

### Раздел 3 ЯЗЫКОЗНАНИЕ

УДК 821.16:81'373

DOI: 10.22213/2618-9763-2021-2-104-113

Р. А. Верняева, кандидат филологических наук

Ижевский государственный технический университет имени М. Т. Калашникова, г. Ижевск, Россия

#### СТАТИСТИЧЕСКИЕ ДАННЫЕ УПОТРЕБЛЕНИЯ ЛЕКСЕМЫ СЪЧА В РУССКИХ ЛЕТОПИСЯХ (на материале корпуса русских летописей ИАС «Манускрипт»)

*Материалом для исследования в данной работе стал электронный корпус древнейших русских летописей, опубликованный на портале «Манускрипт: Славянское письменное наследие».*

*В статье описывается различное понимание понятия «коллокация» в современной лингвистике и представлен количественно-статистический анализ лексической единицы сЪча в древнейших русских летописях. Инструментом для получения необходимых выборок стал модуль n-грамм информационно-аналитической системы (ИАС) «Манускрипт», позволяющий выявить текстовые сочетания с разным количеством компонентов. Модуль дает возможность провести статистический анализ языковых единиц с помощью мер ассоциации. В рамках данной работы были использованы меры T-score и Log-Likelihood. Первая позволяет определить, насколько неслучайна сила связанности между словами в словосочетании, вторая позволяет сравнивать наблюдаемые и ожидаемые значения для двух наборов данных. В ходе исследования были сопоставлены статистические данные употребления в летописном тексте лексемы сЪча и семантически близкой лексемы брань.*

*Анализ статистических данных подтверждает эффективность применения мер ассоциации при работе с древнерусским текстом.*

*В работе делается вывод о том, что количественно-статистический метод анализа древнерусского текста позволяет получить лингвистически обоснованные результаты.*

**Ключевые слова:** летописи; меры ассоциации; статистические данные; биграмы; электронный текстовый корпус.

#### Введение

*Русские летописи как объект лингвистического исследования*

Летописи – важнейшие исторические источники, самые значительные памятники общественной мысли и культуры Древней Руси. Как правило, в летописи излагалась русская история от ее начала, иногда летописные тексты открывались библейской историей и продолжались античной, византийской и русской. Данные памятники играли важную роль в идеологическом обосновании княжеской власти в Древней Руси и пропаганде единства русских земель.

Материалом для настоящего исследования являются три списка древнейших русских летописей – Лаврентьевский (далее – ЛЛ), Ипатьевский (далее – ИЛ) и Радзивилловский (далее – РЛ), электронные транскрипции которых представлены на портале «Манускрипт: славянское письменное наследие»<sup>1</sup>. В настоящее время электронный корпус русских летописей осна-

щен модулем выборок и запросов, модулем статистики, модулем n-грамм, что позволяет автоматизировать выборку интересующих единиц.

Несмотря на разное время создания списков и оригинальные их особенности, тексты содержат в том числе и разные редакции одного произведения – Повести временных лет (далее – ПВЛ), что позволяет изучить 1) развитие интересующих языковых особенностей в некотором временном интервале, а также 2) языковые особенности конкретного периода. В каждом из задействованных нами в качестве материала памятников в той или иной степени отражается язык переписчика, живое произношение, синтаксис разговорной речи [1].

Изучением лексики, синтаксиса летописей, а также особенностей функционирования тех или иных языковых единиц в данном типе текстов занимались А. А. Шахматов, В. И. Борковский, Е. В. Душечкина, И. П. Еремин, П. А. Лавровский, Д. С. Лихачев, А. С. Львов,

Н. С. Ковалев, Н. П. Некрасов, М. Д. Приселков, О. В. Творогов, Ф. П. Филин и многие другие ученые (например, работы [2–17]).

В настоящем исследовании проведен анализ языковых единиц с целью оценки эффективности некоторых мер ассоциации при работе с древнерусским текстом.

*Количественно-статистические методы  
в корпусной лингвистике*

Полнотекстовый корпус, оснащенный инструментами для лингвистического анализа, в настоящее время позволяет изучать не только функционирование каждой отдельной языковой единицы в тексте, но и анализировать лексическую сочетаемость текстовых прецедентов, что, в свою очередь, дает возможность исследователю зафиксировать и проанализировать, например, функционирование устойчивых сочетаний либо сочетаний, претендующих на статус устойчивых, т. е. *коллокаций*.

В современной лингвистике термин *коллокация* занимает ведущее место. По определению Е. В. Ягуновой и Л. М. Пивоваровой, коллокация – это «сочетание двух или более слов, имеющих тенденцию к совместной встречаемости» [18].

Несомненно, в силу активного изучения коллокаций в современных текстах в настоящее время в научной литературе представлены различные определения для данного языкового явления (например, работы [19–22 и др.]). Стоит отметить, что большая часть их основана на семантико-грамматической связности компонентов словосочетания.

По мнению И. А. Мельчука: «...под сочетаемостью языковой единицы понимается число других элементов, с каждым из которых данный элемент может вступать в определенное отношение» [23].

Коллокации в корпусной лингвистике, по мнению Т. Ю. Павельевой, определяют как «статистически устойчивые словосочетания. При этом такое статистически устойчивое сочетание может быть как фразеологизированным, так и свободным» [24]. Как мы отмечали ранее [25], нельзя однозначно говорить об идентичности между коллокацией и фразеологизмом в силу того, что фразеологизм, прежде всего, идеоматичен, тогда как коллокации (биграмм) являются следствием процесса изменения значений слов в результате переосмысления отдельных компонентов сочетания. Главная особенность коллокации – это «сохранение лексического значения одним из ее компонентов при устойчивости

словосочетания в целом и отсутствие или ослабление значения образности и эмоционально оценочной нагрузки» [26].

В работе [27] представлено, по нашему мнению, наиболее точное определение для коллокации: «коллокации – это слабоидиоматичные фразеологизмы преимущественно со структурой словосочетания, в которых семантически главный компонент (база) употреблен в своем прямом значении, а сочетаемость со вспомогательным компонентом (коллокатором) может быть задана в терминах семантического класса, но выбор конкретного слова всегда предопределен узусом».

