

Виктор Аркадьевич Баранов  
Ижевский государственный технический университет  
имени М. Т. Калашникова, Кафедра лингвистики  
ул. Студенческая, д. 7, г. Ижевск, Россия  
[victor.a.baranov@gmail.com](mailto:victor.a.baranov@gmail.com)

## ИНСТРУМЕНТЫ И МЕТОДИКА ПОИСКА КОЛЛОКАЦИЙ В ИСТОРИЧЕСКОМ КОРПУСЕ «МАНУСКРИПТ»

(на примере глаголических рукописей)

В статье описывается методика поиска коллокаций в историческом корпусе «Манускрипт» (*manuscripts.ru*). Модуль n-грамм корпуса, предоставляющий пользователю большое количество параметров поиска, статистической оценки и вывода многокомпонентных сочетаний позволяет осуществлять эксперименты с целью выявления в подкорпусах семантически и/или грамматически устойчивых комплексов словоформ. Методика подготовки запросов и анализа результатов применена к подкорпусу глаголических письменных памятников. Выявлены многокомпонентные сочетания, имеющие различную степень устойчивости.

### 1. Исторический корпус «Манускрипт»

Исторический корпус «Манускрипт» (портал корпуса: [manuscripts.ru](http://manuscripts.ru)) предназначен для анализа средневековых славянских текстов как с помощью традиционных приемов (поиск, отбор, систематизация материала), так и современных, в первую очередь — количественных и статистических методов. Основными направлениями работы над корпусом являются:

- разработка и создание инструментария для хранения, обра-

ботки, демонстрации и анализа транскрипции текстов,

- создание текстовой базы данных и текстологическая, аналитическая, лингвистическая ее разметка.

В настоящее время корпус предоставляет доступ к более чем 100 транскрипциям рукописей и отрывков, работа с которыми возможна в режиме отдельного электронного издания<sup>1</sup>, тематической коллекции<sup>2</sup> или текстового корпуса<sup>3</sup>. Поиск и просмотр результатов запросов возможен с помощью модулей портала — однотекстовой и многотекстовой запросных форм, модуля параллельных корпусов, модулей статистики и n-грамм (о корпусе см., например, [Baranov 2010, Баранов — Дубовцев 2010, Баранов 2011а, Баранов 2011б, Баранов 2011в, Баранов 2012а, Баранов 2012б, Баранов — Аникина 2012а, Баранов — Аникина 2012б, Баранов — Зуга 2012, Баранов 2015]).

Запросные формы и формы вывода отдельного электронного издания и коллекции рукописей (однотекстовая и многотекстовая запросные формы) позволяют просмотреть тексты или их фрагменты, построить простые и сравнительные индексы и конкордансы. Параметры поиска и демонстрации этих форм обеспечивают выбор диапазона рукописей, ввод маски словоформы, леммы или регулярного выражения, поиск на основе специальной разметки (основной — неосновной текст, модификация символов, порча и правка текста и т. п.), выбор грамматических значений, поиск фрагментов и др.

Модуль статистики (<http://manuscripts.ru/mns/cred.stat>) позволяет:

- получить количественные данные о распределении в документе искомой единицы (символа, словоформы, леммы),
- получить сведения о распределении единиц одновременно в нескольких рукописях,
- получить данные по выровненному по фрагментам параллельному корпусу списков одного произведения (более подробно, см., например, в [Баранов — Дубовцев 2012]).

<sup>1</sup> См., например, электронное издание Киевских глаголических листков: [http://manuscripts.ru/mns/main?p\\_text=61350133](http://manuscripts.ru/mns/main?p_text=61350133).

<sup>2</sup> См., например, электронную коллекцию славянских Евангелий: <http://manuscripts.ru/mns/portal.main?p1=30>.

<sup>3</sup> Гиперссылка на запросную форму исторического корпуса (<http://manuscripts.ru/mns/srch.simple>) доступна с титульной страницы портала и из раздела «Инструменты».

## 2. Модуль n-грамм

Модуль n-грамм корпуса «Манускрипт» ([http://manuscripts.ru/mns/cred\\_ngr.stat](http://manuscripts.ru/mns/cred_ngr.stat)) предназначен для поиска коллокаций и коллигаций<sup>4</sup> в отдельных текстах или коллекциях на основе количественных, статистических, лингвистических и/или структурных параметров.

Как известно, поиск коллокаций в современных текстовых корпусах осуществляется с помощью применения к n-граммам (чаще всего — биграммам) статистических мер, оценивающих близость (ассоциацию) компонентов на основе количества всех биграмм в корпусе, реального и ожидаемого количества анализируемой биграммы и каждого из ее компонентов [Evert].

Отличительной особенностью модуля исторического корпуса является наличие в запросной форме не только основных (указание количества компонентов, маски компонентов, абсолютного, относительного количества или меры ассоциации), но и дополнительных параметров,

а) снимающих вариативность написаний средневековых славянских текстов:

- поиск на основе точного или неточного соответствия,
- использование регулярных выражений,
- применение маски к словоформам или леммам;

б) обеспечивающих учет структурно-лингвистических параметров n-грамм:

- расстояния между компонентами,
- морфологического значения компонентов,
- межсловных (пунктуационных) знаков,
- служебных слов (союзов, частиц, предлогов) или местоимений;

в) ограничивающих выборку n-граммами с определенными характеристиками:

- с контактным и/или дистантным расположением компонентов,
- с фиксированным или свободным порядком следования компонентов,
- с ограниченной или неограниченной сочетаемостью компонентов и нек. др.

<sup>4</sup> Коллокация — грамматически и/или семантически устойчивое сочетание слов, компоненты которого (коллокаты) имеют тенденцию к устойчивой (неслучайной) совместной встречаемости. Устойчивость проявляется в регулярном совместном использовании коллокатов, в их ограниченной сочетаемости. Коллигация — конструкция из n-компонентов, характеризующаяся некоторым набором морфологических характеристик словоформ и их грамматических связей.

Все указанные дополнительные параметры могут быть использованы как каждый в отдельности, так и в сочетании друг с другом с одной целью — для поиска устойчивых семантических и/или грамматических сочетаний в больших объемах текстов (более подробно, см., например, в [Баранов 2014, Баранов 2016а, Баранов 2016б]).

### 3. Коллекция глаголических текстов

Коллекция глаголических рукописей корпуса «Манускрипт», созданная совместно коллективом под руководством профессора Хайнца Микласа (Венский университет) и командой корпуса, включает транскрипции 7 документов — 3 кодексов и 4 отрывков (см. список источников).

Доступ к текстам коллекции возможен через страницу Глаголических рукописей (<http://manuscripts.ru/mns/portal.main?p1=40>), страницу Синайских глаголических рукописей (<http://manuscripts.ru/mns/portal.main?p1=55>), а также через модуль статистики и модуль n-грамм.

Объем коллекции — 88 630 словоформ (ЕвАс — 43 287, ПсДС — 30 659, ЛитС — 12 474, КЛ — 1228, МинС — 485, МедЛ — 279, Час — 218).

### 4. Эксперименты по поиску коллокаций

Целью экспериментов, осуществленных с помощью модуля n-грамм, является поиск в коллекции глаголических текстов устойчивых в семантическом или грамматическом отношении сочетаний, а также поиск сочетаний параметров модуля, обеспечивающих выявление таких сочетаний.

Эксперимент 1<sup>5</sup>. Многокомпонентные n-граммы

Поиск семиграмм (всего — 8714) позволил найти повторяющиеся 3 или 2 раза контексты:

- **ДА НЕ ОСТАНАЖТЪ НА КРСТѢ ТѢЛЕСА ВЪ** 3× ЕвАс
- **ѢКО БЛАГЪ ѢКО ВЪ ВѢКЪ МИЛОСТЪ ЕГО** 3× ПсДС
- **БРАША СА СО МЪНОЮ ОТЪ ЮНОСТИ МОЕѦ** 2× ПсДС
- **ДА ПРѢБИЖТЪ ГОЛѢНН НХЪ Н ВЪЗЪМЖТЪ Ѧ** 2× ЕвАс и нек. другие (всего — 19).

Шести- (всего — 14 409) и пятиграммы (всего — 23 511) являются или повторяющимися частями семиграмм, или повторяющимися 3

<sup>5</sup> Эксперименты могут быть воспроизведены с помощью модуля (URL: [http://manuscripts.ru/mns/cred\\_ngr.stat](http://manuscripts.ru/mns/cred_ngr.stat)) после регистрации в корпусе (URL: [http://manuscripts.ru/mns/portal.main?p1=18&p\\_lid=1](http://manuscripts.ru/mns/portal.main?p1=18&p_lid=1)); авторизация не требуется.

или 2 раза более короткими контекстами:

шестиграммы (всего повторяющихся — 76):

- **ДА ПОСТЫДАТЬ СЯ И ПОСРАМАЛЮЮТЪ СЯ** 3× ПсДС
- **ПРИЗЪРИ НА МА И ПОМІЛѸИ МА** 3× ПсДС и некоторые другие (всего — 8)
- **А ІЖЕ ПОГОУБИТЬ ДШЖ СВОЖ МЕНЕ** 2× ЕвАс
- **АЗЪ ЖЕ РЕХЪ ВЪ ѸЖАСѢ МОЕМЪ** 2× ПсДС и некоторые другие (всего — 68),

пятиграммы (всего повторяющихся — 219):

- **ЗАКОНЪ ТВОИ ПОУЧЕНИЕ МОЕ ЕСТЬ** 3× ПсДС
- **ВН СЯ НСЪ ОУЧЕННИКОМЪ СВОИМЪ** 3× ЕвАс и некоторые другие (всего — 18)
- **А ІЖЕ ПОГОУБИТЬ ДШЖ СВОЖ** 2× ЕвАс
- **АЗЪ ЖЕ РЕХЪ ВЪ ѸЖАСѢ** 2× ПсДС
- **Н ВНАѢ БЪ ВКО ДОБРО** 2× ЛитС и многие другие (всего — 201).

Четырехграммы (всего — 36 272, повторяющихся — 662) представляют собой также текстовые формулы:

- **РЕЧЕ ГЪ СВОИМЪ ОУЧЕННИКОМЪ** 6× ЕвАс и другие (всего — 4)
- **ІЖЕ ЕСТЬ НА НЕСѢХЪ** 5× ЕвАс и другие (всего — 3)
- **БЪ МОИ ЕСИ ТЫ** 4× ПсДС и другие (всего — 22)
- **ВЪ ДЕНЪ ПЕЧАЛИ МОЕЯ** 3× ПсДС и другие (всего — 79)
- **НАДЪЮЩЕИ СЯ НА НА** 2× ЛитС и многие другие (всего — 554).

Примеры показывают, что повторение сочетаний является результатом повторения в списках одних и тех же чтений и формул, что связано в первую очередь с составом и структурой рукописей.

В выборках не встретилось ни одного сочетания из 7–4 компонентов-словоформ, которые бы использовались в двух или более текстах. Можно предположить, что возможной причиной этого являются, в частности, различные графико-орфографические особенности рукописей. Чтобы проверить это предположение, был проведен второй эксперимент.

Эксперимент 2. Наиболее частотные n-граммы лемм

Количество повторяющихся шести-, пяти- и четырехграмм, построенных из лемм, больше количества повторяющихся n-грамм из словоформ:

n-граммы	Словоформы	Леммы
7-граммы	21	20
6-граммы	76	95
5-граммы	219	295
4-граммы	662	897

Ожидаемое увеличение количества повторяющихся n-грамм в целом увеличивает и количество повторений. Так, например, в четырехграммах две встретились по 15 раз, 2 — по 13 раз, 1 — 12 раз, 2 — по 11 раз, 1 — 9 раз, 2 — по 8 раз, 5 — по 7 раз, 2 — по 6 раз, 11 — по 5 раз, 44 — по 4 раза, 105 — по 3 раза, 725 — по 2 раза.

По содержанию это также повторяющиеся контексты, речевые формулы и их части. Одновременно есть и существенное отличие от n-грамм, построенных на основе словоформ: один и тот же контекст может повторяться в двух кодексах. В пятиграммах встречается только один контекст, повторяющийся в двух рукописях:

- **НА НЕБО Н НА ЗЕМЛЯ** 2×: ПсДС: **НА НБ̄СИ НА ЗЕМІ** 116.1.1–20<sup>6</sup>, ЕвАс: **НА НБ̄СИ Н НА ЗЕМІ** 73.2.2–3;

в четырехграммах таких случаев уже 29 (см. приложение 1). Такие контексты встречаются в ЛитС и в ПсДС, в ПсДС и в ЕвАс, в ЛитС и в ЕвАс.

Именно они представляют интерес для дальнейшего анализа с точки зрения интертекстуальных отношений, наличия заимствований, формул и коллокаций.

