

ЭВОЛЮЦИЯ МАРКЕМНОЙ ЛЕКСИКИ НА ОСНОВЕ ИССЛЕДОВАНИЯ МАРКЕМ В ХРОНОЛОГИЧЕСКИХ СРЕЗАХ

О.Г. Артемова, Э.П. Комарова, А.А. Кретов

Аннотация. Рассматривается маркемный состав английских текстов второй половины XVII в. Актуальность обсуждаемой проблемы объясняется необходимостью решения задачи объективации и формализации содержательного анализа текста. До сих пор не выработаны четкие параметры, позволяющие, с одной стороны, объективировать характеристику стиля автора, а с другой – проследить взаимосвязь разных авторов и исследовать формирование языка литературы. Применение математических методов в лингвистических исследованиях позволяет выделить закономерности, которым подчиняются все тексты на естественных языках. Универсальные законы Дж. Ципфа, описывающие статистическую структуру текста на естественном языке, позволяют выделить из частотного словаря какого-либо текста или совокупности текстов слова, имеющие ключевое значение. Использование количественных методов для анализа литературных произведений дает возможность не только выявить закономерности в частотности употребления тех или иных слов, но и решить ряд других важных задач лингвистического исследования. В статье предлагается использование метода маркемного анализа как одного из способов формализации содержательного анализа текста. Он дает возможность получить полную картину маркемного состава языка литературных произведений любого хронологического среза или исторического периода, позволяя, при этом, исследовать тексты практически любого объема. Кроме того, метод маркемного анализа дает возможность исследовать специфику маркем отдельных авторов или групп авторов, влияние социокультурных процессов на динамику маркем, изучать эволюцию маркемной лексики на основе исследования маркем в нескольких хронологических срезах, а также посредством выделения общих маркем и определения на их основе коэффициента корреляции авторов устанавливать взаимосвязи различных авторов и, таким образом, обнаруживать преемственность в литературе. В основе метода лежат базовые понятия «маркема» и «индекс текстуальной маркированности» (ИнТеМ). ИнТеМ выражает зависимость между длиной словоформы и частотой ее употребления. Разность, полученная в результате вычитания веса словоформы по длине из ее веса по частоте, представляет собой авторскую составляющую, указывающую на степень весомости отдельной словоформы для конкретного текста. Обработка текстов производится созданным в Воронежском государственном университете программным комплексом тематического анализа лексики «ProTemAl-Engl». С целью определения взаимосвязи английских писателей второй половины XVII в. в созданных ими произведениях выявляется и анализируется маркемная лексика. Маркемами признаются наиболее значимые для автора текста словоформы. Это – имена существительные, как наименее маркированная часть речи, обозначающие важнейшие категории культуры, описываю-

щие внутренний мир человека, а также межличностные и социальные взаимоотношения, базовые понятия и т.д. Потенциальная маркама проходит специальную систему фильтров: грамматический, грамматико-семантический, тематико-семантический, стилистический, диалогический и классификаторный. Указанная система фильтров позволяет исключить попадание в число маркемонимов и одушевленных существительных, кроме словоформы *человек*; стилистически окрашенных слов; названий месяцев и дней недели; названий артефактов, если они не обрели коннотативных или символических значений; лексем, связанных с определенным жанром; обращений и слов-классификаторов. Зависимость величины ИнТеМа от веса словоформы по частоте ограничивает возможность его применения при проведении сопоставительного маркемного анализа, что обусловлено различным количеством как созданных отдельными авторами произведений, так и их доступностью в электронном виде. В этом случае целесообразно использовать *Нормированный ИнТеМ*, определяемый как отношение ИнТеМа маркамы автора к суммарному ИнТеМу всех его маркем, а его значение всегда находится в интервале 0–1. Анализ маркемного состава языка исследуемых литературных произведений предполагает выделение доминантной и вице-доминантной маркем хронологического среза, семантическую характеристику срезовых маркем. Выделение срезовых маркем дает возможность установить общие закономерности, существующие в литературе исследуемого хронологического среза. Параметром, определяющим попадание авторских маркем в число срезовых маркем, служит величина *суммарного нормированного интегрального веса*, вычисляемого как произведение *суммарного нормированного ИнТеМа* на количество авторов, употребивших данную маркаму. Первые 50 маркем сводного маркемного списка, ранжированного в порядке убывания суммарного нормированного интегрального веса, являются маркемами среза. Вычисление коэффициента корреляции в каждой паре авторов среза, базирующееся на значении суммарного нормированного индекса текстуальной маркированности их общих маркем, позволило установить близость авторов относительно друг друга. На основе максимального значения коэффициента корреляции определена сила максимальной векторной связи авторов и предложен способ визуализации основных векторных связей и их силы для авторов среза, что позволяет установить степень маркемной близости авторов. Исходя из количества связей максимальной силы и их направленности, выделены два центра аттракции – большой и малый, а также исследовано их количественное и содержательное маркемное наполнение. Результатом исследования стало проведение стратификации авторов среза, в результате которой установлен статус каждого автора: типичный представитель среза – нуклеар, нетипичный для среза – маргинал, занимающий промежуточное положение – медиар.

Ключевые слова: маркама; индекс текстуальной маркированности; коэффициент корреляции; центр аттракции; визуализация маркемной близости авторов; квантитативная лингвистика; английская литература.

Введение

По мнению историков английской литературы, развитие английской литературы XVII в. приходится на 1620–1690 гг. Последнее деся-

тилетию XVII в. знаменует начало нового периода истории английской литературы, положившего начало становлению просветительских тенденций. Спецификой развития английской литературы XVII в. в целом и его второй половины в частности является тесная связь с общественными процессами данного периода. Исторические события этой революционной эпохи обусловили эволюцию и этапы развития английской литературы. На 60–80-е гг. XVII в. приходится период развития противоречий и борьбы направлений в литературе. Начало ему положила реставрация династии в 1660 г., обусловившая концентрацию при дворе Карла II значительного круга литераторов, подверженных влиянию французского классицизма. Широкое распространение в это время получила теория государства Гоббса, изложенная им в 1651 г. Происходит осмысление результатов революционных и послереволюционных событий, нашедшее воплощение в трех поэмах Мильтона, две из которых, «Возвращенный рай» и «Самсон-борец», являются образцом классицистической поэтики.

В это же время в английской литературе с новой силой возродились барочные тенденции, особенно заметные в творчестве Драйдена. Они явили собой отражение атмосферы неуверенности и ощущения близости надвигающейся катастрофы, сохранявшиеся в английском обществе в период Реставрации. В то же время именно Драйден первым из английских писателей XVII в. высоко поднял национальную английскую традицию и обратил внимание на возможности, которые открывала поэтика классицизма. При этом классицизм Драйдена больше походил на развитие эстетических идей, высказанных Беном Джонсоном в начале века. Поворот английской драматургии в сторону классицизма заметен и в произведениях Т. Отвея и У. Уичерли [1]. Одним из первых против безнравственности английского общества эпохи Реставрации выступил Бэньян, обратившийся в своих произведениях к поиску истинного пути духовного формирования идеального человека, основой которого он считал самосовершенствование. В его творчестве прослеживаются одновременно черты барочного и реалистического направлений. В это время происходили становление и развитие нового этапа истории английской комедии – комедии Реставрации – яркими представителями которой являются Драйден, Уичерли и Этеридж [2]. Наряду с драматургией большой вклад в развитие английской литературы вносит художественная проза в лице Афры Бен, Дж. Мэкензи. Большое значение во второй половине XVII в. приобрели мемуары, представленные именами С. Пеписа и Дж. Ивлина. В этих условиях борьбы направлений шла подготовка новой литературной эпохи. Таким образом, вторая половина XVII в. является переломным моментом в истории английской литературы.