*Результаты статистического исследования*

Статистические сведения в рамках данного исследования были получены с помощью инструментов ИАС «Манускрипт», в частности с помощью модуля *n*-грамм, который дает возможность выявить текстовые сочетания с разным количеством компонентов.

Модуль позволяет провести статистический анализ языковых единиц с применением мер ассоциации (*Mutual Information (MI)*, *Pointwise Mutual Information (PMI)*, *T-score*, *Log-Likelihood*, *Dice*, *Inside* и др.). Мы говорим о статистической ассоциации, которая основана прежде всего на синтаксической и лексической связности. Необходимо отметить, что меры ассоциации учитывают частоту совместной встречаемости единиц, а также частоту каждой отдельной единицы в корпусе текстов.

Использование инструментов ИАС «Манускрипт» для подсчета единиц и автоматического анализа их сочетаемости позволяет установить степень спаянности компонентов внутри синтаксической единицы и обуславливает таким образом актуальность данной работы.

Для получения объективных статических данных нами проведена выборка биграмм с применением мер *T-score*, *Log-Likelihood*.

Как отмечает М. В. Хохлова, «значения мер ассоциации можно считать показателями силы синтагматической связи между элементами словосочетаний» [28].

Мера *T-score* позволяет определить, насколько неслучайна сила связанности между словами в словосочетании. Применяя данную меру, исследователь имеет возможность получить наиболее частотные сочетания.

С помощью меры *Log-Likelihood* пользователь может «оценить слово с помощью сравнения с его среднестатистической, ожидаемой частотой» [29].

Для оценки эффективности мер ассоциации мы взяли частотную и актуальную для летописного текста лексему *сѣча*, которая в тексте ЛЛ употребляется 18 раз, в ИЛ – 17х, в РЛ – 12х.

А. С. Львов в работе «Лексика „Повести временных лет“» пишет о том, что слово *сѣча* в летописном тексте не заимствовано из

церковных книг<sup>1</sup>. С ссылкой на глагольные формы *сѣчи*, *исѣчи*, *высѣчи* ученый утверждает, что словом *сѣча* передавалось значение ‘бой’, «надо полагать преимущественно мечами, может быть, еще и топорами» [30].

Результаты меры *T-score*: лексема *сѣча* – первый компонент биграмм (табл. 1).

*Лексема сѣча в качестве первого компонента биграмм*

Таблица 1. Результаты меры *T-score*

| Ранг | n-грамма        | F  | F(w1) | F(w2) | T-score |
|------|-----------------|----|-------|-------|---------|
| 1    | сѣча зла        | 26 | 47    | 180   | 5,09428 |
| 2    | сѣча велика     | 9  | 47    | 177   | 2,99208 |
| 3    | сѣча силна      | 5  | 47    | 12    | 2,23535 |
| 4    | сѣча избывь     | 3  | 47    | 4     | 1,73174 |
| 5    | сѣча крѣпка     | 2  | 47    | 12    | 1,41307 |
| 7    | сѣча люта       | 2  | 47    | 23    | 1,41203 |
| 8    | сѣчи побѣже     | 2  | 23    | 51    | 1,41184 |
| 9    | сѣчу противу    | 2  | 4     | 523   | 1,40999 |
| 31   | сѣчи разлоучиша | 1  | 23    | 1     | 0,99993 |
| 37   | сѣчи деревлян   | 1  | 23    | 2     | 0,99987 |
| 37   | сѣчи женоуци    | 1  | 23    | 2     | 0,99987 |
| 40   | сѣчѣ побѣже     | 1  | 1     | 51    | 0,99985 |
| 43   | сѣчи исѣкоша    | 1  | 23    | 3     | 0,9998  |
| 44   | сѣча побеже     | 1  | 5     | 15    | 0,99979 |
| 47   | сѣчу идоушоу    | 1  | 4     | 21    | 0,99976 |
| 52   | сѣчи бежаша     | 1  | 23    | 6     | 0,99961 |
| 57   | сѣчи исѣкоша    | 1  | 23    | 14    | 0,99908 |
| 62   | сѣчоу противу   | 1  | 1     | 523   | 0,99851 |
| 64   | сѣча имати      | 1  | 23    | 25    | 0,99836 |

Данные табл. 1 демонстрируют, что самым частотными биграммami стали *сѣча зла* (26х), *сѣча велика* (9х), *сѣча силна* (5х). Именно данные сочетания чаще всего приводятся в контекстах словарных статей для лексемы *сѣча*. Например, в СлРЯ XI–XVII веков даны следующие значения для слова *сѣча* и соответствующие контексты:

1. *Истребление холодным оружием, избиение*: Князь же Олександръ не умедли ни мало, с новгородци и с ладожаны приде на ня [на шведов] и побѣди я <...> и ту бысть велика сѣча свѣмь... и множество много ихъ паде (Новг. I лет.).

2. *Рукопашная схватка, кровопролитное сражение холодным оружием; бой, битва*: И покрыша поле льтское множествьмъ вои и съступиша ся въсходящю слнцю, и быс сѣча зла отинудь и съступиша ся тришьды (Сказ. Бор. Глеб).

3. *Война*: И быс<ть> на Литву сѣча велика (Ипат. лет.).

Приведенная выборка биграмм демонстрирует сочетаемость слова *сѣча* преимущественно с глагольными формами и именами прилагательными.