Для поиска семантически и/или грамматически устойчивых сочетаний из меньшего, чем четыре, количества компонентов необходимо учитывать не только абсолютное количество таких сочетаний в подкорпусе текстов, но и другие параметры — количество всех словоформ и n-грамм в подкорпусе, частоту встречаемости каждого из компонентов, а также ожидаемые частоты конкретной n-граммы и ее компонентов.

Степень близости словоформ в сочетании из n компонентов (чаще всего — из двух) выявляется с помощью так называемых мер устойчивости (мер ассоциации) (см., например, [Evert]). В модуле n-грамм корпуса «Манускрипт» реализовано несколько таких мер: MI (Mutual Information score), PMI (Pointwise mutual Information score), T-score, LL (Log-Likelihood score), LLD (Log-Likelihood<sub>Dunning</sub>), Dice (Dice coefficient), Dice' (Dice' coefficient),  $\chi^2$  (Chi-squared test).

<sup>6</sup> Где 116 — номер листа, 1 — номер страницы, 1 — номер столбца, 20 — номер строки.

### Эксперимент 3. Мера MI

Мера взаимной информации (MI)<sup>7</sup> позволяет выявлять в корпусах единично или редко встречающиеся сочетания, к которым относятся, например, имена собственные, сложные термины, но также и случайные сочетания [Кочеткова, Ягунова—Пивоварова 2013]. Особенностью меры является также то, что она может применяться для оценки а) коллокаций, компоненты которых находятся на расстоянии, и б) коллокаций, имеющих более чем два компонента [Там же].

Запросы, параметрами которых являлись: компоненты — словоформы, количество компонентов — 3, расстояние между компонентами — 0 или 1<sup>8</sup>, учет границ конструкций — «нет» или «да», — дали похожие выборки, в которых в начале перечней находилось большое количество случайных сочетаний, состоящих из словоформ с единичными сокращениями (**а̃т̃ста́нт̃ма** ЕвАс 145.2.1–24), содержащими неп прочитанные символы (**ав̃лю̃бра̃т̃**<sup>у</sup>[...] **м̃в** ЛитС 43.2.1–13) и под.

Иная ситуация при поиске сочетаний на основе лемм. При тех же прочих параметрах в верхней части таблицы с максимальными величинами меры MI находятся предикативные или полупредикативные сочетания [**у̃лов̃к̃з̃ е̃ст̃з̃**]<sup>9</sup> **ва̃ж̃ж̃н̃** [**с̃в̃ьр̃з̃**] **ме̃д̃в̃ьд̃з̃** [**н̃**] **ме̃д̃в̃ьд̃н̃ц̃з̃** ПсДС 1.1.1–10, [**ѣ̃ко**] **а̃сп̃и̃д̃ы̃** **г̃л̃ьх̃ы̃** **з̃а̃т̃з̃к̃а̃ю̃щ̃е̃** [**ѡ̃ш̃и̃** **с̃во̃и̃**] ПсДС 67.2.1–17, [**ѡ̃** **ѣ̃ко** **з̃е̃л̃и̃е̃**] **з̃л̃а̃к̃а** **с̃к̃о̃р̃о̃** **о̃т̃з̃п̃а̃д̃е̃т̃з̃** и некоторые другие конструкции (см. приложение 2).

### Эксперимент 4. Мера T-score

Мера T-score<sup>10</sup> считается одной из лучших для выявления коллокаций [Evert 4.3], в современных текстах позволяет обнаружить наиболее часто используемые предложно-падежные, глагольно-предложные конструкции, сложные предлоги, союзы и обороты [Ягунова—Пивоварова 2013].

Запрос, параметрами которого были: компоненты — словофор-

<sup>7</sup>  $MI = \log_2 \frac{F(w_1, w_2) \times N}{F(w_2) \times F(w_1)}$ , где  $w_1$  — первый компонент,  $w_2$  — второй компонент,  $F(w_1)$ ,  $F(w_2)$  — абсолютные (независимые) частоты  $w_1$  и  $w_2$ ,  $F(w_1, w_2)$  — абсолютная частота n-граммы в подкорпусе,  $N$  — общее число словоформ в корпусе.

<sup>8</sup> Где 0 — контактное расположение компонентов, 1 — дистантное расположение компонентов с одной словоформой между ними.

<sup>9</sup> В квадратных скобках — словоформы, игнорированные модулем при построении триграмм из-за значения 1 параметра «Расстояние».

<sup>10</sup>  $T - score = \frac{F(w_1, w_2) - \frac{F(w_1) \times F(w_2)}{N}}{\sqrt{F(w_1, w_2)}}$  (значения переменных см. в примечании 6).

мы, количество компонентов — 2, расстояние между компонентами — 0, учет границ конструкций — «да», исключить служебные слова — «нет», — позволил выявить двукомпонентные частотные формульные конструкции: предложно-падежные, предикативные сочетания, конструкции глаголов с зависимыми местоимениями, но также и наиболее частотные биграммы знаменательных и служебных слов (первые 20 см. в приложении 3).

Особенностью этого списка является то, что биграммы, характеризующиеся наиболее высоким значением меры, встречаются в нескольких рукописях (далее, за пределами 20 начальных позиций также: **раба твоего** ЛитС, ПсДС, ЕвАС, **на та** ПсДС, ЕвАС, **азъ же** ПсДС, ЕвАС, **и не** Час, ПсДС, **на земли** ЛитС, ПсДС, ЕвАС, **въ домъ** ПсДС, ЕвАС, **на нь** ЛитС, ПсДС, ЕвАС и мн. др.). Анализ результата подтверждает вывод, сделанный на основании современных текстов: эта мера позволяет выявить биграммы, которые характеризуют подкорпус в целом [Ягунова—Пивоварова 2013]. Это справедливо и по отношению к историческому текстовому материалу.

Наличие встречающихся в нескольких списках биграмм тем более показательно, если учесть, что наборы биграмм, имеющих наибольшее значение меры T-score, отдельно в каждой из рукописей практически не совпадают (см. приложение 4), а перечень наиболее частых словоформ совпадает лишь частично, их порядок следования — уникален (см. приложение 5).

Очень показательной является выборка на основе лемм, в которой в подавляющем большинстве присутствуют семантически и/или грамматически значимые, встречающиеся не в одном, а сразу в нескольких кодексах биграммы: атрибутивно-именные, предложно-именные, предикативные конструкции, сочетания глагола с прямым дополнением (местоимением), конструкции с постпозитивным **же** (см. приложение 6).

#### Эксперимент 5. Мера Dice

Мера Dice<sup>11</sup> позволяет найти в текстах симметричные (компоненты употребляются только вместе) или приближающиеся к ним по этому показателю сочетания [Пентус—Пиперски—Сорокин].

Запрос с параметрами: компоненты — словоформы, количество компонентов — 2, расстояние между компонентами — 0, учет границ конструкций — «да», исключить служебные слова — «нет», — возвращает выборку, в которой начало списка состоит из единичных уни-

<sup>11</sup>  $Dice = \frac{2 * F(w_1, w_2)}{F(w_1) + F(w_2)}$  (значения переменных см. в примечании 6).

кальных сочетаний, компоненты которых имеют нестандартную графико-орфографическую форму. Но при этом повторение некоторого сочетания словоформ с высоким показателем симметричности 2 раза и более может свидетельствовать о существенной семантической и/или грамматической связи его компонентов, см., например, **ДѢХА СѢТА** (позиция 5 приложения 7), **НИЩЕТОЖЪ БОГАТНЫМЪ** (позиция 12), **ПЕСЬНАНА МЪХЪ** (позиция 13), **ПОЛІ ТАНЫСЪ** (позиция 14), **ЪГОТОВАТИ ТРЕПЕЗЪ** (позиция 19), **ѠБА ПРНЪСТЕ** (позиция 20). Остальные 14 биграмм этого списка — повторение в рукописях идентичных контекстов.

Эксперимент 6. Мера Dice, примененная к свободным сочетаниям

Запросная форма модуля позволяет искать сочетания, разница между которыми только в порядке следования компонентов. Такие сочетания интерпретируются как идентичные, к такому объединенному сочетанию может быть применена любая статистическая мера.

Запрос с параметрами: компоненты — словоформы, количество компонентов — 2, расстояние между компонентами — 0, учет границ конструкций — «да», исключить служебные слова — «нет», закрепленность — «свободные», следование — «нет» — позволяет найти в коллекции биграмм, компоненты которых встречаются друг с другом чаще, чем с другими компонентами, и порядок следования которых свободный. Это глагольные сочетания с зависимыми существительными или местоимениями, именные с согласованным определением, предикативные конструкции, - все, имеющие высокий показатель семантической и грамматической связи друг с другом (см. приложение 8).

Снятие графических вариантов и игнорирование морфологической формы слов с помощью построения биграмм на основе лемм дает возможность найти сочетания, в которых компоненты семантически и грамматически устойчивых комплексов имеют большую степень связанности: **ПЪРЪВЪ ПОСЛѢДНЪ** в **Мнози же бждтъ пръвин послѣднн · н послѣднн/ пръвнн**, наименование географического объекта **УРМЪНОЕ МОРЕ**, предикативные конструкции **КОУРЪ ВЪЗГЛАНТН**, **РЕЦН ГОСПОДЪ**, глагольные сочетания **ДВЪРЬ ЗАТВОРНТН**, **ПРНЕСТН ДАРЪ**, **БОЛЪТН ЦѢЛНТН**, **ХВАЛА ВЪЗДАТН**, **ЗАВѢТЬ ЗАВѢЩАТН**, именные конструкции с несогласованными и согласованными определениями **НАРОСТЬ ГНЪВЪ**, **КЪТЕРЪ УЛОВЪКЪ**, **БОГЪ НАШЪ**, **ДОУША МОН**, предложно-падежные конструкции **НА ЗЕМЛѢ**, **НА НЕБО**, а также буквенная

цифр и конструкции с **ж** (см. приложение 9).

В последней выборке на основе лемм биграмм с высоким значением меры Dice являются коллокатами с семантически и грамматически связанными компонентами, а их совокупность является одновременно и тематической доминантой глаголической коллекции и может быть использована в качестве ее статистической лексико-тематической характеристики.

Эксперимент 7. Мера Dice (леммы, свободные сочетания) как характеристика текстов

Сравнение биграмм, имеющих высокие значения меры Dice<sup>12</sup>, в разных текстах показывает, что перечень начальных 20-ти совершенно различен (см. приложение 10). Кроме того, различны их максимальные величины меры Dice. Если в АсЕв и ПсДС диапазон значений этих 20-ти биграмм соотносим (0.26–0.08 и 0.31–0.05 соответственно), то аналогичный диапазон ЛитС, несмотря на высокое значение 0.44 первой биграмм, в целом значительно ниже — 0.15–0.01. Из чего следует, что характеристикой текстов (и подкорпусов) является не только определенный перечень биграмм, характеризующихся максимальной мерой ассоциации компонентов, но и сама величина этой меры (и их диапазон и/или частотность) для каждого из списков.

## 5. Заключение

Эксперименты, осуществленные с помощью модуля n-грамм корпуса «Манускрипт», позволили установить, что наиболее чувствительными для результатов являются следующие параметры:

- количество компонентов n-граммы,
- лингвистическая единица коллоката — словоформа или лемма,
- закрепленность или свободное следование компонентов.

В меньшей степени влияли на результат:

- изменение расстояния между коллокатами,
- учет или неучет знаков-разделителей.

Влияние на результат включения в n-граммы служебных слов или

<sup>12</sup> Параметры запроса: компоненты — словоформы, количество компонентов — 2, расстояние между компонентами — 0, учет границ конструкций — «да», исключить служебные слова — «нет», закрепленность — «свободные», следование — «нет».

игнорирования их зависело от применяемой меры: статистические меры в меньшей степени, чем количественные (абсолютное или относительное количество), чувствительны к наличию или отсутствию в n-граммах служебных слов.

Большое влияние на результаты оказывает графико-орфографическая вариативность написаний в кодексах. Она препятствует автоматическому выявлению семантически и/или грамматически идентичных словоформ<sup>13</sup>.

Это ограничение было снято в экспериментах использованием для построения n-грамм не текстовых прецедентов, а их лемм, полученных с помощью автоматической лемматизации текстов рукописей. И этот способ устранения графико-орфографической вариативности и морфологической изменчивости дал существенно иные результаты. В то же время следует учесть то, что вследствие неполной автоматической лемматизации (часть словоформ автоматически не распознается лемматизатором вследствие нестандартных сокращений и отсутствия в базе данных лемматизатора некоторых лемм) полученные в экспериментах результаты не могут считаться окончательными.

