обстоятельственной функции как границе выделения маркем.

Компьютерное извлечение потенциальных маркированных лексических единиц осуществляется посредством программного комплекса автоматической обработки англоязычных текстов *“ProTemAl-Engl”*, разработанного в Воронежском государственном университете (автор – А.С. Гусельникова, научный руководитель – д.т.н. И.Е. Воронина, научный консультант – д.ф.н. А.А. Кретов). Данный программный комплекс анализирует файл формата *.txt*, содержащий в себе нужный для исследования текст или их совокупность, а затем вычисляет объективный показатель информативности лексемы – ИнТеМ (Индекс Текстуальной Маркированности), численно равный разности между весом словоформы по частоте и ее весом по длине: $\text{ИнТеМ} = \text{Ч-вес} - \text{Д-вес}$. Для вычисления ИнТеМа русских словоформ существует аналогичный программный комплекс «ТемАЛ» (автор – Ирина Попова, научный руководитель – д.т.н. И.Е. Воронина, научный консультант – д.ф.н. А.А. Кретов). Исходя из принципа целесообразности, в качестве маркем рассматриваются первые 50 лексем с наибольшим положительным значением ИнТеМа, прошедшие всю многоэтапную систему фильтров, которые и составляют ранжированный маркемный список текста или корпуса текстов.

Видится необходимым рассмотреть систему фильтров для выделения маркем. Так, например, А.А. Кретов предлагает следующую последовательность фильтров или критериев, которые должна пройти языковая единица, для того чтобы стать маркемой:

- 1) частеречный (только существительные);
- 2) грамматический (только словарные формы существительных);

3) грамматико-семантический (отсеиваются онимы и одушевленные существительные, кроме слова *человек*);

4) тематико-семантический (настраивается на конкретное произведение и конкретный жанр, к примеру: *явление* и *действие* в пьесах; воинская терминология; названия жанров; название месяцев и дней недели; имена героев; артефактов, если последние не обрели символических значений; и т.д.);

5) стилистический (отсеиваются стилистически окрашенные слова);

6) диалогический (отфильтровываются обращения, звательные формы: *Боже, старче, матушка* и т.п.);

7) классификационный (исключаются слова-классификаторы, фигурирующие в конструкциях «Сущ. И.п. + Сущ. Р.п.», не несущие собственной семантики: *множество в множество плодов*) [Кретов 2010, 2].

Рассмотрим процесс прохождения лексемы через озвученную систему фильтров на примере исторического романа сэра Вальтера Скотта «Уэверли, или Шестьдесят лет назад» и составим его маркемный список. После прохождения автоматизированной обработки произведения в программе “*ProTemAl-Engl*” мы получаем документ формата *Excel* (.xls /.xlsx) и таблицу с ранжированным по убыванию ИнТеМа списком всех слов в тексте. Рассмотрим первые 9 потенциальных маркем и проведем их через систему фильтрации.

№	Слово	Частота	Длина	Ч-вес	Д-вес	ИнТеМ
1.	LIEUTENANT-COLONEL	5	18	0,7718	0,1642	0,6076
2.	WAVERLEY-HONOUR	54	15	0,974	0,4092	0,5648
3.	NOTWITHSTANDING	22	15	0,9365	0,4092	0,5273
4.	CIRCUMSTANCES	87	13	0,985	0,5668	0,4182
5.	DISTINGUISHED	55	13	0,9746	0,5668	0,4078
6.	NON-COMMISSIONED	4	16	0,7338	0,3281	0,4057
7.	CORRESPONDENCE	12	14	0,888	0,4892	0,3988
8.	REPRESENTATIVE	10	14	0,8661	0,4892	0,3769
9.	DISAPPOINTMENT	9	14	0,8533	0,4892	0,3641

Лексемы № 1 (*lieutenant-colonel*) и № 8 (*representative*), являясь существительными одушевленными, и лексема № 2 (*Waverly-Honour*), являясь онимом – названием семейного поместья, не проходят критерий № 3 (грамматико-семантический). Единицы № 3 (предлог или глагольная форма *notwithstanding*), № 5 (глагольная форма *distinguished*), № 6 (глагольная форма *non-commissioned*) не попадают под частеречный критерий (№ 1). Слово № 4 (*circumstances*) находится во множественном числе, тем самым не отвечая критерию № 2 (грамматический). Лексема № 7 (*correspondence*) так же не может являться маркемой, будучи по контексту произведения артефактом, созданным людским трудом (критерий № 4 – тематико-семантический). Таким образом, только 9 позиция в общем списке потенциальных маркем изучаемого текста – *disappointment* – имеет право войти в маркемный список произведения. Данный процесс проверки продолжается до тех пор, пока маркемный список не наберет 50 наименований.