Обращение к контекстам позволяет отметить, что в качестве частотных выступают биграммы, в которых существительное *сѣча* находится в предикативных отношениях с составным именным сказуемым, выраженным краткой формой имени прилагательного со связкой:

*и' сосу'ниш(с)^а о'бои' и' б(с)^ы сеча зла · первое' побегоша · пориаше · потомъ и'зяславъ дв^двичь (РЛ, л. 183); и' пои'доша противу себе · и' б(с)^ы сеча зла · и' мнози па(д)ша · о(т) и'зяслава · и' ст^ослава · и' всеволода (РЛ, л. 97); и' затворишася болгаре · в городе · и' и'зълезоша болгаре на сечу противу ст^ославу · и' бы'(с) ^сеча велика · и' о'долеваху болгаре (ИЛ, л. 27 об.); и' и'злезоша болгаре на сечю · противу' с^тославу · и' бы(с) ^сеча велика · о'даляху' болгаре · (РЛ, л. 37 об.) и т. п.*

Мы не можем утверждать, что данные сочетания являются в период создания памятников устойчивыми словесными комплексами, претендующими на статус фразеологического оборота в современном его понимании, словари

<sup>1</sup> Далее обратим внимание на слово *брань*.

также не фиксируют их как устойчивые, однако полученная выборка свидетельствует о том, что данные биграммы частотны в летописном тексте и могут претендовать на статус статистически устойчивых сочетаний.

*Лексема с Ъча в качестве второго компонента биграммы*

Наиболее частотным сочетанием, в котором слово с Ъча находится в позиции второго компонента, является *быс с Ъча* (31x, 10x) (табл. 2).

Таблица 2. Результаты меры *T-score*: постпозиция лексемы с Ъча

| Ранг | <i>n</i> -грамма | <i>F</i> | <i>F</i> (w1) | <i>F</i> (w2) | <i>T-score</i> |
|------|------------------|----------|---------------|---------------|----------------|
| 1    | бсы с Ъча        | 31       | 1209          | 47            | 5,5386         |
| 2    | быс с Ъча        | 10       | 206           | 47            | 3,15353        |
| 8    | болгаре с Ъчю    | 2        | 48            | 4             | 1,41383        |
| 17   | бывше с ЪчѢ      | 1        | 2             | 1             | 0,99999        |
| 17   | высечеѢ с Ъче    | 1        | 1             | 5             | 0,99999        |
| 17   | генечевичь с Ъче | 1        | 1             | 5             | 0,99999        |
| 37   | сирѢчь с Ъчю     | 1        | 4             | 4             | 0,99995        |
| 38   | обрѢтошася с Ъчи | 1        | 1             | 23            | 0,99993        |
| 42   | моисии с Ъче     | 1        | 7             | 5             | 0,9999         |
| 44   | болгаре с Ъчоу   | 1        | 48            | 1             | 0,99986        |
| 49   | бсѢ с Ъча        | 1        | 4             | 47            | 0,99946        |
| 51   | бывшее с Ъче     | 1        | 52            | 5             | 0,99926        |
| 57   | время с Ъчю      | 1        | 177           | 4             | 0,99798        |
| 60   | бысть с Ъча      | 1        | 17            | 47            | 0,99772        |

Данные табл. 2 и контексты демонстрируют функционирование анализируемого компонента биграммы в сочетании с глагольной связкой:

*бе бо празникъ · прео'браженио' гн(с)^ю · въ днь · егда си б(с)^ы сеча · избывъ же володимерь сего · постави црк^въ · (ИЛ, л. 47); бе бо въ ть днъ преобращенье гн(с)^е · егда си б(с)^ы сеча · избывъ же володимерь сего · постави црк^въ · (ЛЛ, л. 43) и т. д.*

В целом связочные конструкции менее характерны для летописного текста в силу того, что летопись, как известно, представляет собой компилятивный жанр древнерусской литературы, в котором соединяются информативные фрагменты и повествовательные, которые, в свою очередь, допускают употребление элементов живой разговорной речи, но наша выборка демонстрирует довольно активное употребление глагольной связки при существительном с Ъча.

Однако стоит отметить, что преобладающее большинство контекстов с биграммой *быс с Ъча* включают ее сочетание с именной формой, образуя, таким образом, составное именное сказуемое с присвязочным предикативным членом, выраженным именем:

*на нежтине · ниве · и'сступише(м)ся о'бои'мъ · б(с)^ы сеча зла · и' первие' убиша бориса сн^а вячеславля похв(л)^аш(г)^ася ве(л)^ми' (РЛ, л. 116 об.); и покрыша полетьское обои · о(т) множества вои · бе же пятокъ тогда въсходящю сл^нцю · исступишася обои · б(с)^ы сеча зла · якаже не была в руси (ЛЛ, л. 49); и' поидоста проти'ву · и' бывшимъ и(м)^ на месъте на нежтани ниве и совокупившимъ · о'бои'мъ · б(с)^ы сеча зла · и' первое' убиша бориса сн^а вячеславля · (ИЛ, л. 75); о(т) лица василкова · б^у помогшу' побегоша зли'и погани · и б^ (с)ы сеча лю'та · и гнаша е' · за много поприць · (ИЛ, л. 269) и т. п.*

### Результаты меры *Log-Likelihood*

Мера *Log-Likelihood* – мера правдоподобия – «сравнивает наблюдаемые и ожидаемые значения для двух наборов данных» [31].

Результаты выборки подтверждают данные, полученные при применении меры *T-score*: в выборку попали в качестве самых частотных сочетания с Ъча *зла* (26x), с Ъча *велика* (9x), с Ъча *силна* (5x), *быс с Ъча* (31x) (табл. 3, 4).