Выборка и ранжирование многокомпонентных n-грамм позволяет выявить повторяющиеся в одном или нескольких кодексах контексты, речевые формулы или их части. Оценка биграмм с помощью мер MI, T-score и Dice позволила установить те предикативные, атрибутивно-именные, предложно-именные конструкции, сочетания глагола с прямым дополнением (местоимением), некоторые сложные онимы, которые имеют тесные семантические и/или грамматические связи компонентов.

Последний эксперимент позволил установить, что характеристикой кодексов (в нашем случае — АсЕв, ПсДС и ЛитС) является индивидуальный перечень биграмм каждого с наиболее высокими показателями связи компонентов, а также значения меры Dice этих перечней.

Лексико-семантическая, синтаксическая, фразеологическая, жанрово-тематическая оценка выявленных коллокаций и сочетаний, близких к ним, — предмет специальных работ.

<sup>13</sup> Использующиеся в других модулях корпуса процедуры устранения вариативности с помощью транслитерации, применения правил приравнивания букв или их сочетаний друг к другу в настоящее время в модуль статистики еще не встроены.

## Приложение 1.

### Четырехграммы, повторяющиеся в нескольких текстах<sup>14</sup>

**въ родъ н родъ** 12×: ЛитС: **въ родъ ѿ родъ** 52.2.1-13, ПсДС: **въ родъ і родъ** 42.2.1-12, **въ родъ і родъ** 60.1.1-19, **Въ родъ і родъ** 89.2.1-1, **въ родъ і родъ** 95.1.1-23, **въ родъ і родъ** 95.2.1-6, **въ родъ і родъ** 104.2.1-10, **Въ родъ і родъ** 105.1.1-10, **Въ родъ і родъ** 110.2.1-9, **въ родъ і родъ** 122.2.1-22, **въ родъ і родъ** 131.2.1-14, **въ родъ і родъ** 142.1.1-11

**н къ томоу не** 6×: ПсДС: **і къ томѡ не** 51.1.1-15, ЕвАс: **н къ томоу не** 15.2.1-18, **н къ томоу не** 78.1.1-3, **н къ томоу не** 96.2.2-5, **н къ томоу не** 97.1.1-13, **н къ томоу не** 156.2.1-29

**въ домъ отъць мон** 4×: ПсДС: **въ домѡ бѣа моего** 145.1.1-14, ЕвАс: **Въ домоу ѡца моего** 29.1.1-12, **въ дом бѣа моего** 55.2.1-14, **Въ домоу ѡца моего** 92.2.2-28

**възнеси са на небо** 4×: ПсДС: **Възнеси са на небеса** 67.1.1-14, **Възнеси са на небеса** 67.2.1-3, **Възнеси са на небеса** 113.1.1-4, ЕвАс: **вънесе са на небо** 154.2.1-4

**н възврати са въ** 4×: ПсДС: **его възврати са во** 141.2.1-10, ЕвАс: **н възвратиша са въ** 28.2.2-20, **н възвратише са въ** 54.2.2-23, **н възврати са въ** 145.1.1-3

**отъць н сынъ н** 4×: ЛитС: **бѣа н сѣна н** 48.1.1-5, **бѣа ѿ сѣна н** 56.1.1-9, **бѣа н сѣна н** 56.2.1-10, ЕвАс: **ѡца н сѣна н** 112.2.1-19

**речи господь господь мон** 4×: ПсДС: **рѣхъ гѣн гѣ мон** 18.2.1-11, **рече гѣ гю моему** 114.2.1-20.1, ЕвАс: **рече гѣ гю моему** 49.1.2-26, **рече гѣ гю моему** 116.1.1-21

**благословенъ господь богъ нзранлевъ** 3×: ПсДС: **блнъ гѣ бѣ ілевъ** 80.2.1-7, **блнъ гѣ бѣ ілевъ** 111.1.1-16, ЕвАс: **блнъ гѣ бѣ нзлвъ** 149.2.1-11

**велни по всь земля** 3×: ПсДС: **велеі по всей земли** 58.2.1-6, **велеі по всей земли** 101.1.1-1, ЕвАс: **велеі по всей земн** 147.1.1-11

**въ днь въ нже** 3×: ПсДС: **Въ дн въ наже** 21.2.1-3, ЕвАс: **въ днн въ наже** 145.1.1-9, **въ днн въ наже** 149.1.1-5

**господь сила съ мзи** 3×: ЛитС: **гѣ силъ съ нами** 52.2.1-12, ПсДС: **гѣ силъ съ нами** 58.1.1-13, **гѣ силъ съ нами** 58.1.1-24

**небо н на земля** 3×: ПсДС: **нбси на земли** 116.1.1-20, **небеси на земли** 131.1.1-19, ЕвАс: **неси н на земли** 73.2.2-3

**аще не обрати са** 2×: ПсДС: **Аще не обрати са** 9.1.1-17, ЕвАс: **Аште не обрати са** 42.2.2-25

<sup>14</sup> Компоненты — словоформы, количество компонентов — 4, расстояние между компонентами — 0.

**БЪИТИ НА ГЛАВА** н 2х: ПсДС: **ЕСИ НА ГЛАВЪ ЕГО** 28.2.1-4, ЕвАС: **БЪ НА ГЛАВЪ ЕГО** 155.2.1-22

**ВЪ ВЪСЪ ПРЕДЪЛЪ** н 2х: ПсДС: **БО ВЪСА ПРЕДЪЛЪ ИХЪ** 108.2.1-17, ЕвАС: **ВЪ ВЪСЪХЪ ПРЕДЪЛЪХЪ ЕНА** 135.1.1-5

**ВЪ НМА ОТЪЦЪ** н 2х: ПсДС: **ВЪ НМА ОТЦА** н 1.1.1-2, ЕвАС: **ВЪ НМА ОЦА** н 112.2.1-18

**НА НЕБО** н **НА** 2х: ПсДС: **НА НБСИ** **НА** 116.1.1-20, ЕвАС: **НА НБСИ Н НА** 73.2.2-3

**ДА НЕ ВЪЗВРАТИТИ СЯ** 2х: ПсДС: **ДА НЕ ВЪЗВРАТИТЬ СЯ** 83.1.1-1, ЕвАС: **ДА НЕ ВЪЗВРАТИТЬ СЯ** 83.2.1-15

**Н ВЕСЕЛИТИ СЯ ВЪ** 2х: МинС: **Н ВЕСЕЛА СЯ ВЪ** 1.2.1-2, ПсДС: **Н ВЪСЕЛИШИ СЯ ВЪ** 6.2.1-17

**Н ВЪЗВРАТИТИ СЯ ВЪСПАТЬ** 2х: ПсДС: **И ВЪЗВРАТАТЬ СЯ ВЪСПАТЬ** 129.1.1-23, ЕвАС: **НЕ ВЪЗВРАТИТЬ СЯ ВЪСПАТЬ** 83.2.1-15

**Н ВЪСЪ НЖЕ БЪИТИ** 2х: ПсДС: **И ВЪСЪ ВЪЖЕ СЖТЬ** 141.2.1-16, ЕвАС: **Н ВЪСА ИЖЕ БВАХЪ** 50.2.2-14

**Н ДАТИ Н БЪИТИ** 2х: ПсДС: **ЕГО ДАЛЪ ЕМОЪ ЕСИ** 28.1.1-20, ЕвАС: **Н ДАСТЪ Н СЖШТИНМЪ** 53.1.1-22

**Н ДЕНЬ Н НОЩЬ** 2х: ПсДС: **НХЪ ДЕНЪ Н НОЩЪ** 2.1.1-15, ЕвАС: **НЕМОУ ДЕНЬ Н НОШТЬ** 67.1.1-22

**Н ПОКЛОНИТИ СЯ** н 2х: ПсДС: **И ПОКЛОНАТЬ СЯ ЕМОЪ** 80.1.1-10, ЕвАС: **Н ПОКЛОНИ СЯ ЕМОУ** 26.1.1-12

**НЪИНЪ Н ПРНСНО** н 2х: ЛитС: **НЪИНЪ Н ПРНСНО Н** 51.2.1-13, ПсДС: **НЪИНЪ Н ПРНСНО** н 2.1.1-12

**ОТЪ ВЪСТОКЪ Н ЗАПАДЪ** 2х: ПсДС: **ОТЪ ВЪСТОКЪ И ЗАПАДЪ** 111.1.1-23, ЕвАС: **УТЪ ВЪСТОКЪ Н ЗАПАДЪ** 63.1.1-18

**ПРЕДАТИ БЪИТИ ВЪ РЖКА** 2х: ПсДС: **ПРЕДАНЪ БЖДЕТЬ ВЪ РЖЦЪ** 14.2.1-12, ЕвАС: **ПРЕДАНЪ БЖДЕТЬ ВЪ РЖЦЪ** 78.1.2-7

**ПРНСКЪРЪБЪНЪ БЪИТИ ДОУША МОН** 2х: ПсДС: **ПРНСКЪРЪБЪНА ЕСИ ДШЕ МОЪ** 54.2.1-16, ЕвАС: **ПРНСКЪРЪБЪНА ЕСТЬ ДША МОЪ** 89.2.2-23

**СВОИ ЗАПОВЪДЪТИ О ТЪ** 2х: ПсДС: **СВОИМЪ ЗАПОВЪСТЪ О ТЕБЪ** 98.2.1-20, ЕвАС: **СВОИМЪ ЗАПОВЪСТЪ О ТЕБЪ** 140.2.1-4

Приложение 2.

Результат применения меры MI<sup>15</sup> к леммам<sup>16</sup>

№	N-грамма	F	MI	Адрес
1	ВАЗАТИ МѢДВѢДЬ МѢДВѢДНЦА	1	14.85	ПсДС: ВАЖЖИ [СВѢРЪ] МѢДВѢДЪ [Н] МѢДВѢДНЦЖ 1.1.1-10
2	АСПИДА ГЛОУХЪ ЗАТЪКАТИ	1	14.43	ПсДС: АСПИДЪ ГЛОУХЪ ЗАТЪКАЮЩЕИ 67.2.1-17
3	МЕУЪ ОБЮДОУ ОСТРЪ	1	13.85	ПсДС: МЕУИ ОБЮДОУ ОСТРИ 144.2.1-10
4	ЗЕЛНИ ЗЛАКЪ ОТЪПАСТИ	1	12.85	ПсДС: ЗЕЛИЕ ЗЛАКА [СКОРО] ОТЪПАДЕТЪ 47.1.1-8
5	ТОУКЪ ПЪШЕННУЪНЪ НАСЪЩАТИ	1	12.26	ПсДС: ТЪКА ПЪШЕНИУЪНА НАСЪЩАНА 142.2.1-21
6	ПАМФНЛЪ ПАВЪЛЪ ПОРФНЛЪ	1	12.11	ЕвАС: ПАМФНЛА [ОУАЛЕНТА] ПАВЛА ПОРФНЛА 142.2.1-14
7	СТАРЪЦЪ СЕМЕОНЪ ПРОВЪЗВѢСТНТИ	1	12.11	ЛитС: СТАРЪЦЮ СЕМЕОНОУ [...] ПРОВЪЗВѢСТИ 56.1.1-18
8	НОВЪ МѢСАЦЪ ТРЖБА	1	12.04	ПсДС: НОВЪ МѢСАЦЪ ТРЖБОУ 90.2.1-3
9	СРѢДА СКОУМЪНЪ ПОСЪПАТИ	1	12.04	ПсДС: СРѢДЪ СКОМЪНЪ ПОСЪПАХЪ 67.1.1-11
10	ЛЪВЪ ВЪСЪХЪТАТИ РНКАТИ	1	11.97	ПсДС: ЛЪВЪ ВЪСЪХЪТАНА [И] РНКАНА 30.1.1-17
11	КЪЛЪ ДАЛЕУЕ ОТЪСТОЪТИ	1	11.79	ПсДС: КЪЛЪ ДАЛЕУЕ ОТЪСТОЪТЪ 105.2.1-18
12	ОУТВЪРЪЖЕННИ ХЛѢБЪНЪ СЪТЪРТИ	1	11.68	ПсДС: ОУТВЪРЪЖЕНИЕ ХЛѢБНО СЪТЪРЪ 108.1.1-14
13	ЗЕЛНИ ЗЛАКЪ СКОРО	1	11.53	ПсДС: ЗЕЛИЕ ЗЛАКА СКОРО 47.1.1- 8
14	СЪВНТЪКЪ КЪНИЖЪНЪ ПИСАТИ	1	11.48	ПсДС: СЪВИТЪЦЪ КЪНИЖЪНЪМЪ ПИШЕТЪ 51.2.1- 15
15	РОВЪ НЗДЪРЪТИ НСКОПАТИ	1	11.39	ПсДС: РОВЪ НЗДЪРЪН НСКОПА 9.2.1-8

<sup>15</sup> Компоненты — леммы, количество компонентов — 3, расстояние между компонентами от 0 до 1.