Историография вопроса и методология

Использование аналитических методов, основанных на применении математического аппарата в лингвистических исследованиях, имеет свою историю, началом которой принято считать 50-е гг. XX в., хотя уже в 1851 г. английский логик Август де Морган впервые предложил использовать математический аппарат для решения задачи установления авторства литературных текстов [3]. В лингвистике появление количественных методов связывают, прежде всего, с возникновением направления структурной лингвистики, т.е. началом XX в. Однако длительное время среди ученых, занимавшихся вопросами применения математических методов в лингвистических исследованиях, не существовало единства взглядов на цели и задачи математического направления в лингвистике. Так, немецкий ученый Э. Форстеманн [4], русские математики В.Я. Буняковский [5] и А.А. Марков [6, 7], английский статистик Джордж Юл [8] рассматривали языковые элементы как наиболее подходящий иллюстративный материал для количественных методов. При этом значимость результатов такого исследования для лингвистов находилась вне сферы их внимания. А. Росс [9] рассматривал теорию вероятностей и математическую статистику в качестве математической модели для проверки и подтверждения тех лингвистических выводов, которые допускают числовую трактовку [10].

Выделению закономерностей, которым подчиняются все тексты на естественных языках, посвящены работы ряда отечественных и зарубежных ученых. Эти исследования, в частности, выявили необходимость формализации статистических закономерностей языка, что нашло свое отражение в трудах Ж. Эсту [11], Э. Кондона [12], В. Болдрини [13], Дж. Ципфа [14, 15], Б.Н. Головина [16] и др. Именно Дж. Ципф вывел два универсальных закона, описывающих статистическую структуру любого текста на естественном языке. Согласно этим законам, произведение вероятности обнаружения некоторого слова в тексте на ранг его частоты – константа, сохраняющая свое значение для всех текстов на данном языке, а частота и количество слов, входящих в текст с данной частотой, связаны между собой. Эти закономерности позволяют выделить из частотного словаря какого-либо текста или совокупности текстов слова, имеющие ключевое значение. Также свой вклад в развитие количественных методов внесли американцы К. Шэннон и У. Уивер [17], применившие понятия теории информации для лингвистического описания, Г. Хердан, изучавший количественную лингвистику [18], отечественный филолог и автор трудов по математической лингвистике Р.Г. Пиотровский [19, 20], немецкий математик В. Фукс [21] и др.

Первоначально количественные методы интенсивно применялись при составлении частотных словарей, но еще в начале 40-х гг. прошлого столетия был предложен ряд приемов количественной оценки содержания текста (методы контент-анализа), оказавшиеся, к сожалению, малоэффективными ввиду необходимости наличия однородного языкового материала значительного объема [22].

Большой вклад в развитие квантитативной лингвистики внесла Жозефина Майлз, осуществив анализ лексического состава английских поэтических произведений XVI–XX вв. Использование для этой цели квантитативных методов позволило выявить закономерности в частотности употребления тех или иных слов. В частности, ею были отмечены такие особенности английских лексем, как преобладание в базовой лексике всех периодов односложных слов и переход слов из состава дополнительной лексики предшествующего хронологического периода в состав основной лексики последующего периода. При этом было установлено, что лексем, принадлежащие к основной лексике, более стабильны на протяжении нескольких веков, нежели принадлежащие дополнительной лексике, что и позволяет обеспечивать преемственность языка поэзии [23]. Таким образом, полученные Жозефиной Майлз данные, а также используемые в работе методы анализа материала и способы представления результатов исследования заложили основы для проведения подобных исследований.

В настоящее время лингвисты проявляют интерес к квантитативному исследованию языка английской поэзии [24]. В своих работах А.Б. Борунов [25, 26] исследует авторский корпус американского прозаика Рэгу Н. Митры. Сопоставительное исследование на материале английской поэзии и прозы, охватывающее творчество 14 авторов, принадлежащих XIV–XX вв., осуществили Л.Г. Кишинская и С.В. Кишинский [27]. Актуальность нашей работы определяется отсутствием квантитативных исследований языка английской прозы, позволяющих детально изучить литературные связи авторов-современников на лексическом уровне.

Метод формализованного изучения текстов – маркемный анализ – предложил А.А. Кретов. Применение метода маркемного анализа [28, 29] дает возможность получить полную картину маркемного состава языка английской прозы любого хронологического среза или исторического периода, позволяя, при этом, исследовать тексты практически любого объема: как правило, нижней границей пригодного для маркемного анализа текста являются 70 тыс. словоформ.

Основу данного метода исследования составляют понятия «маркема» и «индекс текстуальной маркированности» (ИнТеМ). Выделение маркем в текстах автора осуществляется на основе вычисления ИнТеМа и ранжирования словоформ в порядке его убывания. Обработка со-

бранных в единый файл английских текстов производится с использованием программного комплекса тематического анализа лексики «ProTemAL-Engl» (автор – А.С. Гусельникова, научный руководитель – д-р техн. наук И.Е. Воронина, научный консультант – д-р филол. наук А.А. Кретов).

Для проведения маркемного анализа в тексте выделяются 50 маркем – прошедшие через все фильтры словоформы (имена существительные) с максимальным положительным значением ИнТеМа, представляющим собой разность между весом слова по частоте и длине. Полученные списки словоформ проходят специальную систему фильтров, подвергаются контекстной проверке значений многозначных слов и исключению грамматической омонимии, в результате чего составляются индивидуальные маркемные списки авторов (подробно об используемой системе фильтров см.: [30]).

Поскольку величина ИнТеМа зависит от длины текста, его использование в качестве инструмента сопоставительного маркемного анализа, а также установления динамики маркем возможно при исследовании текстов одинаковой длины, что по ряду объективных причин не всегда возможно. К числу таких причин можно отнести различное количество созданных каждым автором произведений и их доступность в электронном виде.

В этом случае корректно провести маркемный анализ позволяет применение *Нормированного ИнТеМа* (НормИнТеМа), определяемого как отношение ИнТеМа маркемы автора к СумИнТеМу всех его маркем. Таким образом, НормИнТеМ вычисляется по формуле: $\text{НормИнТеМ} = \text{ИнТеМ} : \text{СумИнТеМ}$, а его значение всегда находится в интервале 0–1.

Разнообразие литературных направлений второй половины XVII в. оказало влияние на формирование маркемных систем авторов среза. Выделение срезовых маркем (СМ) позволяет установить общие закономерности, существующие в литературе исследуемого хронологического среза. Определение СМ осуществляется на основе сводного маркемного списка, составленного из маркемных списков авторов среза и ранжированного в порядке убывания *суммарного нормированного интегрального веса* (СумНормИнВес) его маркем. Величина СумНормИнВеса определяется произведением *СумНормИнТема* на *количество авторов*, употребивших данную маркему. Первые 50 маркем ранжированного по убыванию СумНормИнВеса сводного списка являются маркемами среза.

Исследование

Цель данной работы – выявление маркемной лексики английской прозы второй половины XVII в. Задача исследования – выделить СМ в прозаических текстах второй половины XVII в. (далее – срез 17-2) и определить стратификацию авторов среза.

Объектом исследования являются доступные в электронном виде произведения 16 писателей: Джона Бэньяна (John Bunyan – Bun), Афры Бен (Afra Behn – Bn), Ричарда Брэттуэйта (Richard Brathwaite – Br), Томаса Гоббса (Thomas Hobbes – Hb), Джона Драйдена (John Dryden – Dr), Джона Ивлина (John Evelyn – Ev), Абрахама Каули (Abraham Cowley – Cl), Джона Локка (John Locke – Lck), Джона Мильтона (John Milton – Ml), Джорджа Мэкензи (Sir George Mackenzie – Mck), Томаса Отвея (Thomas Otway – Ot), Сэмюэла Пеписа (Samuel Pepys – Ps), Уильяма Сейлза (William Sales – Sl), Уильяма Уичерли (William Wucherley – Wch), Ричарда Хэда (Richard Head – Hd), Джорджа Этериджа (George Etherege – Eth) (в скобках указаны оригинальные имена авторов и принятые в работе сокращенные обозначения авторов).