Таблица 3. Результаты меры *Log-Likelihood*: лексема с Ъча в качестве первого компонента биграммы

| Ранг | <i>n</i> -грамма   | <i>F</i> | <i>F</i> (w1) | <i>F</i> (w2) | <i>LL</i>  |
|------|--------------------|----------|---------------|---------------|------------|
| 22   | с Ъчи разлоучишася | 1        | 23            | 1             | 3417,69741 |
| 25   | с Ъче хи           | 1        | 5             | 6             | 3323,39857 |
| 29   | с Ъчи деревлян     | 1        | 23            | 2             | 3171,69741 |
| 29   | с Ъчи женоуци      | 1        | 23            | 2             | 3171,69741 |
| 31   | с ЪчѢ побѢже       | 1        | 1             | 51            | 3135,07702 |

Окончание табл. 3

| Ранг | <i>n</i> -грамма | <i>F</i> | <i>F</i> ( <i>w</i> 1) | <i>F</i> ( <i>w</i> 2) | <i>LL</i>  |
|------|------------------|----------|------------------------|------------------------|------------|
| 34   | сЪча избывь      | 3        | 47                     | 4                      | 3061,96557 |
| 35   | сЪчи исъсЪкоша   | 1        | 23                     | 3                      | 3027,79664 |
| 37   | сЪче побеже      | 1        | 5                      | 15                     | 2998,20425 |
| 39   | сЪчно идоушоу    | 1        | 4                      | 21                     | 2957,98357 |
| 40   | сЪчи золотая     | 1        | 23                     | 4                      | 2925,69741 |
| 40   | сЪчи лоутшии     | 1        | 23                     | 4                      | 2925,69741 |
| 43   | сЪча силна       | 5        | 47                     | 12                     | 2853,35833 |
| 45   | сЪчи бежаша      | 1        | 23                     | 6                      | 2781,79664 |
| 50   | сЪча крЪпка      | 2        | 47                     | 12                     | 2528,16402 |
| 51   | сЪчи горожане    | 1        | 23                     | 14                     | 2481,0881  |
| 51   | сЪчи исЪкоша     | 1        | 23                     | 14                     | 2481,0881  |
| 53   | сЪча зла         | 26       | 47                     | 180                    | 2477,3771  |
| 58   | сЪчоу противоу   | 1        | 1                      | 523                    | 2308,94954 |
| 59   | сЪча люта        | 2        | 47                     | 23                     | 2297,26855 |
| 61   | сЪчи имати       | 1        | 23                     | 25                     | 2275,30879 |
| 63   | сЪчи побЪже      | 2        | 23                     | 51                     | 2268,28078 |
| 65   | сЪчи бЪжаша      | 1        | 23                     | 30                     | 2210,60232 |
| 66   | сЪчи роуками     | 1        | 23                     | 34                     | 2166,18155 |
| 68   | сЪчно боу        | 1        | 4                      | 204                    | 2151,07702 |
| 69   | сЪчи воЪ         | 1        | 23                     | 36                     | 2145,89586 |
| 70   | сЪча велика      | 9        | 47                     | 177                    | 2106,83538 |
| 71   | сЪчи варягы      | 1        | 23                     | 43                     | 2082,83628 |
| 72   | сЪчно противоу   | 2        | 4                      | 523                    | 2062,94954 |
| 73   | сЪче створи      | 1        | 5                      | 245                    | 2006,88641 |

Таблица 4. Результаты меры *Log-Likelihood*: лексема сЪча в качестве второго компонента биграмм

| Ранг | <i>n</i> -грамма | <i>F</i> | <i>F</i> ( <i>w</i> 1) | <i>F</i> ( <i>w</i> 2) | <i>LL</i>  |
|------|------------------|----------|------------------------|------------------------|------------|
| 6    | бывшЪ сЪчЪ       | 1        | 2                      | 1                      | 4284,49365 |
| 16   | ВысечецЪ сЪче    | 1        | 1                      | 5                      | 3959,29934 |
| 16   | генечевичъ сЪче  | 1        | 1                      | 5                      | 3959,29934 |
| 27   | сирЪчь сЪчно     | 1        | 4                      | 4                      | 3546,49365 |
| 28   | обреЪтошася сЪчи | 1        | 1                      | 23                     | 3417,69741 |
| 28   | отрокмо сЪчи     | 1        | 1                      | 23                     | 3417,69741 |
| 33   | моисии сЪче      | 1        | 7                      | 5                      | 3268,69003 |
| 35   | болгаре сЪчоу    | 1        | 48                     | 1                      | 3156,59288 |
| 36   | отроком сЪчи     | 2        | 5                      | 23                     | 3092,5031  |
| 38   | бывъши сЪчи      | 1        | 3                      | 23                     | 3027,79664 |
| 39   | болгаре сЪчно    | 2        | 48                     | 4                      | 2910,59288 |
| 42   | бсЪ сЪча         | 1        | 4                      | 47                     | 2672,06479 |
| 44   | бывшее сЪче      | 1        | 52                     | 5                      | 2556,99117 |
| 50   | Ярослава сЪче    | 1        | 125                    | 5                      | 2245,71641 |
| 51   | время сЪчно      | 1        | 177                    | 4                      | 2201,46269 |
| 54   | бысть сЪча       | 1        | 17                     | 47                     | 2158,54894 |
| 55   | почаша сЪчи      | 3        | 105                    | 23                     | 2155,89379 |
| 56   | бы сЪча          | 10       | 206                    | 47                     | 2090,37998 |
| 57   | бывши сЪчи       | 1        | 45                     | 23                     | 2066,70155 |
| 61   | бсы сЪча         | 31       | 1209                   | 47                     | 1863,85585 |
| 68   | велика сЪча      | 2        | 177                    | 47                     | 1573,03383 |
| 70   | своеи сЪчи       | 1        | 259                    | 23                     | 1445,56257 |
| 72   | бЪ сЪча          | 1        | 1151                   | 47                     | 662,57145  |

Статистическая мера *Log-Likelihood* является эффективной для выделения как редких, так и высокочастотных сочетаний, поэтому выборка с применением данной меры является достаточ-

но показательной в нашем случае: мы можем подтвердить данные, полученные мерой *T-score*, выявив статистически устойчивые сочетания, а также увидеть редкие сочетания, которые не

имеют ни статистической, ни семантической устойчивости.

По мнению А. С. Львова, в летописях слово *сѢча* становится менее употребительным, вместо него начинает употребляться слово *брань*, что ученый связывает с желанием «писать так же, как и в церковных книгах, в которых *брань* ‘битва, бой’ [32].