<sup>16</sup> В этом и в следующих приложениях приводятся первые 20 записей списков.

№	N-грамма	F	MI	Адрес
16	оукрашѣти страстьнъ побѣда	1	11.39	ЛитС: оукрашѣна страстьнами побѣдами 63.2.1-11
17	злакъ скоро отъпасти	1	11.11	ПсДС: злака скоро отъпадетъ 47.1.1-9
18	прелюбодѣи оучасти полагати	1	11.04	ПсДС: прелюбодѣю оучасти [свое] полагаше 61.2.1-18
19	провѣщати ганаани испытати	1	11.04	ПсДС: Провѣщаю ганаани испытати 85.1.1-24
20	вина неразтворенъ растворенни	1	10.98	ПсДС: вина неразтворена [испльнъ] раствореніе 83.1.1- 25

### Приложение 3.

#### Результат применения меры T-score к словоформам<sup>17</sup>

№	N-грамма	F	F (w <sub>1</sub> )...F (w <sub>x</sub> )*	T-score	Рукопись
1	въ вѣкъ	94	вѣ (2099) вѣкъ (111)	9.42	ПсДС, ЕвАс
2	реѹе гѣ	67	реѹе (449) гѣ (116)	8.11	ЕвАс
3	онъ же	61	онъ (83) же (1185)	7.67	ЛитС, ЕвАс
4	реѹе емоу	62	реѹе (449) емоу (396)	7.62	ЕвАс
5	къ немоу	57	къ (227) немоу (63)	7.53	ЕвАс
6	да не	76	да (728) не (1282)	7.51	ЛитС, ПсДС, ЕвАс, КЛ
7	н реѹе	92	н (4541) реѹе (449)	7.19	ЛитС, ЕвАс
8	реѹе нмъ	53	реѹе (449) нмъ (211)	7.13	ЕвАс
9	#####** #####	50	##### (125) ##### (125)	7.05	ЛитС
10	гѣа емоу	49	гѣа (133) емоу (396)	6.92	ЕвАс
11	ѣоу гѣа	47	ѣоу (60) гѣа (152)	6.84	ЕвАс
12	къ тебѣ	42	къ (177) тебѣ (279)	6.40	МинС, ЛитС, ПсДС, ЕвАс, КЛ
13	же реѹе	52	же (1185) реѹе (449)	6.39	ЛитС, ЕвАс
14	на ма	50	на (1046) ма (522)	6.20	ПсДС, ЕвАс
15	са въ	102	са (1793) въ (2099)	5.90	МинС, ПсДС
16	гѣж вамъ	35	гѣж (73) вамъ (183)	5.89	ЕвАс
17	се го радѣ	34	се го (89) радѣ (106)	5.81	ПсДС, ЕвАс

<sup>17</sup> Компоненты — словоформы, количество компонентов — 2, расстояние между компонентами — 0, учет границ конструкций — «да», исключить служебные слова — «нет».

№	N-грамма	F	F (w <sub>1</sub> )...F (w <sub>x</sub> )*	T-score	Рукопись
18	реуе же	44	реуе (449) же (1185)	5.73	ЕвАс
19	отъ мене	34	отъ (533) мене (169)	5.66	ПсДС, ЕвАс
20	са тебѣ	41	са (1793) тебѣ (279)	5.52	ПсДС, ЕвАс

\* Частота каждого из коллокатов в коллекции.

\*\* Порча, нечитаемые словоформы.

#### Приложение 4. Первые биграммы трех рукописей (T-score)<sup>18</sup>

№	АсЕв		ПсДС		ЛитС	
	N-грамма	T-score	N-грамма	T-score	N-грамма	T-score
1	реуе гѣ	8.05	вѣ вѣкѣ	9.31	молимъ се	5.11
2	онъ же	7.51	на ма	6.15	бже нашъ	4.33
3	къ немоу	7.51	къ тебѣ	5.85	раба твоего	4.10
4	реуе емоу	7.40	и не	5.63	гї бже	3.36
5	реуе нмъ	7.01	да не	5.52	хѣ бже	3.30
6	гѣ емоу	6.83	са тебѣ	5.41	вѣ нстннж	3.27
7	ѣоу гѣ	6.83	азъ же	5.41	гдемь	3.15
					нашнмъ	
8	гѣж вамъ	5.87	отъ мене	5.40	на земн	2.80
9	же реуе	5.84	са вѣ	5.20	се гї	2.67
10	гѣ нмъ	5.36	на та	5.11	дхомъ	2.63
					твонмъ	
11	ѣвѣ ш <sup>т</sup>	5.36	дша моѣ	5.07	нъ н	2.45
12	же роди	5.31	восъ денъ	4.68	стааго дха	2.44
13	реуе же	5.24	срѣдъце мое	4.55	тебѣ	2.43
					прноснмъ	
14	вѣ нма	5.14	на земи	4.46	гдемь нѣ	2.43
15	н реуе	5.13	сего ради	4.35	нсноу	2.36
16	на десѣте	5.05	постыдѣтъ	4.32	н реуе	2.32
			са			
17	емоу нсѣ	4.98	ѣко ты	4.22	ты бо	2.23
18	да не	4.97	законъ твои	4.22	отъ грѣхъ	2.21
19	азъ есмъ	4.96	ѣъ мои	4.16	сна твоего	2.21
20	ѣвѣ о <sup>т</sup>	4.94	ма отъ	4.11	вѣко гї	2.21

<sup>18</sup> Компоненты — словоформы, количество компонентов — 2, расстояние между компонентами — 0, учет границ конструкций — «да», исключить служебные слова — «нет».

Приложение 5.

Первые 20 наиболее частых словоформ АсЕв, ПсДС и ЛитС<sup>19</sup>

№	АсЕв Словоформа	Freq*	ПсДС Слово- форма	Freq	ЛитС Словоформа	Freq
1	<b>н</b>	0.13099	<b>н</b>	0.11302	<b>н</b>	0.07079
2	<b>въ</b>	0.02710	<b>въ</b>	0.03506	<b>твон</b>	0.01708
3	<b>бытн</b>	0.02659	<b>сѧ</b>	0.03187	<b>се (сѧ)</b>	0.01451
4	<b>же</b>	0.02239	<b>мон</b>	0.03004	<b>въ</b>	0.01371
5	<b>сѧ</b>	0.02014	<b>твон</b>	0.02648	<b>да</b>	0.01227
6	<b>азъ</b>	0.01386	<b>азъ</b>	0.02518	<b>нашь</b>	0.00954
7	<b>рѣшн</b>	0.01375	<b>бытн</b>	0.02097	<b>богъ</b>	0.00946
8	<b>ѧко</b>	0.01236	<b>господь</b>	0.02029	<b>бытн</b>	0.00858
9	<b>на</b>	0.01107	<b>ѧко</b>	0.01941	<b>на</b>	0.00802
10	<b>нсовъ</b>	0.00996	<b>на</b>	0.01615	<b>нже</b>	0.00762
11	<b>нже</b>	0.00792	<b>отъ</b>	0.01318	<b>господь</b>	0.00746
12	<b>да</b>	0.00751	<b>богъ</b>	0.01311	<b>вьсь</b>	0.00633
13	<b>глаголатн</b>	0.00742	<b>тъ</b>	0.01076	<b>отъ</b>	0.00553
14	<b>свон</b>	0.00725	<b>свон</b>	0.01001	<b>мъ</b>	0.00545
15	<b>къ</b>	0.00702	<b>вьсь</b>	0.00930	<b>тъ</b>	0.00537
16	<b>съ</b>	0.00670	<b>да</b>	0.00711	<b>съ</b>	0.00433
17	<b>ы</b>	0.00647	<b>земля</b>	0.00581	<b>же</b>	0.00393
18	<b>вьсь</b>	0.00647	<b>въкъ</b>	0.00541	<b>дохъ</b>	0.00369
19	<b>отъ</b>	0.00584	<b>же</b>	0.00502	<b>ѧко</b>	0.00329
20	<b>мон</b>	0.00582	<b>тъ</b>	0.00483	<b>свон</b>	0.00305

\* Относительная частота в списке.

<sup>19</sup> Компоненты — словоформы, количество компонентов — 1, учет границ конструкций — «да», исключить служебные слова — «нет».

Приложение 6.  
 Результат применения меры T-score к леммам<sup>20</sup>

№	N-грамма	F	F (w <sub>1</sub> )...F (w <sub>x</sub> )	T-score	Адрес
1	<b>ОНЪ ЖЕ</b>	106	<b>ОНЪ</b> (289) <b>ЖЕ</b> (2366)	9.55	ЛитС, ЕвАс
2	<b>ДОУША МОН</b>	99	<b>ДОУША</b> (244) <b>МОН</b> (2319)	9.31	ПсДС, ЕвАс
3	<b>НА ЗЕМЛЯ</b>	61	<b>НА</b> (1080) <b>ЗЕМЛЯ</b> (232)	7.45	ЛитС, ПсДС, ЕвАс
4	<b>БЪ БЪКЪ</b>	116	<b>БЪ</b> (9595) <b>БЪКЪ</b> (333)	7.42	ЛитС, ПсДС, ЕвАс
5	<b>РЕЩН ГОСПОДЬ</b>	90	<b>РЕЩН</b> (1229) <b>ГОСПОДЬ</b> (1511)	7.28	ЛитС, ПсДС
6	<b>БОГЪ НАШЬ</b>	63	<b>БОГЪ</b> (1093) <b>НАШЬ</b> (555)	7.08	Час, МинС, ПсДС, ЕвАс, КЛ
7	<b>НА НЕБО</b>	52	<b>НА</b> (1080) <b>НЕБО</b> (165)	6.9	ЛитС, ПсДС, ЕвАс
8	<b>НСПОВЪДЪТН СА</b>	51	<b>НСПОВЪДЪТН</b> (73) <b>СА</b> (2935)	6.80	ПсДС
9	<b>СРЪДЪЦЕ МОН</b>	53	<b>СРЪДЪЦЕ</b> (175) <b>МОН</b> (2319)	6.65	МинС, ПсДС
10	<b>МНЛОСТЬ ТВОИ</b>	53	<b>МНЛОСТЬ</b> (225) <b>ТВОИ</b> (2201)	6.51	ЛитС, ПсДС, КЛ
11	<b>БЪСЪ МНРЪ</b>	40	<b>БЪСЪ</b> (1126) <b>МНРЪ</b> (137)	6.05	МинС, ЛитС, ПсДС, ЕвАс
12	<b>РАБА ТВОИ</b>	47	<b>РАБА</b> (224) <b>ТВОИ</b> (2201)	6.04	ЛитС, ПсДС, ЕвАс
13	<b>ОТЬЦЪ МОН</b>	48	<b>ОТЬЦЪ</b> (271) <b>МОН</b> (2319)	5.91	ЛитС, ПсДС, ЕвАс
14	<b>БРАГЪ МОН</b>	42	<b>БРАГЪ</b> (150) <b>МОН</b> (2319)	5.88	ПсДС
15	<b>ГОСПОДЬ БОГЪ</b>	65	<b>ГОСПОДЬ</b> (1511) <b>БОГЪ</b> (1093)	5.75	МинС, ЛитС, ПсДС, ЕвАс
16	<b>БЪЗВЕСЕЛНТН СА</b>	39	<b>БЪЗВЕСЕЛНТН</b> (99) <b>СА</b> (2935)	5.72	ПсДС, ЕвАс
17	<b>БЪЗВРАТНТН СА</b>	37	<b>БЪЗВРАТНТН</b> (108) <b>СА</b> (2935)	5.50	ПсДС, ЕвАс
18	<b>МОЛНТВА МОН</b>	33	<b>МОЛНТВА</b> (67) <b>МОН</b> (2319)	5.44	ПсДС
19	<b>ОУПЪБАТН НА</b>	33	<b>ОУПЪБАТН</b> (145) <b>НА</b> (1080)	5.44	ПсДС
20	<b>НСОУСЪ ЖЕ</b>	50	<b>НСОУСЪ</b> (433) <b>ЖЕ</b> (2366)	5.44	ЕвАс

<sup>20</sup> Компоненты — леммы, количество компонентов — 2, расстояние между компонентами — 0, учет границ конструкций — «да», исключить служебные слова — «нет».