№	Маркема	Частота	Длина	Ч-вес	Д-вес	ИнТе	Значение
1.	DISAPPOINTMENT	9	14	0,853	0,489	0,3641	разочарование
2.	CONSIDERATION	13	13	0,897	0,566	0,3305	размышление
3.	RECOLLECTION	32	12	0,957	0,639	0,3174	воспоминание
4.	RECONCILIATION	6	14	0,799	0,489	0,3103	примирение
5.	CONSCIOUSNESS	9	13	0,853	0,566	0,2865	осознание
6.	EMBARRASSEMEN	9	13	0,853	0,566	0,2865	смущение
7.	INDIFFERENCE	18	12	0,921	0,639	0,2821	безразличие
8.	SATISFACTION	17	12	0,918	0,639	0,2786	удовлетворение
9.	DETERMINATION	8	13	0,838	0,566	0,2719	решимость
10.	GRATIFICATION	8	13	0,838	0,566	0,2719	удовлетворение
11.	OPPORTUNITY	53	11	0,973	0,706	0,2672	возможность
12.	IMAGINATION	45	11	0,968	0,706	0,2624	воображение
13.	DISPOSITION	30	11	0,953	0,706	0,2474	склонность
14.	HOSPITALITY	22	11	0,936	0,706	0,2302	гостеприимство
15.	APPREHENSION	10	12	0,866	0,639	0,2265	опасение
16.	OBSERVATION	16	11	0,913	0,706	0,2074	наблюдение
17.	UNDERSTANDING	5	13	0,771	0,566	0,205	понимание
18.	INCONVENIENCE	5	13	0,771	0,566	0,205	неудобство
19.	ENCOURAGEMENT	5	13	0,771	0,566	0,205	ободрение
20.	MORTIFICATION	5	13	0,771	0,566	0,205	унижение, стыд
21.	DIFFICULT(Y/IES)	39	10	0,964	0,765	0,1994	трудности
22.	REMONSTRANCE	8	12	0,838	0,639	0,1991	возражение
23.	SUPERSTITION	8	12	0,838	0,639	0,1991	суеверие
24.	INDIGNATION	14	11	0,904	0,706	0,1978	негодование
25.	ASSISTANCE	36	10	0,961	0,765	0,1961	помощь

26.	SELF-	3	14	0,681	0,489	0,1919	самоуверенность
27.	RESPONSIBILITY	3	14	0,681	0,489	0,1919	ответственность
28.	IMPORTANCE	32	10	0,957	0,765	0,1916	значимость
29.	DISSATISFACTIO	2	15	0,600	0,409	0,1911	недовольство
30.	MISCONSTRUCTI	2	15	0,600	0,409	0,1911	недопонимание
31.	PRINCIPLES	28	10	0,951	0,765	0,1856	принципы
32.	AFFECTATION	12	11	0,888	0,706	0,1817	манерность
33.	ASTONISHMENT	7	12	0,820	0,639	0,1811	изумление
34.	DISCIPLINE	26	10	0,946	0,765	0,1807	дисциплина
35.	ATTACHMENT	25	10	0,943	0,765	0,1785	привязанность
36.	PROTECTION	24	10	0,941	0,765	0,1761	защита
37.	CONFIDENCE	24	10	0,941	0,765	0,1761	уверенность
38.	RESOLUTION	23	10	0,939	0,765	0,1739	решительность
39.	ENTERTAINMENT	4	13	0,733	0,566	0,167	развлечение
40.	CONTRADICTION	4	13	0,733	0,566	0,167	противоречие
41.	EXPOSTULATION	4	13	0,733	0,566	0,167	упрек
42.	IMPRESSION	19	10	0,925	0,765	0,1602	впечатление
43.	INTELLIGENCE	6	12	0,799	0,639	0,1599	разум
44.	INTERFERENCE	6	12	0,799	0,639	0,1599	вмешательство
45.	APPROBATION	10	11	0,866	0,706	0,1598	одобрение
46.	DISPLEASURE	10	11	0,866	0,706	0,1598	недовольство
47.	EARNESTNESS	10	11	0,866	0,706	0,1598	серьезность
48.	DISTINCTION	10	11	0,866	0,706	0,1598	признание
49.	CURIOSITY	57	9	0,975	0,816	0,1591	любопытство
50.	REPUTATION	18	10	0,921	0,765	0,1563	репутация