Таблица 5. Абсолютное и относительное количество единиц

| Лексема | ЛЛ                    |                          | ЛИ                    |                          | РЛ                    |                          |
|---------|-----------------------|--------------------------|-----------------------|--------------------------|-----------------------|--------------------------|
|         | абсолютное количество | относительное количество | абсолютное количество | относительное количество | абсолютное количество | относительное количество |
| сѢча    | 18                    | 0,01                     | 17                    | 0,009                    | 12                    | 0,01                     |
| брань   | 16                    | 0,01                     | 42                    | 0,02                     | 10                    | 0,01                     |

Данные табл. 5 демонстрируют относительно равное количество употреблений слов *сѢча* и *брань* в летописных списках. Следует отметить, что в ЛЛ и РЛ лексема *сѢча* незначительно преобладает. Исключение составляет ИЛ:

1) Лексема *брань* преобладает в данном списке в сравнении с другими двумя. Данное очевидное количественное расхождение между рукописями может быть связано как со временем или местом создания списков, так и с их объемом (известно, что ИЛ больше по объему).

2) Важным представляется показатель преобладания слова *брань* над словом *сѢча*. Происхождение, состав и формирование текста летописи определяются специалистами по-разному. Лето-

пись имеет мало сходств с другими летописными списками. Можем предположить, что динамика в употреблении анализируемых единиц связана с тем, что ИЛ в отличие от других взятых нами списков включает в свой состав еще Киевскую и Галицко-Волынскую летописи.

Необходимо сделать вывод о том, что ЛЛ и РЛ не демонстрируют предпочтения слова *брань* лексеме *сѢча*, что связано, вероятно, с тем, что летописный текст отражает смешение книжной и живой языковой традиции.

Несмотря на активное употребление существительного *брань* в тексте летописи, нами зафиксировано единичное количество биграмм с этим компонентом (табл. 6, 7).

Таблица 6. Результаты меры *T-score*: лексема *брань* в качестве первого компонента биграммы

| Ранг | <i>n</i> -грамма | <i>F</i> | <i>F</i> ( <i>w</i> 1) | <i>F</i> ( <i>w</i> 2) | <i>T-score</i> |
|------|------------------|----------|------------------------|------------------------|----------------|
| 3    | брань мыслити    | 2        | 180                    | 87                     | 1,38257        |
| 11   | брань предати    | 1        | 180                    | 105                    | 0,94599        |

Таблица 7. Результаты меры *T-score*: лексема *брань* в качестве второго компонента биграммы

| Ранг | <i>n</i> -грамма | <i>F</i> | <i>F</i> ( <i>w</i> 1) | <i>F</i> ( <i>w</i> 2) | <i>T-score</i> |
|------|------------------|----------|------------------------|------------------------|----------------|
| 28   | бѢ брань         | 1        | 6647                   | 180                    | -2,4193        |
| 30   | въ брань         | 1        | 18622                  | 180                    | -8,57938       |

Выборка биграмм демонстрирует сочетание лексемы *брань* преимущественно с глагольной формой. Биграмм, в которых бы лексема *брань* сопровождалась какой-либо качественной характеристикой (как лексема *сѢча*), не зафиксировано.

### Выводы

Проанализированные статистические данные и рассмотренные нами контексты позволили сделать вывод о том, что статистически устойчивыми сочетаниями можно считать сочетания, характеризующиеся разной степенью слитности компонентов с точки зрения семантики и грам-

матики. Использование статистических мер для извлечения коллокаций из древнерусских текстов является эффективным методом при комплексном их анализе.

### Список источников и сокращений

#### Словари

СлРЯ – Словарь русского языка (XI–XVII вв.) // Этимология и история слов русского языка. URL: <http://etymolog.ruslang.ru/index.php?act=xi-xvii>, свободный.

#### Электронный корпус русских летописей

ЛЛ – Лаврентьевская летопись 1377 г. (РНБ, Ф.п.IV.2) // Манускрипт: Славянское письменное насле-

дие. URL: [http://manuscripts.ru/mns/main?p\\_text=32500902](http://manuscripts.ru/mns/main?p_text=32500902), требуется регистрация.

ИЛ – Ипатьевская летопись перв. пол. (сер.?) XV в. (БАН, 16.4.4) // Манускрипт: Славянское письменное наследие. URL: [http://manuscripts.ru/mns/main?p\\_text=32151080](http://manuscripts.ru/mns/main?p_text=32151080), требуется регистрация.

РЛ – Радзивилловская летопись кон. XV в. (БАН, 34.5.30) // Манускрипт: Славянское письменное наследие. URL: [http://manuscripts.ru/mns/main?p\\_text=43296853](http://manuscripts.ru/mns/main?p_text=43296853), требуется регистрация. ПВЛ – Повесть временных лет об. – оборотный лист л. – лицевой лист.

### Библиографические ссылки

1. *Истрин В. М.* Очерк истории древнерусской литературы домосковского периода (XI–XIII вв.). Москва : Академия, 2003. С. 215.

2. *Борковский В. И.* О языке Суздальской летописи по Лаврентьевскому списку // Тр. Комис. по рус. яз. АН СССР, 1931. Т. 1. С. 1–91.

3. *Душечкина Е. В.* Художественная функция чужой речи в русском летописании // Уч. зап. Тарт. ун-та. 1973. Вып. 306. Тр. по рус. и славян. филологии. Т. 21. С. 65–104.

4. *Еремин И. П.* «Повесть временных лет» как памятник литературы // Литература Древней Руси: Этюды и характеристики. Москва ; Ленинград : Изд-во АН СССР, 1966. С. 43–47.

5. *Лавровский П. А.* О языке северных русских летописей. Санкт Петербург, 1852. 163 с.

6. *Лихачев Д. С.* Русские летописи и их культурно-историческое значение. Москва ; Ленинград : Изд-во АН СССР, 1947. 499 с.

7. *Лихачев Д. С.* Поэтика древнерусской литературы. Москва : Наука, 1979. 359 с.

8. *Львов А. С.* Лексика «Повести временных лет». Москва : Наука, 1975. 367 с.