Приложение 7.  
 Результат применения меры *Dice* к словоформам<sup>21</sup>

№	N-грамма	F	F(w <sub>1</sub> )...F(w <sub>x</sub> )	Dice	Адрес
1	БРАДЖ АРЎНЖ	2	БРАДЖ (2) АРЎНЖ (2)	1.00	ПсДС: 130.2.1-19, 130.2.1-20
2	ВЪЗВЪША ВЪТРН	2	ВЪЗВЪША (2) ВЪТРН (2)	1.00	ЕвАс: 35.2.2-2, 35.2.2-18
3	ВЪНЕГЪДА ПРІДЕ	2	ВЪНЕГЪДА (2) ПРІДЕ (2)	1.00	ПсДС: 62.1.1-9, 63.1.1-20
4	ДРОЖДЪЯ ПРНВАЗАТН	2	ДРОЖДЪЯ (2) ПРНВАЗАТН (2)	1.00	МедЛ: 2.1.1-4, 2.2.1-3
5	ДЪХА СЪТА	2	ДЪХА (2) СЪТА (2)	1.00	ЕвАс: 131.1.1-19, 131.2.1-2
6	ЗЪЛЪІХЪ ӨКОУСЪНЪІХЪ	2	ЗЪЛЪІХЪ (2) ӨКОУСЪНЪІХЪ (2)	1.00	ЛитС: 14.1.1-20, 71.1.1-18
7	ІДЪМЪЮ ПРОСТЪРЖ	2	ІДЪМЪЮ (2) ПРОСТЪРЖ (2)	1.00	ПсДС: 69.2.1- 8,113.1.1-17
8	ІОНОПЛЕМЕНЪНІЦІ ПОКОРИША	2	ІОНОПЛЕМЕНЪНІЦІ (2) ПОКОРИША (2)	1.00	ПсДС: 69.2.1- 9,113.1.1-18
9	КЕСАРНЯ ФІАНПОВЪН	2	КЕСАРНЯ (2) ФІАНПОВЪН (2)	1.00	ЕвАс: 78.1.2-18, 149.2.1-27
10	КОЕЖ СЪМРЪТНІЖ	3	КОЕЖ (3) СЪМРЪТНІЖ (3)	1.00	ЕвАс: 27.1.2-1, 31.1.2-1, 101.2.2-12
11	МОЪВЪ КОНОБЪ	2	МОЪВЪ (2) КОНОБЪ (2)	1.00	ПсДС: 69.2.1- 6,113.1.1-16
12	НИЩЕТОЖ ӨБОГАТНМЪ	2	НИЩЕТОЖ (2) ӨБОГАТНМЪ (2)	1.00	ЛитС: 53.2.1-17, 62.1.1-13
13	ПЕСЪЯЯ МЪХЪІ	2	ПЕСЪЯЯ (2) МЪХЪІ (2)	1.00	ПсДС: 87.1.1- 24,108.2.1-16
14	ПОЛІ ТАНЪЎСЪ	2	ПОЛІ (2) ТАНЪЎСЪ (2)	1.00	ПсДС: 86.1.1- 4,87.1.1-20
15	ПОМОРИЪ ТУРЬСКА	2	ПОМОРИЪ (2) ТУРЬСКА (2)	1.00	ЕвАс: 123.2.1-9, 150.2.1-9
16	ПРОСТЪРЖ САПОГЪ	2	ПРОСТЪРЖ (2) САПОГЪ (2)	1.00	ПсДС: 69.2.1-8, 113.1.1-17
17	СЕЛОМЪ РАЗМЪРЖ	2	СЕЛОМЪ (2) РАЗМЪРЖ (2)	1.00	ПсДС: 69.2.1-3, 113.1.1-12

<sup>21</sup> Компоненты — словоформы, количество компонентов — 2, расстояние между компонентами — 0, учет границ конструкций — «да», исключить служебные слова — «нет».

№	N-грамма	F	F (w <sub>1</sub> )...F (w <sub>x</sub> )	Dice	Адрес
18	<b>сѣврѣшено</b> <b>любнѣн</b>	2	<b>сѣврѣшено</b> (2) <b>любнѣн</b> (2)	1.00	ЛитС: 14.2.1-7, 72.1.1-4
19	<b>ѣготовати</b> <b>трѣпѣзж</b>	2	<b>ѣготовати</b> (2) <b>трѣпѣзж</b> (2)	1.00	ПсДС: 86.1.1-17, 86.1.1-21
20	<b>ѣба прнѣстѣ</b>	2	<b>ѣба</b> (2) <b>прнѣстѣ</b> (2)	1.00	ЛитС: 39.2.1-14, 40.1.1-12

### Приложение 8.

Результат применения меры *Dice* к словоформам со свободным порядком их следования в биграмме<sup>22</sup>

№*	N-грамма	F	F (w <sub>1</sub> )...F (w <sub>x</sub> )	Dice	Адрес
2	<b>побѣдамн</b> <b>снѣтн</b>	2	<b>побѣдамн</b> (4) <b>снѣтн</b> (2)	0.67	ЛитС: 47.2.1-23, 66.1.1-16
3	<b>принѣжтѣ</b> <b>дарѣ</b>	2	<b>принѣжтѣ</b> (3) <b>дарѣ</b> (3)	0.67	ПсДС: 75.2.1-12, 80.1.1-8
5	<b>нзвлѣѣ</b> <b>ножѣ</b>	2	<b>нзвлѣѣ</b> (3) <b>ножѣ</b> (5)	0.50	ЕвАс: 90.2.2-12, 100.1.2-6
7	<b>свонмѣ</b> <b>оѣѣннкомѣ</b>	17	<b>свонмѣ</b> (45) <b>оѣѣннкомѣ</b> (33)	0.44	ЕвАс: 14.1.1-25, 22.1.2-16, 30.2.2-11, 34.1.1-25, 45.2.2- 15.12, 60.1.2-20, 69.1.1-15, 87.1.1-17, 120.2.1-19, 121.2.1-17, 124.2.1-13, 128.2.1-4, 128.2.1-27, 151.1.1-13, 156.2.1-10, 157.2.1-1, 157.2.1-4
10	<b>мѣногю</b> <b>силю</b>	2	<b>мѣногю</b> (2) <b>силю</b> (8)	0.40	ПсДС: 42.2.1-22, 74.2.1-24
12	<b>многѣ</b> <b>прѣлѣстѣтѣ</b>	2	<b>многѣ</b> (9) <b>прѣлѣстѣтѣ</b> (4)	0.31	ЕвАс: 48.2.1-24, 49.1.1-4
13	<b>оѣѣннкомѣ</b> <b>свонм</b>	3	<b>оѣѣннкомѣ</b> (12) <b>свонм</b> (8)	0.30	ЕвАс: 41.2.1-13, 49.2.1-17, 67.2.1-21

<sup>22</sup> Компоненты — словоформы, количество компонентов — 2, расстояние между компонентами — 0, учет границ конструкций — «да», исключить служебные слова — «нет», закрепленность — «свободные», следование — «нет».

№*	N-грамма	F	F(w <sub>1</sub> )...F(w <sub>x</sub> )	Dice	Адрес
14	ГЛѢЖ ВАМЪ	37	ГЛѢЖ (73) ВАМЪ (183)	0.29	ЕвАс: 8.2.2-20, 11.1.2-21, 13.1.1-13, 14.1.1-9, 23.1.1-4, 33.1.2-17, 33.1.2-29, 34.2.1-8, 36.1.2-29, 43.1.2-22, 45.2.2-21, 48.1.2-25, 62.2.2-4, 62.2.2-29, 67.1.1-25, 67.2.1-10, 69.1.2-15, 69.1.2-24, 72.2.2-18, 76.2.2-27, 84.2.1-6, 85.1.2-10, 87.2.2-12, 91.2.1-12, 92.2.1-15, 93.1.2-18, 93.2.1-2, 97.1.2-13, 97.2.1-1, 97.2.1-7, 115.1.2-1, 119.1.1-12, 124.1.1-16, 124.2.1-28, 132.1.1-28, 143.1.1-19, 147.1.1-7
16	СТРАХЪ НАПАДЕ	2	СТРАХЪ (11) НАПАДЕ (4)	0.27	ПсДС: 109.1.1-6, ЕвАс: 148.1.1-14
17	ѠБО БІМЪ	2	ѠБО (15) БІМЪ (8)	0.26	ПсДС: 65.1.1-20, 91.1.1-8
18	ХВАЛНМЪ ТЕ	3	ХВАЛНМЪ (4) ТЕ (20)	0.25	ЛитС: 19.1.1-10, 40.1.1-20, 57.1.1-19
21	РЕЧЕ ГЪ	68	РЕЧЕ (449) ГЪ (116)	0.24	ЕвАс: 8.1.2-12, 9.1.2-5, 9.2.2-7, 13.1.1-23, 13.1.2-27, 13.2.2-6, 14.1.1-25, 15.1.1-13, 23.1.1-20, 23.2.1-24, 27.1.2-28, 29.2.1-13, 30.1.2-25, 30.2.1-6, 32.2.2-2, 33.2.1-21, 34.1.1-25, 34.2.2-10, 36.1.1-27, 39.2.1-22, 44.1.2-11, 47.2.1-16, 49.1.2-26, 49.2.1-17, 53.2.1-14, 55.1.1-3, 55.2.2-13, 60.1.1-6, 60.1.2-20, 61.1.1-7, 61.1.2-7, 61.2.2-15, 62.1.2-24, 64.2.2-14, 65.2.1-27, 66.2.2-8, 67.1.2-3, 67.2.1-21, 68.1.1-22, 69.1.1-15, 69.2.2-5, 71.2.2-24, 73.1.1-3, 73.2.2-24, 76.2.1-22, 87.1.2-9, 92.1.2-24, 114.1.1-2, 114.2.1-2, 115.1.1-2, 116.1.1-21, 116.2.2-16, 116.2.2-22, 116.2.2-27, 117.1.1-4, 118.2.1-24, 120.1.1-24, 120.2.1-19, 123.1.1-25, 124.1.1-5, 126.1.1-11, 127.2.1-17, 128.2.1-4, 128.2.1-27, 141.2.1-10, 143.2.1-4, 151.1.1-13, 158.1.1-10

№*	N-грамма	F	F (w <sub>1</sub> )...F (w <sub>x</sub> )	Dice	Адрес
23	<b>МНОГЪ</b> <b>НАРОДЪ</b>	4	<b>МНОГЪ</b> (12) <b>НАРОДЪ</b> (23)	0.23	ЕвАс: 37.1.1-28, 52.1.2-15, 82.1.2-17, 90.2.1-15
24	<b>АЗЪ ЕСМЪ</b>	28	<b>АЗЪ</b> (152) <b>ЕСМЪ</b> (95)	0.23	ЕвАс: 11.1.1-11, 13.2.2-9, 13.2.2- 19, 23.1.1-4, 23.1.1-12, 24.2.1- 26, 26.2.2-25, 28.1.1-15, 29.1.2- 2, 32.2.1-16, 48.2.1-23, 71.1.2- 9, 80.2.1-19, 82.2.2-27, 89.1.1-9, 93.1.1-10, 93.1.1-19, 94.2.1-20, 94.2.2-19, 100.1.1-8, 100.1.1-13, 100.1.1-23, 102.1.1-25, 114.1.1- 16, 114.1.2-7, 114.2.1-2, 125.1.1- 27, 148.2.1-11
25	<b>МОНХЪ</b> <b>СЛОВЕСЪ</b>	3	<b>МОНХЪ</b> (18) <b>СЛОВЕСЪ</b> (10)	0.21	ЕвАс: 76.2.2-17, 94.1.1-25, 120.2.1-8.15
26	<b>ИСТИНѢ ТВОЕИ</b>	2	<b>ИСТИНѢ</b> (5) <b>ТВОЕИ</b> (24)	0.21	ПсДС: 94.1.1-2, 133.2.1-10
27	<b>АЗЪ ЕСМЪ</b>	21	<b>АЗЪ</b> (117) <b>ЕСМЪ</b> (95)	0.20	ЛитС (MSC, Cod. Sin. slav. 5/N): 43.2.1-20, ПсДС: 34.2.1-3, 44.2.1-14, 51.1.1- 12, 58.1.1-21, 61.1.1-15, 77.2.1- 6, 93.2.1-10, 95.1.1-15, 114.2.1-1, 120.1.1-1, 121.2.1-14, 123.1.1-6, 124.1.1-19, 124.2.1-22, 137.1.1- 10, ЕвАс: 28.1.1-29, 29.1.1-23, 42.1.1-13, 89.1.1-29, 99.1.2-20
28	<b>НМАТЪ</b> <b>ЖИВОТЪ</b>	7	<b>НМАТЪ</b> (42) <b>ЖИВОТЪ</b> (29)	0.20	ЕвАс: 8.1.2-22, 9.1.2-14, 9.2.1- 1, 13.2.1-5, 14.1.1-18, 32.2.1-21, 117.1.1-17
29	<b>МНЛОСТЬ</b> <b>ТВОЮЖ</b>	3	<b>МНЛОСТЬ</b> (7) <b>ТВОЮЖ</b> (24)	0.19	ЛитС: 22.2.1-10, 27.1.1-19, ПсДС: 17.1.1-1
30	<b>ГРѢШЪНИЦИ</b> <b>ПОГЫБЕНЖТЪ</b>	2	<b>ГРѢШЪНИЦИ</b> (15) <b>ПОГЫБЕНЖТЪ</b> (6)	0.19	ПсДС: 48.1.1-2, 74.1.1-22

\* Из списка удалены случаи типа **ЛЮБО МАСЛО ЛЮБО ЛОИ** (МедЛ, 3.2.1-3), в которых между синтагмами отсутствует знак-разделитель, что приводит к автоматическому построению свободной биграммы **ЛЮБО МАСЛО** (2×).