Общее количество исследованных в этом хронологическом срезе словоформ – 8 620 801.

Отнесение исследуемых текстов ко второй половине XVII в. в известной степени условно, поскольку некоторые текстов отдельных авторов формально принадлежат как первой, так и второй половине XVII в.

В срезе 17-2 отсутствует общая для всех исследуемых авторов маркема. Доминантная маркема *satisfaction* ‘удовлетворение’ (15; 7,718; первая цифра – количество авторов, употребивших маркему, вторая – ее СумНормИнВес) не отмечена в маркемном списке Д. Мильтона. Вице-доминантная маркема *understanding* ‘осмысление’ (13; 6,494) отсутствует в маркемных списках Дж. Мэкензи, У. Сейлза и Дж. Этериджа. Распределение НормИнТеМа доминантной и вице-доминантной маркем в срезе 17-2 представлено на гистограмме (рис. 1).

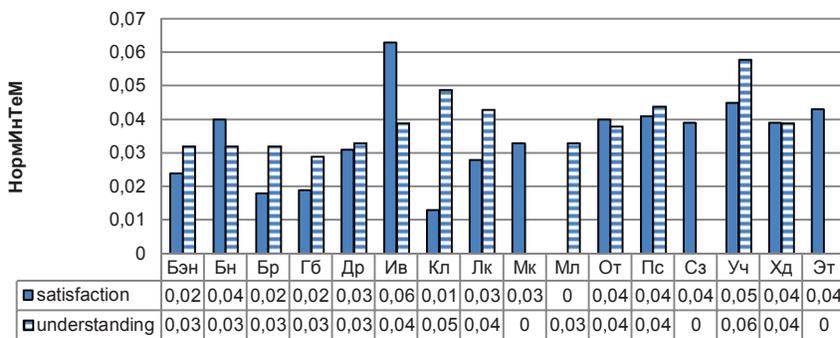


Рис. 1. Распределение НормИнТеМа маркем *satisfaction* / *understanding* во второй половине XVII в.

Высокий СумНормИнТеМ *understanding* обусловлен ее доминированием у А. Каули и Д. Локка. У Р. Брэттуэйта, С. Пеписа, У. Уичерли, Р. Хэда – это вице-доминантная маркема. Ранг *understanding* в маркемном списке Дж. Мэкензи, У. Сейлза и Дж. Этериджа отсутствует.

кемных списках других авторов находится в пределах 3–8. При этом ранг доминантной СМ *satisfaction* в авторских маркемных списках находится, в основном, в пределах 3–4 (А. Бен, Дж. Мэкензи, Т. Отвей, С. Пепис, У. Сейлз, У. Уичерли, Р. Хэд). В качестве доминантной она выделена только в списке Д. Ивлина. Достаточно высокий ранг маркемы отмечен в списках Д. Драйдена – 6, Д. Локка – 10, Д. Бэньяна – 13. И только у Т. Гоббса, Р. Брэтзэйтта ранг *satisfaction* находится в третьем десятке, у А. Каули – в четвертом десятке. Минимальный ранг – 47 – *satisfaction* занимает в маркемном списке Дж. Этериджа. Высокое значение СумНормИнТеМа *understanding* в срезе 17-2, очевидно, связано с особенностями рационализма эстетики Ренессанса, барокко и классицизма. Для эпохи Возрождения характерны представления о гармонии разума и чувства, в то время как для представителей барокко разум – это сила, дающая возможность противостоять хаосу жизни, а для классицизма характерно противопоставление рационального эмоциональному [31]. Поэтому вполне закономерно выделение в качестве доминантной маркемы «эмоциональной» *satisfaction*, а в качестве вице-доминантной – «ментальной» *understanding*.

В целом в маркемной лексике второй половины XVII в. высока доля «ментальных» и «эмоциональных» маркем – 22. Однако в числе СМ наблюдается незначительное преобладание «эмоциональных» маркем – 12. Тем не менее распространенность общих «ментальных» маркем в списках большинства авторов среза выше. Помимо вице-доминантной маркемы среза *understanding* в маркемных списках большинства авторов встречаются *conscience* ‘совесть’ (14; 4,581), *confidence* ‘вера, доверие’ (11; 2,152), *consideration* ‘размышление’ (11; 4,840), *expectation* ‘ожидание’ (11; 3,110), *pleasure* ‘желание, воля’ (11; 1,942), *remembrance* ‘память, воспоминание’ (10; 2,169), *imagination* ‘воображение’ (9; 1,377), *observation* ‘наблюдение’ (8; 1,272). И только *contemplation* ‘раздумье’ (6; 0,997) встречается менее чем у половины авторов среза. Среди «эмоциональных» маркем, вошедших в маркемные списки не менее восьми авторов среза, наряду с *satisfaction*, выделим *entertainment* ‘развлечение’ (12; 3,870), *happiness* ‘счастье, радость’ (11; 2,386), *pleasure* ‘удовольствие’ (11; 1,942), *acknowledgement* ‘признательность’ (8; 0,727). Остальные маркемы – *indifference* ‘равнодушие’ (6; 0,589), *repentance* ‘сожаление’ (6; 0,608), *admiration* ‘восторг’ (5; 0,524), *disappointment* ‘разочарование’ (5; 0,864), *displeasure* ‘недовольство’ (5; 0,438), *indignation* ‘негодование’ (5; 0,360), *astonishment* ‘удивление’ (4; 0,369) – характеризуют творчество около трети авторов среза. Среди «эмоциональных» маркем пять – *indifference*, *repentance*, *disappointment*, *displeasure*, *indignation* – имеют отрицательную окраску. Максимальное число отрицательных маркем в авторских списках СМ – 3 – выделено у Д. Бэньяна, А. Бена, С. Пеписа. Общей для них является *disappointment*. *Indignation* объединяет Д. Бэньяна и А. Бена и при этом имеет равный НормИнТеМ – 0,014. Общей для А. Бена и С. Пеписа является *indif-*

ference. Пара *displeasure – indignation* выделена у двух авторов – Д. Мильтона и У. Сейлза. По две отрицательных маркеме в разной комбинации выделены у Д. Драйдена – *indifference, displeasure*, у Д. Ивлиина – *repentance, displeasure*, Дж. Мэкензи – *indifference, disappointment*, у Дж. Этериджа – *repentance, indignation*. По одной отрицательной маркеме выделено в текстах Р. Брэттуэйта, Д. Локка – *indifference* – и Т. Гоббса, Т. Отвея, Р. Хэда – *repentance*, а также У. Уичерли – *disappointment*. И только в текстах А. Каули отрицательные «эмоциональные» маркемы отсутствуют. Наиболее полно «эмоциональные» маркемы представлены у С. Пеписа – 9, а также у А. Бен и У. Сейлза – 8.

Таким образом, срез в целом характеризуется преобладанием положительных эмоциональных состояний, вызывающих *восторг, изумление, удовлетворение, удовольствие, радость* и чувство *признательности*.