9. *Ковалев Г. Ф.* Собираемые этнонимы в «Повести временных лет» // Сравнительно-исторические исследования русского языка. Воронеж : Изд-во Воронеж. ун-та, 1980, С. 69–74.

10. Некрасов Н. П. Заметки о языке «Повести временных лет» по Лаврентьевскому списку // Изв. Отд. рус. яз. и словесности. 1986. Т. 1. С. 832–927 ; Т. 2, кн. 1. С. 104–174.

11. *Приселков М. Д.* История русского летописания XI–XV вв. Санкт Петербург, 1996. 280 с.

12. *Творогов О. В.* Литература Древней Руси. Москва : Просвещение, 1981. 128 с.

13. *Творогов О. В.* Лексический состав «Повести временных лет» (словоуказатели и частотный словарь) // Лаврентьевская летопись. (Полное собрание русских летописей. Том первый.). 2-е изд. Москва : Языки славянской культуры, 2001. 496 с.

14. *Филин Ф. П.* Лексика русского литературного языка древнекиевской эпохи: (По материалам летописей) // Учен. зап. Ленингр. пед. ин-та. Т. 80, 1949. 288 с.

15. *Шахматов А. А.* Обзорение нескольких летописных сводов XIV–XVI вв. Москва : 1938. 372 с.

16. *Шахматов А. А.* «Повесть временных лет» и ее источники // Тр. Отд. древнерус. лит-ры, 1940. Т. 4. С. 9–140.

17. *Шахматов А. А.* История русского летописания. Т. 1. Кн. 2. Повесть временных лет и древнейшие русские летописные своды. Санкт Петербург : Наука, 2003. 1024 с.

18. *Ягунова Е. В., Пивоварова Л. М.* Природа коллокаций в русском языке. Опыт автоматического извлечения и классификации на материале новостных текстов // Научно-техническая информация. Серия 2. Информационные процессы и системы. № 6. Москва, 2010. URL: [http://webground.su/services.php?param=priroda\\_collac&part=priroda\\_collac.htm](http://webground.su/services.php?param=priroda_collac&part=priroda_collac.htm) (дата обращения: 15.08.2021).

19. *Zakharov V.* Evaluation and combining association measures for collocation extraction // ACM International Conference Proceeding Series. Сер. “Proceedings of the International Conference on Internet and Modern Society, IMS 2017”. 2017.

20. *Хохлова М. В.* Исследование лексико-синтаксической сочетаемости в русском языке с помощью статистических методов (на базе корпусов текстов) : автореф. Санкт-Петербург, 2010. 27 с. URL: <https://clck.ru/RfVmT> (дата обращения: 20.08.2021).

21. *Борисова Е. Г.* Коллокации. Что это такое и как их изучать. Москва : Филология, 1995. 49 с. URL: <https://search.rsl.ru/ru/record/01001721026> (дата обращения: 20.08.2021).

22. *Баранов В. А.* Корреляционный анализ биграмм русских евангельских списков XI–XIV веков // Интеллектуальные системы в производстве. 2018. Т. 16, № 4. С. 65–74. ISSN 1813–7911.

23. *Мельчук И. А.* О терминах «устойчивость» и «идиоматичность» // Вопросы языкознания. 1960. № 4.

24. *Павельева Т. Ю.* Изучение коллокаций на основе лингвистических корпусов текстов // Вестник Тамбовского университета. Серия «Гуманитарные науки». 2016. URL: <https://clck.ru/X4YZm> (дата обращения: 15.08.2021 г). DOI: 10.20310/1810-0201-2016-21-3/4(155/156)-56-61

25. *Веряева Р. А., Зайнуллина С. Р.* Статистически устойчивые глагольно-именные сочетания в русских летописях // Интеллектуальные системы в производстве. 2020. Т. 18. № 4. С. 54–64. DOI: 10.22213/2410-9304-2020-4-54-64

26. *Павельева Т. Ю.* Изучение коллокаций на основе лингвистических корпусов текстов // Вестник Тамбовского университета. Серия «Гуманитарные науки». 2016. URL: <https://clck.ru/X4YZm> (дата обращения: 15.08.2021). DOI: 10.20310/1810-0201-2016-21-3/4(155/156)-56-61

27. *Баранов А. Н., Добровольский Д. О.* Основы фразеологии. Москва : Флинта, 2013. С. 73.

28. *Хохлова М. В.* Исследование лексико-синтаксической сочетаемости в русском языке с помощью статистических методов (на базе корпусов текстов) : автореф. Санкт-Петербург, 2010. 27 с. URL: <https://clck.ru/RfVmT> (дата обращения: 20.08.2021).

29. Баранов В. А. Статистически значимые слова как характеристика средневекового славянского текста (на материале коллекции Апостолов исторического корпуса «Манускрипт») // Гуманитарное образование и наука в техническом вузе : сб. докл. Всерос. науч.-практ. конф. с междунар. участием (Ижевск, 24–27 октября 2017 г.). Ижевск : Изд-во ИЖГТУ имени М. Т. Калашникова, 2017. С. 359–369. 34,6 МБ (PDF). ISBN 978-5-7526-0781-3.

30. Львов А. С. Лексика «Повести временных лет». Москва : Наука, 1975. С. 306.

31. Конотев М. Введение в корпусную лингвистику : учеб. пособие для студ. филол. и лингв. спец. ун-тов. «Animedia», 2014. URL: <https://clck.ru/X4Z6V> (дата обращения: 10.08.2021).

32. Львов А. С. Лексика «Повести временных лет». Москва : Наука, 1975. С. 306–307.