Приложение 9.  
 Результат применения меры *Dice* к леммам со свободным порядком следования словоформ  
 в биграмме<sup>23, 24</sup>

№	N-грамма	F	F (w <sub>1</sub> )...F (w <sub>n</sub> )	Dice	Адрес
2	пѣрьъ послѣдѣнь	4	пѣрьъ (26) послѣдѣнь (16)	0.19	ЕвАс: прѣвъи послѣдѣньнн 34.2.2-6, послѣдѣньнн прѣвъи 34.2.2-8, послѣдѣньнн прѣвъи 144.1.1-25, прѣвъи послѣдѣньнн 144.1.1-26
4	урѣмьнъ морѣ	5	урѣмьнъ (5) морѣ (50)	0.18	ПсДс: урѣмьнъное морѣ 109.2.1-16, урѣмьнънѣмѣ морю 109.2.1-19, морі урѣмьнънѣмѣ 110.1.1-19, урѣмьнъное морѣ 132.2.1-3, урѣмьнъное морѣ 132.2.1-8
5	коуръ възгласнтн	3	коуръ (10) възгласнтн (30)	0.15	ЕвАс: коуръ възгласи 92.1.1-17, възгласитъ коуръ 92.2.2-20, коуръ възгласн 101.1.2-16
6	дѣрь затворнтн	4	дѣрь (38) затворнтн (18)	0.14	ЕвАс: дѣрьемъ затвореномъ 7.1.1-5, дѣрьемъ затвореномъ 7.1.2-27, затворитъ дѣрь 62.2.2-11, затвори дѣрь 73.2.1-1
7	принестн дара	3	принестн (38) дара (7)	0.13	ПсДс: принежтъ дара 75.2.1-12, дара принежтъ 80.1.1-8, ЕвАс: принеси дара 36.1.1-22
8	болѣтн цѣлнтн	2	болѣтн (20) цѣлнтн (11)	0.13	ЕвАс: болаштара цѣлнтѣ 125.2.1-19, цѣлнтн болаштара 150.2.1-21
9	нѣ	7	нѣ (76) 7(36)	0.13	ЕвАс: нѣ 7(21.1.1-18, 7(127.1.1-12, 7(129.2.1-20, нѣ 7(32.2.1-6, 7(152.1.1-9, нѣ 7(158.1.1-7, нѣ 7(158.2.1-3
10	наростъ гнѣбѣтн	2	наростъ (25) гнѣбѣтн (12)	0.11	ПсДс: вростъ гнѣба 77.1.1-17, гнѣба врості 93.1.1-11

<sup>23</sup> Компоненты — леммы, количество компонентов — 2, расстояние между компонентами — 0, учет границ конструкций — «да», исключить служебные слова — «нет», закреплённость — «свободные», следование — «нет».

<sup>24</sup> Из списка удалены случаи типа **любю масло лои** (МедЛ, 3.2.1-3), в которых между синтагмами отсутствует знак-разделитель, что приводит к автоматическому построению свободной биграммы **сюбо масло** (2×).

№	N-грамма	F	F (w)...F (w)	Dice	Адрес
11	Н-теръ УЛОБЪКЪ	13	Н-теръ (51) УЛОБЪКЪ (211)	0.10	ЕвАс: Җтеръ Улкъ 14.2.1-7, Җтера Улка 32.2.2-21, Үкь Җтеръ 43.1.1-10, Үлкь Җтеръ 46.1.2-26, Үлкь Җтеръ 55.1.1-3, Үкь Җтеръ 57.1.2-24, Үкь Җтеръ 59.1.1-17, Үлкь Җтеръ 61.2.2-16, Үлкь Җтеръ 63.2.1-21, Үкь Җтеръ 64.1.2-4, Үлкь Җтеръ 68.1.1-24, Үлкь Җтеръ 69.2.2-7, Үлкь Җтеръ 77.1.2-24 ЛитС: на Җемн 43.2.1-10, на Җемн 44.1.1-10, на Җемн 44.1.1-21, на Җемн 44.2.1-5, на Җемн 44.2.1-12, на Җемн 44.2.1-16, на Җемн 44.2.1-10, на Җемн 54.2.1-20, на Җемн 69.1.1-18, ПсЦС: на Җемн 8.2.1-6, на Җемн 18.2.1-14, на Җемн 20.2.1-14, на Җемн 37.1.1-23, на Җемн 42.2.1-19, на Җемн 52.2.1-18, на Җемн 58.1.1-16, на Җемн 58.1.1-23, на Җемн 67.2.1-11, на Җемн 68.1.1-11, на Җемн 74.1.1-5, на Җемн 74.1.1-11, на Җемн 79.2.1-25, на Җемн 80.1.1-23, Җемн на 80.1.1-23, на Җемн 81.2.1-15, на Җемн 96.2.1-26, на Җемн 97.1.1-11, на Җемн 104.2.1-24, на Җемн 107.2.1-5, на Җемн 108.1.1-14, на Җемн 115.2.1-13, на Җемн 116.1.1-20, на Җемн 116.2.1-24, на Җемн 120.1.1-1, на Җемн 122.2.1-16, на Җемн 131.1.1-20, Җемн на 132.1.1-16, на Җемн 133.1.1-8, на Җемн 136.1.1-10, на Җемн 137.1.1-3, на Җемн 137.1.1-26, на Җемн 138.1.1-13, на Җемн 140.2.1-17, ЕвАс: на Җемн 11.1.1-15, на Җемн 33.1.2-20, на Җемн 33.1.2-24, на Җемн 33.2.1-2, на Җемн 39.2.1-8, на Җемн 47.2.1-1, на Җемн 51.2.1-10, на Җемн 53.2.2-7, на Җемн 71.1.2-26, на Җемн 73.2.2-3, на Җемн 74.1.2-8, на Җемн 76.1.1-29, на Җемн 77.2.1-25, на Җемн 98.1.1-20, на Җемн 100.1.1-15, на Җемн 112.2.1-14, на Җемн 124.1.1-25, на Җемн 133.1.1-19, на Җемн 150.1.1-20, на Җемн 150.1.1-23
12	на Җемла	63	на (1080) Җемла (232)	0.10	ЕвАс: Җҗ127.1.1-9, Җҗ129.2.1-16, Җҗ132.2.1-4, Җҗ135.2.1-11, Җҗ152.1.1-26, Җҗ156.2.1-6 ЕвАс: Җҗ127.1.1-20, Җҗ128.2.1-26, Җҗ135.2.1-16, Җҗ150.2.1-23, Җҗ150.2.1-26 ПсЦС: Җҗҗамъ Җҗалъ 66.2.1-14, ЕвАс: Җҗҗалъ Җҗҗалъ 11.1.2-6, Җҗҗалъ Җҗҗалъ 22.1.2-8, Җҗҗалъ Җҗҗалъ 41.2.1-11, Җҗҗалъ Җҗҗалъ 89.1.2-12
13	Җҗ	6	Җ(55) җ(76)	0.10	
14	Җҗ	5	Җ(37) җ(76)	0.09	
15	Җҗалъ Җҗҗалъ	5	Җҗалъ (57) Җҗҗалъ (57)	0.09	

№	N-грамма	F	F (w)...F (w <sub>i</sub> )	Dice	Адрес
16	НА НЕБО	54	НА (1080) НЕБО (165)	0.09	ЛитС: НА НЕБО 16.2.1-18, НА НЕБЕХЪ 19.2.1-18, НА НЕБА 38.1.1-16, ПсЦА: НА НЕБЕСИ 46.2.1-4, НА НЕБЕСА 67.1.1-14, НА НЕБЕСА 67.2.1-3, НА НЕБО 75.2.1-20, НЕБА НА 75.2.1-20, НА НЕБЕСИ 81.1.1-3, НА НЕБЕСИ 81.2.1-14, НА НЕБЕХЪ 95.2.1-2, НА НЕБЕСИ 96.2.1-22, НЕБЕСИ НА 104.2.1-24, НА НЕБЕСИ 106.1.1-7, НА НЕБЕСА 113.1.1-4, НА НЕБИ 116.1.1-20, НА НЕБЕСИ 122.2.1-21, НА НЕБЕСИ 127.2.1-2, НА НЕБО 134.2.1-1, ЕвАс: НА НЕО 28.2.2-18, НА НЕО 30.1.2-5, НА НЕБЕХЪ 32.2.2-10, НА НЕБЕХЪ 32.2.2-13, НА НЕБЕХЪ 33.1.2-22, НА НЕБЕХЪ 33.1.2-27, НА НЕБЕХЪ 33.2.1-9, НА НЕБЕХЪ 34.1.2-5, НА НЕБЕХЪ 34.1.2-12, НА НЕО 41.1.1-28, НА НЕБЕХЪ 45.2.2-7, НА НЕБЕХЪ 47.2.1-3, НА НЕБЕХЪ 60.2.1-10, НА НЕБЕХЪ 64.2.1-10, НА НЕО 67.2.1-4, НА НЕО 70.1.2-3, НА НЕО 70.1.2-23, НА НЕБЕХЪ 73.1.1-15, НА НЕБИ 73.2.1-26, НА НЕБИ 73.2.2-3, НА НЕБИ 74.1.2-17, НА НЕО 77.1.1-28, НА НЕБИ 84.1.1-29, НА НЕО 97.2.2-28, НА НЕБЕХЪ 113.2.2-22, НА НЕО 117.1.1-4, НА НЕБИ 117.1.1-6, НА НЕБЕХЪ 119.1.1-8, НА НЕБЕХЪ 123.1.1-23, НА НЕБЕХЪ 124.2.1-4, НА НЕБЕХЪ 129.1.1-19, НА НЕО 133.1.1-22, НА НЕБЕХЪ 150.1.1-14, НА НЕБЕХЪ 150.1.1-24, НА НЕО 154.2.1-4
17	ЗАБЕГЪ ЗАБЪЦАТИ	2	ЗАБЕГЪ (45) ЗАБЪЦАТИ (4)	0.09	ПсЦА: ЗАБЕГЪ ЗАБЪЦАША 91.2.1-19, ЗАБЪЦАХЪ ЗАБЕГЪ 95.2.1-3
18	ОНЪ ЖЕ	107	ОНЪ (289) ЖЕ (2366)	0.09	ЛитС: ОНЪ ЖЕ 43.2.1-19, ЕвАс: ОНА ЖЕ 3.1.1-28, ОНЪМА ЖЕ 3.2.2-8, ОНА ЖЕ 4.1.1-23, ОНЪ ЖЕ 6.1.2-8, ОНЪ ЖЕ 7.1.2-13, ОНИ ЖЕ 8.1.1-12, ОНЪ ЖЕ 11.1.1-10, ОНЪ ЖЕ 14.2.2-13, ОНЪ ЖЕ 24.2.1-25, ОНЪ ЖЕ 24.2.2-21, ОНЪ ЖЕ 25.1.1- 10, ЖЕ ОНЪ 25.1.2-29, ОНИ ЖЕ 25.2.1-17, ОНЪ ЖЕ 26.1.1-10, ОНИ ЖЕ 28.1.2-14, ОНА ЖЕ 35.1.2-19, ОНИ ЖЕ 38.2.1-14, ОНИ ЖЕ 41.1.1-17, ОНЪ ЖЕ 41.1.1-21, ОНИ ЖЕ 41.2.1-8, ОНЪ ЖЕ 42.1.1-16, ОНЪ ЖЕ 43.2.1- 25, ОНЪ ЖЕ 44.1.1-20, ОНЪ ЖЕ 44.2.2-3, ОНА ЖЕ 45.1.2-5, ОНЪ ЖЕ 45.2.1-4, ОНИ ЖЕ 46.1.2-11, ОНИ ЖЕ 47.2.2-8, ОНЪ ЖЕ 48.1.1-27, ОНЪ ЖЕ 48.1.2-21, ОНЪ ЖЕ 53.1.2-14, ОНЪ ЖЕ 53.1.2-21, ОНЪ ЖЕ 53.2.1-6, ОНЪ ЖЕ 53.2.2-16, ОНИ ЖЕ 54.2.1-3, ОНЪ ЖЕ 55.2.1-26, ОНЪ ЖЕ 56.1.1-26, ОНЪ ЖЕ 56.2.1-16, ОНЪ ЖЕ 58.1.1-28, ОНЪ ЖЕ 58.1.2-5, ОНЪ ЖЕ 58.1.2-16, ОНЪ ЖЕ 58.2.2-27, ОНЪ ЖЕ 58.2.2-31, ОНЪ ЖЕ 59.1.1-11, ОНЪ ЖЕ 59.2.1-8, ОНЪ ЖЕ 59.2.2-6, ОНЪ ЖЕ 62.2.1-28, ОНИ ЖЕ 63.2.2-3, ОНЪ ЖЕ 64.1.2-26,