Значимыми для трети и более авторов оказываются и маркемы, характеризующие личностные качества человека – *resolution* ‘решительность’ (9; 1,371), *covetousness* ‘алчность’ (7; 1,104), *greatness* ‘величие, мощь’ (7; 0,817), *dissimulation* ‘лицемерие’ (6; 0,938), *righteousness* ‘праведность’ (6; 1,304), *jealousy* ‘ревность’ (5; 0,461). Несколько меньше представлена маркема *self-preservation* ‘самосохранение’ (4; 0,514). В той или иной степени эти маркемы присутствуют в текстах всех авторов среза. Максимально маркемы этой категории выделены у Р. Брэттуэйта – 5. В текстах А. Бен, Д. Драйдена, Р. Хэда обнаружено четыре маркемы, а у Т. Гоббса, А. Каули, Дж. Мэкензи, Т. Отвея, У. Уичерли и Дж. Этериджа – три. У остальных авторов отмечены одна (Д. Ивлин, Д. Локк, С. Пепис) или две (Д. Бэньян, Д. Мильтон, У. Сейлз) маркемы.

Широко представлена в числе СМ группа маркем, выражающих межличностные отношения, – *friendship* ‘дружба’ (11; 3,056), *assistance* ‘помощь’ (9; 1,516), *contradiction* ‘противоречие, конфликт’ (7; 1,123), *reconciliation* ‘примирение’ (7; 1,340), *encouragement* ‘ободрение’ (6; 1,113), *affection* ‘привязанность’ (5; 0,531), *commendation* ‘одобрение’ (5; 0,531), *forgiveness* ‘прощение’ (5; 0,510). Они характерны для всех авторов, но максимальное представление получили в маркемном списке Д. Драйдена – 7. В текстах этого автора отсутствует только *affection*. Следующими по количеству маркем идут Р. Хэд – 6, А. Бен – 5, Д. Бэньян, Д. Ивлин, Дж. Мэкензи, У. Сейлз – 4, Т. Отвей, С. Пепис, Дж. Этеридж – 3. Меньше всего этих маркем у Р. Брэттуэйта, Т. Гоббса, А. Каули, Д. Локка, Д. Мильтона, У. Уичерли – 2.

Высокочастотной оказалась одна маркема, выражающая морально-нравственные ценности – *innocence* ‘невинность, целомудрие’ (8; 1,072). Важной она является для А. Бен, Д. Драйдена, А. Каули, Дж. Мэкензи, Д. Мильтона, Т. Отвея, У. Сейлза и У. Уичерли.

Кроме того, важными для большинства представителей среза маркемами являются маркемы, выражающие фундаментальные поня-

тия, связанные, в основном, с категорией бытийных отношений – *opportunity* ‘благоприятная возможность, шанс’ (14; 5,619), *experience* ‘(жизненный) опыт’ (13; 2,813), *knowledge* ‘знание, знания’ (11; 2,132), *punishment* ‘наказание’ (11; 2,079), *destruction* ‘разрушение’ (10; 2,841), *advantage* ‘преимущество’ (9; 1,310), *difficult(y/ies)* ‘трудность, трудности’ (9; 1,755), *necessity* ‘необходимость’ (9; 1,211), *perfection* ‘безукоризненность’ (9; 1,261), *occasion* ‘удобный случай, возможность’ (8; 0,891). Более чем у трети авторов среза выделены *misfortune* ‘беда, несчастье’ (7; 0,851), *creature* ‘человек, создание’ (6; 0,440), *providence* ‘провидение, Бог’ (6; 0,671).

Опираясь на количество всех общих маркем (ОМ) в каждой паре авторов, можно установить типичность авторов относительно друг друга. С этой целью был вычислен коэффициент корреляции (КК) авторов среза. Величина КК равна произведению СумНормИнТеМов ОМ двух авторов. В табл. 1 представлены значения КК в каждой паре авторов, округленные до сотых.

Мы использовали КК в качестве единицы измерения силы связи между авторами. Установив в каждой горизонтальной строке максимальное значение КК, мы, тем самым, определили силу максимальной векторной связи авторов. Учет только связей максимальной силы для каждой пары авторов позволил построить граф, отражающий основные векторные связи и их силу для авторов среза 17-2 (рис. 2). Для удобства представления и восприятия информации значения КК на графе представлены в процентном формате и округлены до целых.

Таблица 1
Значение коэффициента корреляции в парах авторов в срезе 17-2

Ав-торы	Bun	Bn	Br	Hb	Dr	Ev	Cl	Lck	Mck	MI	Ot	Ps	Sl	Wch	Hd	Eth
Bun		0,13	0,08	0,14	0,18	0,10	0,12	0,05	0,04	0,06	0,04	0,07	0,08	0,09	0,16	0,10
Bn	0,13		0,14	0,07	0,34	0,22	0,21	0,13	0,25	0,08	0,28	0,26	0,27	0,30	0,42	0,12
Br	0,08	0,14		0,06	0,16	0,10	0,19	0,08	0,07	0,04	0,06	0,16	0,13	0,11	0,19	0,10
Hb	0,14	0,07	0,06		0,21	0,04	0,06	0,07	0,01	0,05	0,03	0,05	0,06	0,04	0,13	0,02
Dr	0,18	0,34	0,16	0,21		0,22	0,27	0,22	0,15	0,09	0,20	0,24	0,25	0,19	0,41	0,11
Ev	0,10	0,22	0,10	0,04	0,22		0,11	0,08	0,20	0,11	0,18	0,26	0,27	0,13	0,23	0,16
Cl	0,12	0,21	0,19	0,06	0,27	0,11		0,14	0,14	0,07	0,15	0,12	0,13	0,21	0,27	0,08
Lck	0,05	0,13	0,08	0,07	0,22	0,08	0,14		0,07	0,04	0,03	0,16	0,08	0,07	0,21	0,02
Mck	0,04	0,25	0,07	0,01	0,15	0,20	0,14	0,07		0,04	0,10	0,17	0,18	0,13	0,18	0,12
MI	0,06	0,08	0,04	0,05	0,09	0,11	0,07	0,04	0,04		0,08	0,04	0,10	0,06	0,09	0,09
Ot	0,04	0,28	0,06	0,03	0,20	0,18	0,15	0,03	0,10	0,08		0,10	0,16	0,19	0,19	0,24
Ps	0,07	0,26	0,16	0,05	0,24	0,26	0,12	0,16	0,17	0,04	0,10		0,15	0,20	0,26	0,09
Sl	0,08	0,27	0,13	0,06	0,25	0,27	0,13	0,08	0,18	0,10	0,16	0,15		0,10	0,25	0,10
Wch	0,09	0,30	0,11	0,04	0,19	0,13	0,21	0,07	0,13	0,06	0,19	0,20	0,10		0,27	0,18
Hd	0,16	0,42	0,19	0,13	0,41	0,23	0,27	0,21	0,18	0,09	0,19	0,26	0,25	0,27		0,17
Eth	0,10	0,12	0,10	0,02	0,11	0,16	0,08	0,02	0,12	0,09	0,24	0,09	0,10	0,18	0,17	

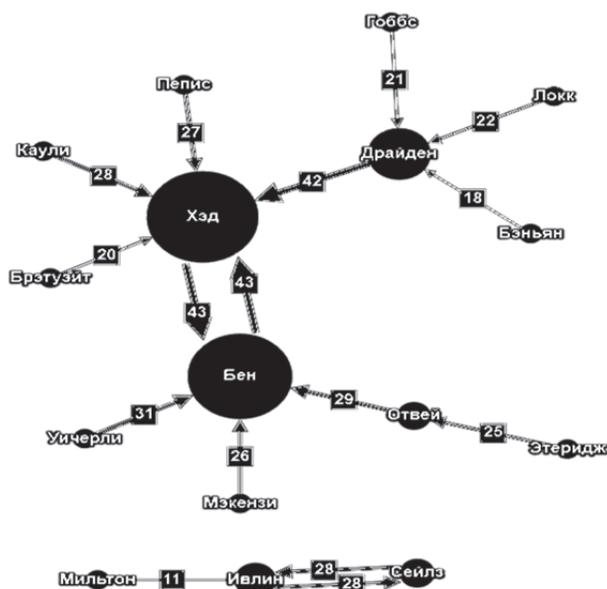


Рис. 2. Предпочтительные связи английских писателей второй половины XVII в. (по данным КК)