Приведем несколько комментариев относительно некоторых моментов, благодаря которым были отсеяны определенные словоформы в представленном списке и изменены первоначальные показатели программы (жирный шрифт в таблице). Во-первых, длина слова в нашем случае считается согласно буквам в данном слове, поэтому показатель длины для маркемы *self-importance* был снижен на 1 из-за дефиса, и, следовательно, вес по длине, а затем и значение ИнТеМа тоже претерпевают изменения. Далее, нельзя не обратить внимания на маркему *understanding* в виду ее комплексной структуры. Данное слово в герундиальной форме не может является маркемой, а значит, показатель частоты был снижен в соответствии с частотностью ее употребления именно в форме существительного. На основе ранжирования маркем по убыванию значения ИнТеМа устанавливаются доминанта и вице-доминанта в тексте. В нашем случае с большим разрывом доминантой первого исторического романа сэра Вальтера Скотта «Уэверли, или Шестьдесят лет назад» является маркема *disappointment* –

разочарование, огорчение, вице-доминантой – маркема consideration – размышление, соображение.

Несмотря на автоматизированный процесс определения ИнТеМа, последующая проверка данных и этапы уточнения значений и контекстов являются чрезвычайно трудоемкой и скрупулезной деятельностью. Маркемологический подход позволяет также произвести распределение маркем по ядру и периферии текста на основе среднего значения ИнТеМа, сортировку маркем относительно семантических блоков. Таким образом, маркемологический подход к формализованному содержательному анализу текста представляет собой синтез объективного, автоматизированного и традиционного инструментов для структурированного исследования текстов с целью выявления ключевых тем и интенций, заложенных в произведениях литературы.

Список использованной литературы:

1. Артемова О.Г. Языковые ключи к английской литературе от Шекспира до Фаулза: монография. Воронеж: НАУКА-ЮНИПРЕСС, 2020. 596 с.
2. Кретов А. А. Динамика маркем в русской литературе XVIII – начала XX веков как отражение социокультурных процессов // Политическая лингвистика. 2010. № 3 (33). С. 141–150.
3. Кретов А.А. Понятие маркемы: методика выявления и практика использования // Универсалии русской литературы. Воронеж. 2010. С. 138–153.
4. Титов В.Т. Общая количественная лексикология романских языков монография. Воронеж: Изд-во Воронеж. гос. ун-та, 2002. 240 с.
5. Miles J. The Continuity of Poetic Language: Studies in English Poetry from the 1540's to the 1940's. Berkeley, CA: University of California Press, 1951. 572 p.
6. Oxford English Dictionary [Electronic Resource]. URL: <https://www.oed.com> (дата обращения 30.03.2024).

7. Sir Walter Scott. Waverley, or, 'Tis Sixty Years Since. In Three Volumes. Edinburgh: J.Ballantyne and Co, 1814. 365 p.

УДК 37.02

Zhao Xixi
Belarusian National Technical University (BNTU)
Scientific supervisor: Malashenko E.A.,
Ph.D. in Education, Associate Professor
Minsk, Republic of Belarus
Belarus State Economic University (BSEU)
malashenko@bseu.by

DEVELOPMENT OF ESSENTIAL PROFESSIONAL COMPETENCIES FOR FUTURE PSYCHOLOGY EDUCATORS

Abstract: *This study examines the historical and cultural foundations of psychology education in China and the Republic of Belarus, highlighting how cultural values and national identity shape educational practices and societal attitudes toward mental health. In China, psychology education reflects the influence of philosophical traditions like Confucianism and Taoism, whereas in Belarus, it combines Soviet educational influences with post-Soviet reforms. The research explores the key professional competencies required for future psychology educators, including cognitive, pedagogical, interpersonal, and applied skills. It underscores the importance of cross-cultural communication, ethical awareness, and emotional management in fostering effective educational practices.*

Key words: *Psychology education, communicative competencies, cognitive competencies, pedagogical competencies, cross-cultural communication, teacher training.*

Чжао Сиси
Белорусский национальный технический университет (БНТУ)
Научный руководитель: Малашенко Е.А.,
кандидат педагогических наук, доцент
Минск, Республика Беларусь
Белорусский государственный экономический университет
malashenko@bseu.by

РАЗВИТИЕ КЛЮЧЕВЫХ ПРОФЕССИОНАЛЬНЫХ КОМПЕТЕНЦИЙ У БУДУЩИХ ПРЕПОДАВАТЕЛЕЙ ПСИХОЛОГИИ

Аннотация: *В исследовании рассматриваются исторические и культурные основы психологического образования в Китае и Республике Беларусь, подчеркивая, как культурные ценности и национальная идентичность*