№	N-грамма	F	F (w)...F (w <sub>2</sub> )	Dice	Адрес
19	БОГЪ НАШЪ	66	БОГЪ (1093) НАШЪ (555)	0.09	<p> <b>О</b>НЪ ЖЕ 64.2.1-12, <b>О</b>НЪ ЖЕ 64.2.2-5, <b>О</b>НЪ ЖЕ 65.2.1-3, <b>О</b>НЪ ЖЕ 65.2.1-13, <b>О</b>НЪ ЖЕ 70.2.2-7, <b>О</b>НЪ ЖЕ 70.2.2-20, <b>О</b>НЪ ЖЕ 71.1.1-15, <b>О</b>НН ЖЕ 75.1.1-18, <b>О</b>НЪ ЖЕ 77.2.1-13, <b>О</b>НЪ ЖЕ 77.2.2-2, <b>О</b>НН ЖЕ 78.1.2-24, <b>О</b>НА ЖЕ 79.1.1-1, <b>О</b>НА ЖЕ 79.1.1-16, <b>О</b>НН ЖЕ 80.1.1-20, <b>О</b>НА ЖЕ 80.2.2-14, <b>О</b>НН ЖЕ 87.2.2-29, <b>О</b>НЪ ЖЕ 88.1.1-15, <b>О</b>НЪ ЖЕ 89.1.1-10, <b>О</b>НЪ ЖЕ 91.2.2-21, <b>О</b>НН ЖЕ 100.1.1-18, <b>О</b>НЪ ЖЕ 104.1.1-17, <b>О</b>НН ЖЕ 107.1.1-15, <b>О</b>НН ЖЕ 108.1.1-16, <b>О</b>НН ЖЕ 108.1.2-1, <b>О</b>НН ЖЕ 111.1.2-24, <b>О</b>НЪ ЖЕ 112.1.1-9, <b>О</b>НИ ЖЕ 112.1.2-17, <b>О</b>НЪ ЖЕ 116.2.2-6, <b>О</b>НН ЖЕ 117.2.1-23, <b>О</b>НН ЖЕ 118.1.1-2, <b>О</b>НЪ ЖЕ 119.2.1-16, <b>О</b>НЪ ЖЕ 119.2.1-26, <b>О</b>НЪ ЖЕ 121.2.1-8, <b>О</b>НЪ ЖЕ 126.1.1-24, <b>О</b>НИ ЖЕ 126.1.1-26, <b>О</b>НЪ ЖЕ 126.1.1-28, <b>О</b>НЪ ЖЕ 128.1.1-25, <b>О</b>НИ ЖЕ 131.2.1-27, <b>О</b>НН ЖЕ 134.1.1-5, <b>О</b>НН ЖЕ 134.1.1-22, <b>О</b>НЪ ЖЕ 134.2.1-22, <b>О</b>НЪ ЖЕ 135.1.1-21, <b>О</b>НЪ ЖЕ 139.1.1-6, <b>О</b>НЪ ЖЕ 140.1.1-23, <b>О</b>НН ЖЕ 143.2.1-17, <b>О</b>НЪ ЖЕ 144.1.1-15, <b>О</b>НА ЖЕ 145.1.1-20, <b>О</b>НЪ ЖЕ 147.1.1-25, <b>О</b>НИ ЖЕ 150.1.1-2, <b>О</b>НА ЖЕ 153.1.1-14, <b>О</b>НА ЖЕ 153.1.1-15, <b>О</b>НЪ ЖЕ 153.1.1-26, <b>О</b>НА ЖЕ 153.2.1-25, <b>О</b>НИ ЖЕ 154.1.1-1, <b>О</b>НН ЖЕ 154.2.1-5, <b>О</b>НА ЖЕ 156.1.1-19, <b>О</b>НЪ ЖЕ 156.2.1-26                      Час: <b>Б</b>ЖЕ <b>НА</b>ШЪ 1.2.1-10, <b>Б</b>ЖЕ <b>НА</b>ШЪ 1.2.1-17, <b>Б</b>ЖЕ <b>НА</b>ШЪ 1.2.1-19, MS: <b>Б</b>Ъ <b>НА</b>ШЪ 2.2.1-7, АлтС: <b>Б</b>ЖЕ <b>НА</b>ШЪ 14.1.1-25, <b>Б</b>А <b>НА</b>ШЕГО 17.1.1-12, <b>Б</b>ЖЕ <b>НА</b>ШЪ 19.1.1-10, <b>Б</b>ЖЕ <b>НА</b>ШЪ 19.1.1-15, <b>Б</b>ЖЕ <b>НА</b>ШЪ 27.1.1-7, <b>Б</b>ЖЕ <b>НА</b>ШЪ 27.1.1-16, <b>Б</b>ЖЕ <b>НА</b>ШЪ 27.2.1-15, <b>Б</b>ЖЕ <b>НА</b>ШЪ 34.2.1-10, <b>Б</b>ЖЕ <b>НА</b>ШЪ 48.1.1-7, <b>Б</b>ЖЕ <b>НА</b>ШЪ 48.2.1-2, <b>Б</b>ЖЕ <b>НА</b>ШЪ 50.1.1-24, <b>Б</b>ЖЕ <b>НА</b>ШЪ 51.2.1-15, <b>Б</b>ЖЕ <b>НА</b>ШЪ 53.2.1-5, <b>Б</b>ЖЕ <b>НА</b>ШЪ 53.2.1-11, <b>Б</b>ЖЕ <b>НА</b>ШЪ 59.2.1-11, <b>Б</b>ЖЕ <b>НА</b>ШЪ 61.1.1-21, <b>Б</b>ЖЕ <b>НА</b>ШЪ 62.1.1-10, <b>Б</b>ЖЕ <b>НА</b>ШЪ 63.1.1-11, <b>Б</b>ЖЕ <b>НА</b>ШЪ 64.1.1-22, <b>Б</b>ЖЕ <b>НА</b>ШЪ 65.1.1-13, <b>Б</b>ЖЕ <b>НА</b>ШЪ 70.1.1-12, ПсДС: <b>Б</b>ЖЕ <b>НА</b>ШЪ 2.1.1-4, <b>Б</b>А <b>НА</b>ШЕГО 24.2.1-2, <b>Б</b>А <b>НА</b>ШЕГО 27.2.1-18, <b>Б</b>А <b>НА</b>ШЕГО 28.1.1-7, <b>Б</b>А <b>НА</b>ШЕГО 51.2.1-1, <b>Б</b>А <b>НА</b>ШЕГО 56.1.1-13, <b>НА</b>ШЪ <b>Б</b>Ъ 58.1.1-13, <b>НА</b>ШЪ <b>Б</b>Ъ 58.1.1-25, <b>Б</b>О <b>НА</b>ШЕМО 58.2.1-13, <b>Б</b>А <b>НА</b>ШЕГО 59.1.1-2, <b>Б</b>А <b>НА</b>ШЕГО 59.1.1-19, <b>Б</b>Ъ <b>НА</b>ШЪ 59.2.1-11, <b>Б</b>Ъ <b>НА</b>ШЪ 61.1.1-1, <b>Б</b>А <b>НА</b>ШЕГО 73.1.1-21, <b>Б</b>Ъ <b>НА</b>ШЪ 75.1.1-19, <b>НА</b>ШЪ <b>Б</b>Ъ 75.1.1-19, <b>Б</b>О <b>НА</b>ШЕМО 84.1.1-11, <b>Б</b>Ъ <b>НА</b>ШЪ 85.1.1-2, <b>Б</b>А <b>НА</b>ШЕГО 98.1.1-20, <b>Б</b>А <b>НА</b>ШЕГО 99.2.1-8, <b>Б</b>Ъ <b>НА</b>ШЪ 99.2.1-11, <b>Б</b>Ъ <b>НА</b>ШЪ 101.1.1-9, <b>Б</b>О <b>НА</b>ШЕМО 102.2.1-11, <b>Б</b>А <b>НА</b>ШЕГО 103.1.1-8, <b>Б</b>ЖЕ <b>НА</b>ШЪ 103.1.1-16, <b>Б</b>А <b>НА</b>ШЕГО 103.1.1-19, <b>Б</b>Ъ <b>НА</b>ШЪ 103.1.1-21, <b>Б</b>Ъ <b>НА</b>ШЪ 103.2.1-3, <b>Б</b>Ъ <b>НА</b>ШЪ 107.2.1-26, <b>Б</b>ЖЕ <b>НА</b>ШЪ 111.1.1-13, <b>Б</b>Ъ <b>НА</b>ШЪ 116.1.1-18, <b>Б</b>Ъ <b>НА</b>ШЪ 117.2.1-9, <b>Б</b>А <b>НА</b>ШЕГО 127.1.1-26, <b>Б</b>О <b>НА</b>ШЕМО 127.2.1-6, <b>Б</b>А <b>НА</b>ШЕГО 131.1.1-4,                 </p>

№	N-грамма	F	F(w)...F(w <sub>i</sub> )	Dice	Адрес
20	А ОУША МОН	102	А ОУША (244) МОН (2319)	0.08	<p>Б̄А НАШЕГО 131.1.1-12, Б̄Ъ НАШЪ 131.1.1-17, Б̄Ъ НАШЕМОУ 142.1.1-14, Б̄Ъ НАШЕМОУ 142.2.1-2, ЕВАС: Б̄Ъ НАШЪ 115.2.1-20, К/Л/ЕЖЕ НАШЪ 2.1.1-7</p> <p>ПсДС: А̄ШН МОН 3.2.1-18, А̄ША МОЕ 71.1.1-14, А̄ШЖ МОЖ 8.1.1-13, А̄ШЖ МОЮ 8.2.1-5, А̄ЮШН МОЕИ 15.1.1-12, А̄ШН МОЕИ 16.2.1-9, МОН А̄ШЖ 20.2.1-8, А̄ШЖ МОЖ 20.2.1-8, А̄ШЖ МОЖ 21.1.1-2, А̄ШЖ МОЖ 31.1.1-2, А̄ША МОЕ 31.2.1-17, А̄ШЖ МОЖ 32.1.1-9, А̄Ш МОЖ 33.2.1-9, А̄ШЖ МОЮ 35.2.1-7, А̄ША МОЕИ 36.1.1-12, А̄ШЖ МОЮ 39.1.1-3, А̄ШЖ МОЮ 40.1.1-4, А̄ША МОЕ 40.1.1-8, А̄ШЖ МОЮ 40.1.1-23, А̄ША МОЕ 43.2.1-4, А̄Ш МОЕИ 44.2.1-14, А̄Ш МОЕИ 45.1.1-1, А̄ША МОЕ 45.1.1-6, А̄Ш МОЕИ 45.1.1-16, А̄ШЖ МОЮ 45.1.1-19, А̄ШЖ МОЮ 45.2.1-9, А̄ША МОЕИ 52.1.1-18, А̄ША МОЕ 53.2.1-5, А̄ША МОЕ 53.2.1-6, А̄ШЖ МОЮ 53.2.1-13, А̄ШЕ МОЕ 53.2.1-18, А̄ША МОЕ 53.2.1-22, А̄ШЕ МОЕ 54.1.1-18, А̄ШЕ МОЕ 54.2.1-16, А̄ШЖ МОЮ 60.2.1-8, А̄ША МОЕИ 64.2.1-6, А̄Ш МОЕИ 64.2.1-9, А̄ШЖ МОЮ 65.2.1-13, А̄ШЖ МОЮ 66.1.1-24, А̄ШЖ МОЖ 66.2.1-15, А̄ША МОЕ 67.1.1-2, А̄ШЖ МОЮ 67.1.1-10, А̄ШЖ МОЮ 67.1.1-17, А̄ШЖ МОЮ 68.1.1-21, А̄ША МОЕ 70.1.1-13, А̄ША МОЕ 70.2.1-1, А̄ЮША МОЕ 71.1.1-15, А̄ША МОЕ 71.1.1-22, А̄ША МОЕИ 71.2.1-1, А̄ШЖ МОЮ 73.1.1-23, А̄Ш МОЕИ 73.2.1-17, А̄ША МОЕИ 76.1.1-5, А̄ШЖ МОЮ 76.2.1-4, А̄Ш МОЕИ 77.1.1-2, А̄ША МОЕ 77.1.1-8, А̄ША МОЕИ 78.1.1-6, А̄ШЖ МОЮ 78.2.1-19, А̄ШЖ МОЮ 78.2.1-25, А̄ША МОЕ 79.2.1-6, А̄ША МОЕ 84.1.1-23, А̄ША МОЕ 92.1.1-22, А̄ШЖ МОЖ 94.1.1-7, А̄ША МОЕИ 94.1.1-10, А̄ША МОЕ 94.2.1-15, А̄ЮШЖ МОЮ 97.1.1-20, А̄ЮШЖ МОЮ 100.2.1-14, А̄ША МОЕ 105.1.1-20, А̄ША МОЕ 105.1.1-22, А̄ША МОЕ 106.1.1-19, А̄ШЖ МОЮ 114.1.1-25, А̄ШЖ МОЮ 117.2.1-8, А̄ШЕ МОЕ 117.2.1-12, А̄ШЖ МОЮ 117.2.1-14, А̄ША МОЕ 120.1.1-3, А̄ША МОЕ 120.1.1-14, А̄ША МОЕ 120.1.1-20, А̄ША МОЕ 122.2.1-3, А̄ША МОЕ 123.2.1-10, А̄ША МОЕ 124.1.1-28, А̄ША МОЕ 125.2.1-25, А̄ША МОЕ 126.1.1-18, А̄ШЖ МОЮ 126.2.1-3, А̄ША МОЕ 126.2.1-15, А̄ША МОЕ 129.2.1-15, А̄ША МОЕ 134.2.1-15, А̄ША МОЕИ 137.1.1-6, А̄ША МОЕИ 137.1.1-24, А̄ШЖ МОЮ 137.2.1-5, А̄ШЖ МОЮ 137.2.1-15, МОИ А̄ША 137.2.1-24, А̄ША МОЕ 137.2.1-24, А̄ШЖ МОЮ 138.1.1-16, А̄Ш МОЕИ 138.1.1-19, А̄ША МОЕ 141.2.1-2, ЕВАС: А̄ШЖ МОЖ 23.1.1-26, А̄ША МОЕ 26.2.2-30, МОЕИ А̄ШЕ 60.1.1-25, А̄ША МОЕ 89.2.2-24, А̄ШЖ МОЖ 92.2.2-14, А̄ШЖ МОЖ 114.1.2-13, А̄ША МОЕ 130.1.1-6, А̄ША МОЕ 144.2.1-24</p>