Всего на графе 16 связей. Одновременный учет количества и векторной ориентированности связей позволил выделить два центра аттракции: большой и малый. Центром аттракции мы называем объединение авторов, имеющих взаимонаправленные связи. В структуру графа также входят авторы, имеющие разнонаправленные связи – Д. Драйден, Т. Отвей, У. Сейлз. Эти авторы образуют центры-посредники (ЦП). Поскольку все ЦП имеют непосредственную связь с ЦА, мы назвали их центрами-посредниками первой степени (ЦП-1). Периферию графа составляют терминальные (конечные) авторы – Т. Гоббс, Д. Локк, Д. Бэньяна, Дж. Этеридж, Дж. Мильтон. Их характеризует наличие только одного вида связи – центробежной. Под центробежной связью мы понимаем связь максимальной силы, которая направлена от данного автора к ближайшему для него по КК автору.

Большой центр аттракции (БЦА) – Хэд – Брэттуэйт – Каули – Пепис – Драйден – Бен – Уичерли – Мэкензи – Отвей. Все авторы БЦА группируются возле двух связующих фигур Р. Хэда и А. Бен, имеющих двустороннюю связь одинаковой силы. Группу авторов, имеющих центробежную связь с Р. Хэдом, образуют С. Пепис, А. Каули, Р. Брэттуэйт и Д. Драйден. Группу А. Бен составляют У. Уичерли, Дж. Мэкензи, Т. Отвей. В состав БЦА входят два ЦП-1, связующими фигурами которых являются Д. Драйден и Т. Отвей. ЦП-1 Д. Драйдена характеризует наличие трех центробежных связей – Т. Гоббса, Д. Локка и Д. Бэньяна, а ЦП-1 Т. Отвея – центробежная связь Дж. Этериджа. С учетом вектор-

ной направленности связи получаем две цепочки: Гоббс – Локк – Бэньян – Драйден – Хэд, замыкающуюся на Р. Хэде, и Этеридж – Отвей – Бен, замыкающуюся на А. Бен. Сила связи Р. Хэда с Д. Драйденом и А. Бен отличается незначительно – 42 и 43. Максимальная для среза сила двусторонней связи – 43 – характеризует пару Хэд – Бен. При этом наличие взаимонаправленной связи Хэд – Бен при отсутствии связей максимальной силы между авторами, группирующимися возле Р. Хэда и А. Бен, позволяет рассматривать Р. Хэда и А. Бен еще как и авторов, образующих два ЦП: ЦП Хэда и ЦП Бен. Суммарная мощность БЦА с учетом силы максимальной связи составляет 461.

Рассматривая количественное и содержательное маркемное наполнение центра относительно Р. Хэда, отметим, что больше всего ОМ с ним имеет А. Бен – 30, из них СМ – 27. Далее следуют Д. Драйден – 29 (СМ – 27), С. Пепис – 23 (СМ – 19), А. Каули – 22 СМ, Р. Брэттуэйт – 20 (СМ – 18). Общими для всех авторов являются семь маркем – «ментальные» *confidence, conscience, consideration, understanding*; «эмоциональные» *entertainment, satisfaction*; «фундаментальная» *opportunity*.

Следующая группа из восьми маркем объединяет три группы авторов: Хэд – Бен – Брэттуэйт – Драйден – Пепис; Хэд – Брэттуэйт – Драйден – Каули – Пепис; Хэд – Бен – Брэттуэйт – Драйден – Каули. Больше всего общих маркем – 4 – представлено в группе Хэд – Брэттуэйт – Драйден – Каули – Пепис. Это – «ментальная» СМ *observation* и «фундаментальные» СМ *knowledge, experience, necessity*. «Эмоциональная» маркема *acknowledgement* и выражающая межличностные отношения маркема *friendship* объединяют авторов группы Хэд – Бен – Брэттуэйт – Драйден – Пепис, а «ментальная» *remembrance* и «эмоциональная» *happiness* – Хэд – Бен – Брэттуэйт – Драйден – Каули.

Самая многочисленная группа маркем – 12 – выделена в группах из четырех авторов. Общими для группы Хэд – Бен – Драйден – Пепис являются «фундаментальные» *advantage, difficult(y/ies)*, «ментальная» *expectation* и «межличностная» *assistance*. Вторая по количеству ОМ группа Хэд – Бен – Драйден – Каули объединена «ментальной» *imagination*, «межличностной» *encouragement* и «фундаментальной» *punishment*. Маркемы, характеризующие личностные качества человека, *covetousness* и *resolution*, объединяют Хэда с Бен, Брэттуэйтом, Драйденом и Каули, причем обе маркемы характеризуют группу Хэд – Брэттуэйт – Драйден. У Бен выделена только *resolution*, а у Каули – *covetousness*.

Еще три общие СМ характеризуют связь Хэда с Каули – Бен – *misfortune*, Каули – Брэттуэйт – *contemplation*, Каули – Драйден – *destruction*. По две маркемы связывают Хэда с Бен – Драйденом – *complaisance* ‘почтительность, услужливость’ (ОМ), *contradiction* (СМ), Бен – Пеписом – *melancholy* ‘уныние, тоска’ (ОМ), *occasion* (СМ).

Общими только для отдельных пар авторов являются, главным образом, маркемы, не вошедшие в число СМ, – *inconstancy* ‘непостоянство’ (Хэд – Бен); *concupiscence* ‘вожделение’, *impression* ‘впечатление’ (Хэд – Брэттуэйт); *commiseration* ‘сочувствие, сострадание’, СМ *self-preservation* (Хэд – Драйден); *convenience* ‘удобство, комфорт’, *kindness* ‘привязанность, нежность’ (Хэд – Пепис).

Таким образом, содержательный состав маркем, обеспечивающих основные маркемные связи авторов большого центра аттракции, составили 11 «фундаментальных», 9 «ментальных», 7 «эмоциональных», 6 маркем «личностных» и 5 «межличностных» маркем.

Малый (изолированный) центр аттракции (МЦА) представляют Ивлин – Мильтон – Сейлз. Связующей фигурой центра является Д. Ивлин. Он же является ключевой фигурой ЦП-1, опосредующего связь Д. Мильтона с У. Сейлзом. Суммарная мощность этого центра аттракции – 67.

По количеству связей центральной фигурой выступает Р. Хэд – шесть непосредственных связей общей мощностью 203 (при определении количества связей учитывается их направленность, поэтому двусторонняя связь считается как две связи). На втором месте А. Бен: пять связей общей мощностью 172. На третьем месте – Д. Драйден с четырьмя связями общей мощностью 103. Следует отметить, что все три автора объединены связями максимальной силы, образуя системообразующую триаду среза – с Хэдом в качестве центральной фигуры.

По суммарному значению максимальной силы связи рейтинг авторов первой половины XVII в. выглядит следующим образом (табл. 2).

Мы сопоставили корреляцию рейтинга авторов, основанного на значениях максимальной силы, со стратификацией авторов, составленной на основе СумНормИнТеМа только общих СМ, представленной на рис. 3. Критерием определения принадлежности автора к одному из классов является значение разности суммы НормИнТеМов СМ каждого автора и среднего значения суммы НормИнТеМов СМ всех авторов среза.