### References

- Istrin V. M. *Oчерk istorii drevnerusskoi literatury domoskovskogo perioda (XI–XIII vv.)* [Essay on the history of Old Russian literature of the pre-Moscow period (XI–XIII centuries)]. Moscow, Academy Publ., 2003, pp. 215. (in Russ.).
- Borkovskii V. I. *O yazyke Suzdal'skoi letopisi po Lavrent'evskomu spisku* [About the language of the Suzdal chronicle according to the Laurentian list]. Proceedings of the Russian Language Commission Academy of Sciences of the USSR, 1931, vol. 1. pp. 1–91. (in Russ.).
- Dushechkina E. V. [The artistic function of someone else's speech in Russian annals]. *Uch. zap. Tart. unta.* 1973. Edition 306. vol. 21, pp. 65–104. (in Russ.).
- Eremin I. P. «Povest' vremennykh let» kak pamyatnik literatury [“The Tale of Bygone Years” as a Literary Monument]. *Literatura Drevnei Rusi: Etyudy i kharakteristiki.* Moscow; Leningrad, 1966, pp. 43–47. (in Russ.).
- Lavrovskii P. A. *O yazyke severnykh russkikh letopisei* [About the language of the northern Russian chronicles]. St. Petersburg, 1852, pp. 163. (in Russ.).
- Likhachev D. S. *Russkie letopisi i ikh kul'turno-istoricheskoe znachenie* [Russian Chronicles and their cultural and historical significance]. Moscow; Leningrad, 1947, pp. 499. (in Russ.).
- Likhachev D. S. *Poetika drevnerusskoi literatury* [Poetics of Old Russian Literature]. Moscow, Science Publ., 1979, pp. 359. (in Russ.).
- L'vov A. S. *Leksika «Povesti vremennykh let»* [Vocabulary “The Tale of Bygone Years”]. Moscow, Science Publ., 1975, pp. 367. (in Russ.).
- Kovalev G. F. *Sobiratel'nye etnonimy v «Povesti vremennykh let»* [Collective ethnonyms in the “Tale of Bygone Years”]. Comparative historical studies of the Russian language. Voronezh, Voronezh University Press, 1980, pp. 69–74. (in Russ.).
- Nekrasov N. P. *Zametki o yazyke «Povesti vremennykh let» po Lavrent'evskomu spisku* [Notes about the language of the “Tale of Bygone Years” according to the Laurentian list]. *Bulletin of the Department of Russian Language and Literature*, 1986, vol. 1. pp. 832–927; vol. 2, book 1, pp. 104–174. (in Russ.).
- Priselkov M. D. *Istoriya russkogo letopisaniya XI–XV vv.* [History of Russian chronicle writing XI–XV centuries]. St. Petersburg, 1996, pp. 280. (in Russ.).
- Tvorogov O. V. *Literatura Drevnei Rusi* [Literature of Ancient Russia]. Moscow, Education Publ., 1981, pp. 128. (in Russ.).
- Tvorogov O. V. *Leksicheskiy sostav «Povesti vremennykh let» (slovoukazateli i chastotnyi slovnik)* [Lexical composition of the “Tale of Bygone Years” (vocabulary and frequency vocabulary)]. *Laurentian Chronicle. (Complete collection of Russian chronicles. Volume one.)*. 2nd ed. Moscow, Languages of Slavic Culture, 2001, pp. 496. (in Russ.).
- Filin F. P. *Leksika russkogo literaturnogo yazyka drevnekievskoi epokhi: (Po materialam letopisei)* [Lexicon of the Russian literary language of the ancient Kiev era: (Based on the materials of the chronicles)]. *Scientific notes of the Leningrad Pedagogical Institute*, vol. 80, 1949, pp. 288. (in Russ.).
- Shakhmatov A. A. *Obozrenie neskol'kikh letopisnykh svodov XIV – XVI vv.* [Review of several chronicles of the XIV–XVI centuries]. Moscow, 1938, pp. 372. (in Russ.).
- Shakhmatov A. A. «Povest' vremennykh let» i ee istochniki [“The Tale of Bygone Years” and its sources]. *Proceedings of the Department of Old Russian Literature*, 1940, vol. 4, pp. 9–140. (in Russ.).
- Shakhmatov A. A. *Istoriya russkogo letopisaniya* [History of Russian annals]. Vol. 1, book 2. *The Tale of Bygone Years and the Most Ancient Russian Chronicle Vaults.* St. Petersburg, Science Publ., 2003, pp. 1024. (in Russ.).
- Yagunova E. V., Pivovarova L. M. *Priroda kollokatsii v russkom yazyke. Opyt avtomaticheskogo izvlecheniya i klassifikatsii na materiale novostnykh tekstov* [The nature of collocations in Russian. Experience in automatic extraction and classification based on news texts]. *Scientific and technical information. Series 2. Information Processes and Systems. No. 6.* Moscow, 2010. Available at: [http://webground.su/services.php?param=priroda\\_collac&part=priroda\\_collac.htm](http://webground.su/services.php?param=priroda_collac&part=priroda_collac.htm) (accessed 15.08.2021). (in Russ.).
- Zakharov V. Evaluation and combining association measures for collocation extraction // *ACM International Conference Proceeding Series. Serp. “Proceedings of the International Conference on Internet and Modern Society, IMS 2017”*, 2017. (in English).
- Khokhlova M. V. *Issledovanie leksiko-sintaksicheskoi sochetaemosti v russkom yazyke s pomoshch'yu statisticheskikh metodov (na baze korpusov tekstov)* [Research of lexical-syntactic collocation in Russian using statistical methods (based on text corpora)]. PhD thesis. St. Petersburg, 2010, 27 p. Available at: <https://clck.ru/RfVmT> (accessed 20.08.2021). (in Russ.).
- Borisova E. G. *Kollokatsii. Chto eto takoe i kak ikh izuchat'* [Collocations. What are they and how to study them]. Moscow, Filologiya Publ., 1995, 49 p. (in Russ.). Available at: <https://search.rsl.ru/ru/record/01001721026> (accessed 20.08.2021 г.). (in Russ.).
- Baranov V. A. [Correlation analysis of the bigrams of the Russian Gospel lists of the XI–XIV centu-

ries]. *Intellektual'nye sistemy v proizvodstve*, 2018, vol. 16, no. 4, pp. 65-74. ISSN 1813-7911 (in Russ.).

23. Mel'chuk I. A. *O terminakh «ustoichivost'» i «idiomatichnost'»* [About the terms “sustainability” and “idiomaticity”]. *Linguistic issues*, 1960, no. 4. (in Russ.).

24. Pavel'eva T. Yu. *Izuchenie kollokatsii na osnove lingvisticheskikh korpusov tekstov* [Studying collocations based on linguistic text corpora]. *Bulletin of the Tambov University. Humanities Series*. 2016. Available at: <https://clck.ru/X4YZm> (accessed 15.08.2021). DOI: 10.20310/1810-0201-2016-21-3/4(155/156)-56-61 (in Russ.).

25. Vernyaeva R. A., Zainullina S. R. *Statisticheski ustoychivye glagol'no-imennye sochetaniya v russkikh letopisyakh* [Statistically stable verbal-nominal combinations in Russian annals]. *Intelligent systems in production*, 2020, vol. 18, no. 4, pp. 54-64. DOI: 10.22213/2410-9304-2020-4-54-64 (in Russ.).

26. Pavel'eva T. Yu. *Izuchenie kollokatsii na osnove lingvisticheskikh korpusov tekstov* [Studying collocations based on linguistic text corpora]. *Bulletin of the Tambov University. Humanities Series*. 2016. Available at: <https://clck.ru/X4YZm> (accessed 15.08.2021). DOI: 10.20310/1810-0201-2016-21-3/4(155/156)-56-61 (in Russ.).

27. Baranov A. N., Dobrovolskii D. O. *Osnovy frazeologii* [Fundamentals of phraseology]. Moscow, Flinta Publ., 2013, pp. 73. (in Russ.).

28. Khokhlova M. V. *Issledovanie leksiko-sintaksicheskoi sochetaemosti v russkom yazyke s pomoshch'yu statisticheskikh metodov (na baze korpusov tekstov)* [Research of lexical-syntactic collocation in Russian using

statistical methods (based on text corpora)]: PhD thesis. St. Petersburg, 2010, 27 p. (in Russ.). Available at: <https://clck.ru/RfVmT> (accessed 20.08.2021 г.). (in Russ.).

29. Baranov V. A. *Statisticheski znachimye slova kak harakteristika srednevekovogo slavjanskogo teksta (na materiale kollekcii Apostolov istoricheskogo korpusa «Manuskript») // Gumanitarnoe obrazovanie i nauka v tehnikom universitete: sbornik dokladov Vserossijskoj nauchno-prakticheskoy konferencii s mezhdunar. uchastiem (Izhevsk, 24–27 oktjabrja 2017 g.)*. [Statistically significant words as a characteristic of the medieval Slavic text (based on the collection of the Apostles of the historical corpus "Manuscript") // Humanitarian education and science in a technical university: a collection of reports of the All-Russian scientific-practical conference from the international participation (Izhevsk, October 24-27, 2017)]. Izhevsk, Izhevsk State Technical University Publ., 2017, pp. 359-369. ISBN 978-5-7526-0781-3.

30. L'vov A. S. *Leksika «Povesti vremennykh let»* [Vocabulary “The Tale of Bygone Years”]. Moscow, Science Publ., 1975, p. 306. (in Russ.).

31. Kopotev M. *Kopotev M. Vvedenie v korpusnuju lingvistiku : uchebnoe posobie dlja studentov filologicheskikh i lingvisticheskikh special'nostej universitetov. «Animedia»* [Kopotev M. Introduction to corpus linguistics: a textbook for students of philological and linguistic specialties of universities. “Animedia”]. Available at: <https://clck.ru/X4Z6V> (accessed 10.08.2021 г.). (in Russ.).

32. L'vov A. S. *Leksika «Povesti vremennykh let»* [Vocabulary “The Tale of Bygone Years”]. Moscow, Science Publ., 1975, pp. 306-307. (in Russ.).

R. A. Vernyaeva, PhD of in Philology  
Kalashnikov Izhevsk State Technical University, Izhevsk, Russia

#### STATISTICAL DATA ON THE USE OF THE LEXEME *сѣча* IN RUSSIAN CHRONICLES (based on the corpus of russian chronicles of the IAS “Manuscript”)

*The material for research in this work was the electronic corpus of the oldest Russian Chronicles, published on the portal “Manuscript: Slavic Written Heritage”.*

*The article describes the different understanding of the concept of “collocation” in modern linguistics and presents a quantitative and statistical analysis of the lexical *сѣча* in the most ancient Russian Chronicles. The tool for obtaining the necessary samples was the n-gram IAS “Manuscript” module, which makes it possible to identify text combinations with a different number of components. The module makes it possible to carry out a statistical analysis of linguistic units using association measures. Within the framework of this work, the T-score and Log-Likelihood measures were used. The first allows you to determine how non-random the strength of the connection between words in a phrase is, the second allows you to compare the observed and expected values for two datasets. In the course of the scientific research, the statistical data on the use of the lexeme *сѣча* and the semantically close lexeme *брань* in the annalistic text were compared.*

*The analysis of statistical data confirms the effectiveness of the application of association measures when working with the Old Russian text.*

*The paper concludes that the quantitative-statistical method of analysis of the Old Russian text allows one to obtain linguistically substantiated results.*

**Keywords:** chronicle; measures of association; statistics; bigrams; electronic text corpus.

Получено: 31.08.2021

**Образец цитирования**

*Верняева Р. А.* Статистические данные употребления лексемы *сѣча* в русских летописях (на материале корпуса русских летописей ИАС) «Манускрипт» // Социально-экономическое управление: теория и практика. 2021. № 3 (46). С. 104–113. DOI: 10.22213/2618-9763-2021-2-104-113.

**For Citation**

Vernyaeva R. A. [Statistical Data on the Use of the Lexeme *сѣча* in Russian Chronicles (Based on the Corpus of Russian Chronicles of the IAS “Manuscript”)]. *Social'no-jekonomicheskoe upravlenie: teorija i praktika*, 2021, no. 3 (46), pp. 104-113 (in Russ.). DOI: 10.22213/2618-9763-2021-2-104-113.