Типичными представителями своего среза – *нуклеарями* – являются Р. Хэд, Д. Драйден, А. Бен. Список авторов полностью совпадает со списком, выделенным в рейтинге, с той разницей, что Драйден и Бен поменялись местами. Это обусловлено тем, что данные рейтинга опираются только на величины максимальной силы и не учитывают другие связи между авторами среза и вектор их направления. Нетипичными для среза – *маргиналами* – определяются Д. Бэньян, Д. Локк, Т. Гоббс, Д. Мильтон. Отличие от рейтинга касается места в нем Р. Брэттуэйта – 14. Однако, делая поправку на то, что данные рейтинга показывают лишь основные связи авторов, это отличие можно считать несущественным. Промежуточное положение занимают *медиары* – Д. Ивлин, С. Пепис, У. Сейлз, У. Уичерли, А. Каули, Т. Отвей, Дж. Мэкензи, Дж. Этеридж, Р. Брэттуэйт. В рейтинге они также занимают промежу-

точное положение. В целом при учете количества и вектора связей авторов, а также их отношения к центрам аттракции, наблюдается полное совпадение графических и аналитических данных.

Таблица 3

Рейтинг авторов первой половины XVII в. по суммарной силе связи

№	Автор	Количество связей	Сила связи
1	Хэд	6	203
2	Бен	5	172
3	Драйден	4	103
4	Ивлин	3	67
5	Сейлз	2	56
6	Отвей	2	54
7	Уичерли	1	31
8	Каули	1	28
9	Пепис	1	27
10	Мэкензи	1	26
11	Этеридж	1	25
12	Локк	1	22
13	Гоббс	1	21
14	Брэттуэйт	1	20
15	Бэньян	1	18
16	Мильтон	1	11

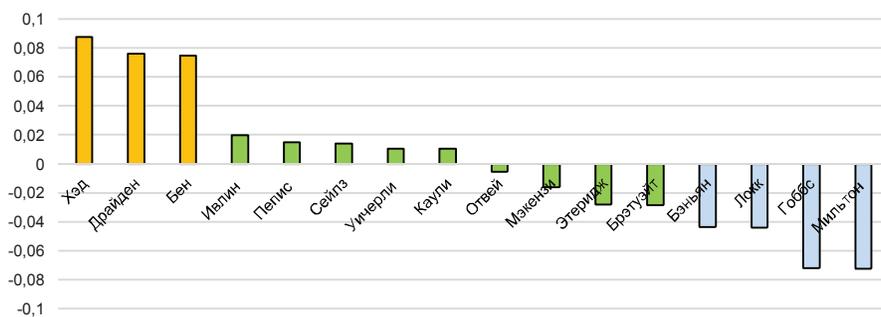


Рис. 3. Стратификация английских авторов второй половины XVII в. относительно среднего значения

Заключение

Проведенное маркерное исследование текстов английских авторов второй половины XVII в. позволило установить:

1. Доминантой данного хронологического среза является маркема *satisfaction*, вице-доминантой – маркема *understanding*.

2. В срезе отсутствуют общие для всех авторов маркемы.

3. Срез характеризуется незначительным преобладанием «эмоциональных» маркем над «ментальными». В маркерных списках авторов «ментальные» маркемы представлены шире.

4. В срезе высока доля маркем, характеризующих личностные качества человека, межличностные отношения и фундаментальные понятия.

5. В срезе 17-2 выделяются два центра аттракции. Связующими фигурами большого центра аттракции являются Р. Хэд и А. Бен, малого – Д. Ивлин и У. Сейлз.

6. Структуру большого центра аттракции образуют четыре центра-посредника – ЦП Хэда, ЦП Бен, ЦП Драйдена, ЦП Отвея. ЦП Хэда и ЦП Бен являются центрами-посредниками относительно друг друга вследствие наличия взаимонаправленной связи Р. Хэда и А. Бен и отсутствия маркемных связей максимальной силы между авторами, группирующимися около каждого из них.

7. В составе малого центра аттракции выделен один центр-посредник – ЦП Ивлины – опосредующий связь Д. Мильтона и У. Сейлза.

8. В результате стратификация авторов среза установлен статус каждого автора: *нуклеары* – Р. Хэд, Д. Драйден, А. Бен; *медиары* – Д. Ивлин, С. Пепис, У. Сейлз, У. Уичерли, А. Каули, Т. Отвей, Дж. Мэкензи, Дж. Этеридж, Р. Брэттуэйт; *маргиналы* – Д. Бэньян, Д. Локк, Т. Гоббс, Д. Милтон.

Литература

1. **Бердников Г.П.** (гл. ред.) История всемирной литературы : в 9 т. М. : Наука, 1989. Т. 4. URL: <http://litlife.club/br/?b=204319&p=101>
2. **Ерофеева Н.Е.** Зарубежная литература XVII век : учеб. для пед. вузов. URL: <http://17v-euro-lit.niv.ru/17v-euro-lit/erofeeva-uchebnik/index.htm>
3. **Сидоров Ю.В.** Математическая и информационная поддержка методов обработки литературных текстов на основе формально-грамматических параметров : дис. ... канд. техн. наук. Петрозаводск, 2002. 127 с.
4. **Forstemann E.** Commentary on the Maya manuscript in the royal public library of Dresden. Published by the Museum, 1906. 266 p. URL: [http://нэб.рф/search/?f_field\[authorbook\]=f/authorbook/Forstemann+E](http://нэб.рф/search/?f_field[authorbook]=f/authorbook/Forstemann+E).
5. **Буняковский В.Я.** Основания математической теории вероятности. М. : Книга по Требованию, 2017. 502 с. URL: <https://www.bookvoed.ru/files/3515/36/75/0.pdf>
6. **Марков А.А.** Исследования замечательного случая зависимых испытаний // Известия Императорской академии наук. Сер. VI. 1907. Т. 1, № 3. С. 60–80. URL: http://www.mathnet.ru/php/archive.phtml?wshow=paper&jmid=im&paperid=7832&option_lang=rus
7. **Марков А.А.** Пример статистического исследования над текстом «Евгения Онегина», иллюстрирующий связь испытаний в цепь // Известия Императорской академии наук. Сер. VI. 1913. Т. 7, № 3. С. 153–162. URL: http://www.mathnet.ru/php/archive.phtml?jmid=im&option_lang=rus&paperid=6612&wshow=paper
8. **Yule Udny G.** A Statistical Study of Vocabulary. Cambridge, 1944. Vol. IX. 306 p.
9. **Ross A.** Philological Probability Problems // Statist. Soc. Ser. B. 1950. Vol. XII. P. 19–59.
10. **Верхошин С.С.** О статусе количественных методов в лингвистике // Вестник Иркутского государственного лингвистического университета. 2013. Вып. № 3 (24). С. 145–150.
11. **Estoup J.** Gammestenographique. Paris, 1916.
12. **Condon E.V.** Statistics of Vocabulary // Science. 1928. Vol. 67. P. 1733.
13. **Boldrini V.** Le staticheletterarie e ifonemielementarinellapoesia. Milano, 1948.

14. Zipf G.K. Human Behavior and the Principle of Least Effort. Cambridge : Addison-Wesley Press, 1949. 533 p.
15. Zipf G.K. The Psycho-Biology of Language an introduction to dynamic philology. Cambridge : Mass. MITPress, 1965. 336 p.
16. Головин Б.Н. Язык и статистика. М. : Просвещение, 1971. 192 с.
17. Shannon C., Weaver W. The Mathematical Theory of Communication. Univ of Illinois Press, 1949. 117 p.
18. Herdan G. Language as Choice and Chance. Noordhoff : Groningen, 1956. 356 p.
19. Пиотровский Р.Г. Математическая лингвистика : учеб. пособие М. : Высш. шк., 1977. 383 с.
20. Пиотровский Р.Г. Инженерная лингвистика и теория языка. Л. : Наука, 1979. 111 с.
21. Fucks W. Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen. Köln und Oplade, 1955. 110 p.
22. Марусенко М.А. Атрибуция анонимных и псевдонимных текстов методами прикладной лингвистики // Прикладное языкознание. СПб., 1996.
23. Miles J. The continuity of poetic language; the primary language of poetry, 1540's–1940's. NY : Octagon Books, 1965.
24. Слугина О.В. Лингвокогнитивные основы поэтической квантитативности: на материале английского языка : дис. ... канд. филол. наук. Саранск, 2009. 200 с.
25. Борунов А.Б. Параметрическая стратификация лексики авторского англоязычного корпуса (на материале авторского корпуса американского писателя Рэгу Н. Митры) // Филологические науки. Вопросы теории и практики. 2016. № 12 (66) : в 4 ч. Ч. 1. С. 65–68.
26. Борунов А.Б. Квантитативное исследование англоязычного авторского корпуса Рэгу Н. Митры // Litera. 2017. № 1. С. 76–87. URL: http://e-notabene.ru/fil/article_20090.html
27. Кишинская Л.Г., Кишинский С.В. Выявление опережающей роли поэтической речи в развитии языка с помощью корреляционного анализа // Ученые записки Тартуского государственного университета. Лингвостатистика и квантитативные закономерности текста. Труды по лингвистике. 1980. Т. VI. С. 47–58.
28. Кретов А.А. Понятие маркемы: методика выявления и практика использования // Универсалии русской литературы. Воронеж, 2010. С. 138–153.
29. Гусельникова А.С., Воронина И.Е., Кретов А.А. Выделение тематически маркированной лексики в текстах на естественных языках // Материалы XV Международной научно-методической конференции «Информатика: проблемы, методология, технологии», 12–13 февраля 2015 г. Воронеж, 2015. Т. 3. С. 245–250.
30. Фаустов А.А., Кретов А.А. Понятие маркемы и предварительные итоги маркемного анализа русской литературы // Вестник Воронежского государственного университета. Сер. Лингвистика и межкультурная коммуникация. 2017. № 4. С. 16–31.
31. История зарубежной литературы XVII века : учеб. / под общ. рук. З.И. Плавскина. URL: <https://www.libfox.ru/367790-zahariy-plavskin-istoriya-zarubezhnoy-literatury-xvii-veka.html>

Источники

1. <http://www.gutenberg.org/>
2. <https://quod.lib.umich.edu/e/eebo/>
3. <https://archive.org/>
4. <http://oll.libertyfund.org/>

Сведения об авторах:

Артемова Ольга Григорьевна – кандидат филологических наук, доцент кафедры иностранных языков и технологии перевода Воронежского государственного технического университета (Воронеж, Россия). E-mail: olgaartemova65@yandex.ru

Комарова Эмилия Павловна – доктор педагогических наук, профессор кафедры иностранных языков и технологии перевода Воронежского государственного технического университета (Воронеж, Россия). E-mail: vivtkmk@mail.ru

Кретов Алексей Александрович – доктор филологических наук, профессор, заведующий кафедрой теоретической и прикладной лингвистики, директор Научно-методического центра компьютерной лингвистики Воронежского государственного университета (Воронеж, Россия). E-mail: tipl@rgph.vsu.ru

Поступила в редакцию 29 мая 2018 г.

ENGLISH REVOLUTION THROUGH THE PRISM OF MARKEMES

Artemova O.G., Ph.D (Philology), Associate Professor, the Department of Foreign Languages and Translation Technology, Voronezh State Technical University (Voronezh, Russia). E-mail: olgaartemova65@yandex.ru

Komarova E.P., Dr.Sc. (Edu), Professor, the Department of Foreign Languages and Translation Technology, Voronezh State Technical University (Voronezh, Russia). E-mail: vivtkmk@mail.ru

Kretov A.A., Dr.Sc. (Philology), Professor, the Head of the Department of Theoretical and Applied Linguistics, Voronezh State University (Voronezh, Russia). E-mail: tipl@rgph.vsu.ru

DOI: 10.17223/19996195/42/2

Abstract. This paper studies markeme composition in the texts of English writers of the latter half of the seventeenth century. The significance of the discussed problem is proved by the need to objectify and formalize semantical analysis of the text. Clear parameters that provide both the objectification of the author literary style characteristics, determination of the interlinkage between different authors and the study of the literary language formation process haven't been formulated yet. The use of mathematical methods in linguistic research allows revealing regularities all texts written in natural language are complied with. Zipf's universal laws that describe the statistic structure of a natural language text allow extracting key words from frequency word book of any text or a number of texts. Quantitative methods used to analyze literary works make it possible both to reveal regularities in any word frequency and to solve a number of other tasks critical for linguistic research. The method of markeme analysis is suggested to be used as one of the means to formalize the semantical analysis of the text. It provides a means of presenting a complete picture of literary works language markeme composition of any chronological interval or historical period. It also gives the possibility to analyze texts practically of any wordage. Besides the method of markeme analysis allows analyzing markeme specifications of individual authors or groups of authors and the influence of social and cultural processes on markeme dynamics, studying the evolution of markeme vocabulary through several chronological intervals and establishing links between different authors due to a correlation coefficient. Its value depends on the quantitative parameters of mutual markemes. The links established between authors provide an opportunity to reveal the continuity in literature. The method of markeme analysis is based on the terms *markeme* and *Index of Textual Markedness* (InTeM). InTeM expresses the correspondence between the word length and its frequency. The difference between the word weight due to its frequency and the word weight due to its length represents the author's component that marks the level of significance for each individual word in the text. The texts are processed with word thematic analysis program "ProTemAl-Engl" developed in Voronezh State University. To determine the inter linkage between English writers of the latter half of the seventeenth century it is necessary to identify and analyze markeme vocabulary. Those words that are the most significant for the author are considered as markemes. They are nouns, which are the least marked part of speech, that denote the most important categories of culture, describe the inward man and his interpersonal and human relations, basic notions, etc. The potential

markeme should undergo special system of filtration including grammatical, grammar and semantic, thematic and semantic, dialogical and classifying filters. This system of filters prevents including onyms and animate nouns, except words *man*, *creature*; connotative words; names of months and days of week; names of artifacts in case they do not possess connotative or symbolic meanings; words associated with a definite literary genre; addresses and classifying words. The dependence of InTeM value on the weight of word frequency restricts potential for its use in comparative markeme analysis. This is due to the fact that too often there is great difference both in number of works and their length in words written by different authors and their availability in digital form. In such cases it is appropriate to use *Normalized InTeM* instead of InTeM itself. Normalized InTeM is defined as the ratio of a markeme InTeM to total InTeM of all markemes of the author. The value of Normalized InTeM always belongs to the interval 0–1. The analysis of markeme composition of the studied literary works language requires the establishing of dominant and vice-dominant markemes of the chronological interval and semantic characteristic of the interval markemes. The selection of the interval markemes allows establishing general regularities that exist in the literature of the given chronological interval. *Total normalized integral weight* is a parameter that determines which author markemes will become the interval markeme. Its value equals the product of total normalized InTeM and the number of authors who used this markeme. Thus the interval markemes are the first 50 markemes of a general markeme list that is sorted in descending order. The computation of the correlation coefficient in each pair of authors belonging to this chronological interval is based on the value of total normalized index of textual markedness of their mutual markemes. It allows determining the similarity of authors in relation to one another. The largest value of the correlation coefficient determines the power of maximum vector link between the authors and serves as a means to visualize basic vector links and their power for each author. This allows determining the degree of markeme similarity between the authors. Based on the number of maximum power links and their orientation two centres of attraction – large and small – were distinguished and their quantitative and semantic markeme composition was analyzed. The research results in stratification of the authors. Due to this stratification the authors get the following statuses: typical representatives of the chronological interval – nuclear writers, non-typical to the chronological interval – fringe writers, those who obtain intermediate position – medium writers.

Keywords: markeme; index of textual markedness; correlation coefficient; centre of attraction; authors' markemesimilarity visualization; quantitative linguistics; English literature.

References

1. Istorija vseмирnoj literatury: v 9 tomakh [Elektronniy resurs]: Vol. 4 / Berdnikov G.P. idr. – URL: <http://litlife.club/br/?b=204319&p=101>
2. Erofeeva N.E. Zarubezhnaja literatura XVII vek [Elektronniy resurs]: uchebnik dlja pedagogicheskikh vuzov [Textbook for pedagogical universities] – URL: <http://17v-euro-lit.niv.ru/17v-euro-lit/erofeeva-uchebnik/index.htm>
3. Sidorov U. (2002) Matematicheskaya i informatsionnaya podderzhka metodov obrabotki literaturnykh tekstov na osnove formalno-grammaticheskikh parametrov [Mathematical and informational support of methods for processing literary texts on the basis of formal grammatical parameters]. Techn. sciences cand. diss. Petrozavodsk.
4. Forstemann E. (1906) Commentary on the Maya manuscript in the royal public library of Dresden. Published by the Museum. URL: [http://нэб.рф/search/?f_field\[authorbook\]=f/authorbook/Forstemann+E](http://нэб.рф/search/?f_field[authorbook]=f/authorbook/Forstemann+E).
5. Bunakovskiy V. (2017) Osnovaniya matematicheskoi teorii veroyatnosti [Foundations of the mathematical theory of probability]. M.: Kniga po Trebovaniyu. URL: <https://www.bookvoed.ru/files/3515/36/75/0.pdf> (in Russian).
6. Markov A. (1907) Issledovaniya zamechatelnogo sluchaya zavisimyykh ispytaniy [Studies of a remarkable case of dependent trials] // Izv. Imp. akad. nauk. Ser. VI. Vol. 1. 3.

- pp. 60–80. URL: http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=im&paperid=7832&option_lang=rus(in Russian).
7. Markov A. (1913) Primer staticheskogo issledovanija nad tekstom "Evgeniya Onegina", ilustrirujushchij svyaz ispytaniy v tsep [An example of a static study of the text "Evgeniy Onegin" illustrating the connection of tests in a chain] // *Izv. Imp. akad. nauk. Ser. VI. Vol. 1. 3.* pp. 153–162. URL: http://www.mathnet.ru/php/archive.phtml?jrnid=im&option_lang=rus&paperid=6612&wshow=paper
 8. Yule Udny G. (1944) *A Statistical Study of Vocabulary*. Cambridge. IX. 306 p.
 9. Ross A. (1950) *Philological Probability Problems*. *Statist. Soc. Ser. B. Vol. XII.* pp. 19–59.
 10. Verkhozin S. (2013) On quantitative methods in linguistics // *Vestnik Irkutskogo gosudarstvennogo lingvisticheskogo universiteta. Irkutsk. 3 (24).* pp. 145–150.
 11. Estoup J. (1916) *Gamme stenographique*. Paris.
 12. Condon E. (1928) *Statistics of Vocabulary*. *Science. Vol. 67.* 1733.
 13. Boldrini V. (1948). *Le staticheletterarie e ifonemielementar in el la poesia*. Milano.
 14. Zipf G. (1949) *Human Behavior and the Principle of Least Effort*. Cambridge: Addison-Wesley Press. 533 p.
 15. Zipf G. (1965) *The Psycho-Biology of Language an introduction to dynamic philology*. Cambridge: Mass. MIT Press. 336 p.
 16. Golovin B. (1971) *Jazyk i statistika [Language and statistics]*. M.: Prosveshchenie. 192 p.
 17. Shannon C., Weaver W. (1949) *The Mathematical Theory of Communication*. Univ. of Illinois Press. 117 p.
 18. Herdan G. (1956) *Language as Choice and Chance*. Noordhoff : Groningen. 356 p.
 19. Piotrovskij R. (1977) *Matematicheskaya lingvistika: ucheb. posobie [Mathematical linguistics: textbook]*. M.: Vysshaya Shkola. 383 p.
 20. Piotrovskiy R. (1979) *Inzhenernaya lingvistika i teoriya jazyka [Engineering linguistics and the theory of language]*. L.: Nauka. 111 p.
 21. Fucks W. (1955) *Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen*. Köln und Oplade. 110 p.
 22. Marusenko M. (1996) *Attribuciya anonimnykh i psevdononimnykh tekstov metodami prikladnoi lingvistiki [Attribution of anonymous and pseudo-anonymous texts using the methods of applied linguistics]* // *Prikladnoe jazykoznanie*. SPb.
 23. Miles J. (1965) *The continuity of poetic language; the primary language of poetry, 1540's-1940's*. NY: Octagon Books.
 24. Slugina O. (2009) *Lingvokognitivnie osnovy poeticheskoi kvantitativnosti: na materiale anglijskogo jazyka [Lingvocognitive bases of poetic quantitateness: on material of the English language]*. *Philology cand. diss.* Saransk.
 25. Borunov A. (2016) *Parametric stratification of the author's English-language corpus vocabulary (by the material of the American writer Raghu N. Mitras's author's corpus)* // *Philological Sciences. Issues of Theory and Practice*, Tambov, Gramota Publishers. 12(66): in 4 vol. Vol.1. pp. 65–68.
 26. Borunov A. (2017) *Kvantitativnoe issledovanie anglojazychnogo avtorskogo korpusa Regu N. Mityr [Quantitative study of the English-speaking author's corpus by Regu N. Mityr]* // *Litera. 1.* pp. 76–87. URL: http://e-notabene.ru/fil/article_20090.html (in Russian).
 27. Kishinskaya L., Kishinskiy S. (1980) On the outstripping role of poetical speech in the development of language // *Uchenye zapiski Tartuskogo gos. un-ta. Lingvostatistika I kvantitativnie zakonomernosti teksta. Trudy po lingvistike. Vol. VI.* pp. 47–58.
 28. Kretov A. (2010) *Ponyatie markemy: metodika vyjavlenija i praktika ispolzovaniya [The concept of markeme: the methodology of identification and practice of use]* // *Universalii russkoy literatury. Voronezh.* pp. 138–153.
 29. Guselnikova A., Voronina I., Kretov A. (2015) *Vydelenie tematicheskimi markirovannoy leksiki v tekstakh na estestvennykh jazykakh [Highlighting of thematically marked vocabulary in the texts in natural languages]* // *Materialy XV Mezhdunarodnoi nauchno-*

- metodicheskoi konferencii “Informatika: problemy, metodologija, tehnologii”. Voronezh, VGU. Vol. 3. pp. 245–250.
30. Faustov A., Kretov A. (2017) Ponyatie markemy i predvaritelnie itogi markemnogo analiza russkoi literatury [The concept of markeme and preliminary results of the markemic analysis of the Russian literature] // Vestnik Voronezhskogo gos. un-ta. Ser. Lingvistika i mezhkulturnaya kommunikaciya. 4. pp. 16–31.
31. Istoriya zarubezhnoi literatury XVII veka [Elektronniy resurs]: uchebnik / pod obshchimrukovodstvom Z.I. Plavskina – URL: <https://www.libfox.ru/367790-zahariy-plavskin-istoriya-zarubezhnoy-literatury-xvii-veka.html>

Received 29 May 2018