№	N-грамма	F	F (w)...F (w <sub>x</sub> )	Dice	Адрес
21	реци господь	94	реци (1229) господь (1511)	0.07	ЛитС: реуе гъ 43.2.1-17, реуе гъ 43.2.1-21, реуе гъ, 44.1.1-12, ПсцД: Гъ реуе 3.1.1-1, рѣхъ гъ 18.2.1-11, рѣхъ гъ 45.1.1-8, рѣхъ гъ 52.2.1-23, реуе гъ 75.1.1-23, реуе гъ 98.2.1-1, реуе гъ 110.2.1-14, реуе гъ 114.2.1-20, гъ рѣхъ 121.2.1-1, рѣхъ гъ 135.2.1-14, гъ рѣхъ 137.1.1-25, ЕвАс: реуе гъ 8.1.2-12, реуе гъ 9.1.2-5, реуе гъ 9.2.2-7, реуе гъ 12.2.2-8, реуе гъ 13.1.1-23, реуе гъ 13.1.2-27, реуе гъ 13.2.2-6, реуе гъ 14.1.1-25, реуе гъ 15.1.1-13, реуе гъ 23.1.1-20, реуе гъ 23.2.1-24, реуе гъ 27.1.2-28, реуе гъ 29.1.1-2, реуе гъ 29.2.1-13, реуе гъ 30.1.2-25, реуе гъ 30.2.1-6, реуе гъ 30.2.1-16, реуе гъ 31.1.2-14, реуе гъ 32.2.2-2, реуе гъ 33.2.1-21, реуе гъ 34.1.1-25, реуе гъ 34.2.2- 10, реуе гъ 35.2.1-19, реуе гъ 36.1.1-27, реуе гъ 39.2.1-22, реуе гъ 44.1.2-11, реуе гъ 47.2.1-16, реуе гъ 49.1.2-26, реуе гъ 49.2.1-17, реуе гъ 51.2.2-3, реуе гъ 53.2.1-14, реуе гъ 55.1.1-3, реуе гъ 55.2.2- 13, реуе гъ 60.1.1-6, реуе гъ 60.1.2-20, гъ реуе 61.1.1-7, реуе гъ 61.1.2-7, реуе гъ 61.2.2-15, реуе гъ 62.1.2-24, реуе гъ 64.2.2-14, реуе гъ 65.2.1-27, реуе гъ 66.2.2-8, реуе гъ 67.1.2-3, реуе гъ 67.2.1-21, реуе гъ 68.1.1-22, реуе гъ 69.1.1-15, реуе гъ 69.2.2-5, реуе гъ 71.2.2-24, реуе гъ 73.1.1-3, реуе гъ 73.2.2-24, реуе гъ 76.2.1-22, реуе гъ 84.2.1-19, реуе гъ 87.1.2-9, реуе гъ 92.1.2-24, реуе гъ 114.1.1-2, реуе гъ 114.2.1-2, реуе гъ 115.1.1-2, реуе гъ 116.1.1-21, реуе гъ 116.2.1-6, реуе гъ 116.2.2-16, реуе гъ 116.2.2-22, реуе гъ 116.2.2-27, реуе гъ 117.1.1-4, реуе гъ 118.2.1-24, реуе гъ 120.1.1-24, реуе гъ 120.2.1-19, реуе гъ 121.1.1-21, реуе гъ 122.1.1-12, реуе гъ 122.2.1-20, реуе гъ 123.1.1-25, реуе гъ 124.1.1-5, реуе гъ 124.2.1-13, реуе гъ 126.1.1-11, реуе гъ 127.2.1-17, реуе гъ 128.2.1-4, реуе гъ 128.2.1-27, реуе гъ 141.2.1-10, реуе гъ 143.2.1-4, реуе гъ 151.1.1-13, реуе гъ 158.1.1-10 ЛитС: же реуе 43.2.1-19, ПсцД: же рѣхъ 39.1.1-11 и др.
22	же реци	123	же (2366) реци (1229)	0.07	

Приложение 10.

Биграммы трех текстов (*Dice*, леммы, свободный порядок)<sup>25</sup>

№	АсЕв		ПсДС		ЛитС	
	Н-грамма	Dice	Н-грамма	Dice	Н-грамма	Dice
1	първъ послѣдънь	0.26	ѹрьмень морѣ	0.31	правьда облещи	0.44
2	двьрь затворити	0.20	принестн даръ	0.29	азъ одръжатн	0.15
3	болѣти цѣлти	0.16	жъзлъ правостъ	0.24	ѹко богъ	0.05
4	коуръ възгласити	0.16	ѹростъ гнѣбатн	0.13	вьсь зъло	0.03
5	н̄ ѣ	0.14	соуктънь глаголатн	0.11	даръ господь	0.03
6	ктеръ ѹловѣкъ	0.13	оупъбатн на	0.11	дароватн мъи	0.03
7	мжжъ грѣшити	0.13	на земля	0.10	тъи о	0.02
8	правьдъникъ прнзъбатн	0.13	доуша мон	0.10	многость твон	0.02
9	сѣти сѣма	0.13	богъ нашъ	0.09	на нже	0.02
10	оухо нгълннъ	0.12	завѣтъ завѣшатн	0.09	тъи господь	0.02
11	ѣ н̄	0.11	непоруѣнь пжтъ	0.09	твон рабъ	0.02
12	рещи господь	0.11	на азъ	0.07	вьсь же	0.02
13	ѣ н̄	0.10	къ тѣи	0.07	образъ тѣи	0.02
14	онъ же	0.10	ѹко благо	0.07	въ н	0.02
15	доуша съпастн	0.09	проходъ ни	0.07	мъи вьсь	0.02
16	мъногъ народъ	0.09	повѣдѣти къ то	0.06	мъи отъ	0.02
17	н̄ з̄	0.08	отъ азъ	0.06	н сѣинъ	0.01
18	же рещи	0.08	благо же	0.06	въ вьсь	0.01
19	къ прнитн	0.08	тъ быти	0.06	въ твон	0.01
20	срьдъцѣ вашъ	0.08	на господннъ	0.05	съ вьсь	0.01

<sup>25</sup> Компоненты — леммы, количество компонентов — 2, расстояние между компонентами — 0, учет границ конструкций — «да», исключить служебные слова — «нет», закрепленность — «свободные», следование — «нет».

## Список источников и сокращений

- ЕвАс — Ассеманиево Евангелие. Ватиканская (Апостолическая) библиотека, Slav. 3 Glag., X/XI вв., 158 л. Подг. Р. Ф. Тублиова, В. А. Баранов (Ижевский государственный технический университет). [http://manuscripts.ru/mns/main?p\\_text=56673793](http://manuscripts.ru/mns/main?p_text=56673793), дата обращения 15.И.2017.
- ПсДС — Псалтырь Димитрия Синайского. Синай/Египет, Монастырь св. Екатерины, Cod. Sin. slav. 3/N, трет. четв. XI в., 145 л. Подг. А. Н. Соболев (РАН/ААН/Марбургский университет), М. МакРоберт (Оксфордский университет), М. Гау (Венский технический университет), под ред. Х. Микласа (Венский университет). <http://manuscripts.ru/mns/portal.main?p1=42>, дата обращения 15.И.2017.
- ЛитС — Литургиарий (ранее «Миссал»). Синай/Египет, Монастырь св. Екатерины, Cod. Sin. slav. 5/N, вт. четв. XI в., ок. 80 л. Подг. В. Садовски, Д. Хюрнер, Ф. Ваңдл, Х. Миклас, под ред. Х. Микласа (Венский университет). <http://manuscripts.ru/mns/portal.main?p1=67>, дата обращения 15.И.2017.
- МедЛ — Медицинские листки. Синай/Египет, Монастырь св. Екатерины, Cod. Sin. slav. 3/N, посл. четв. XI в., 2 л. Подг. Х. Миклас, Д. Хюрнер, под редакцией Х. Микласа (Венский университет). <http://manuscripts.ru/mns/portal.main?p1=59>, дата обращения 15.И.2017.
- МинС — Минея синайская. Синай/Египет, Монастырь св. Екатерины, Cod. Sin. slav. 4/N, посл. четв. XI в., 2 л. Подг. Х. Миклас, Д. Хюрнер (Венский университет), М. Гау (Венский технический университет), под ред. Х. Микласа (Венский университет). <http://manuscripts.ru/mns/portal.main?p1=48>, дата обращения 15.И.2017.
- КЛ — Киевские листки («Киевский миссал»). ЦНБ АН УССР, ДА/П. 328, X/XI в. и XII в., отрывок, 7 л. Подг. В. А. Баранов (Ижевский государственный технический университет). [http://manuscripts.ru/mns/main?p\\_text=61350133](http://manuscripts.ru/mns/main?p_text=61350133), дата обращения 15.И.2017.
- Час — Часослов, фрагмент. Синай/Египет, Монастырь св. Екатерины, первая четверть XI в., 1 л. Подг. Н. Глибетич (Еврейский университет, Иерусалим), под ред. Х. Микласа (Венский университет). <http://manuscripts.ru/mns/portal.main?p1=66>, дата обращения 15.И.2017.

## Список литературы

- Баранов, В. А., С. В. Дубовцев. 2010. Электронное критическое издание средневекового славянского текста: модель данных и визуализация лингвистических единиц. Интеллектуальные системы в производстве 1, *str.* 280–287.
- Баранов, В. А. 2011а. Корпус средневековых рукописей на портале «Манускрипт: славянское письменное наследие»: стандартные функции и новые возможности. Письменное наследие и современные информационные технологии. Ижевск: Удмуртия, 2011, *str.* 5–36.
- Баранов, В. А. 2011б. Software Tools and User Interfaces designed for Historical-Linguistic Purposes of Project «Manuscript». Информационный бюллетень Ассоциации «История и компьютер» 37. Петрозаводск, 2011, *str.* 7–14.
- Баранов, В. А. 2011в. Полнотекстовая коллекция славянских Евангелий проекта «Манускрипт» и специализированные инструменты разметки: модуль фрагментирования. Вестник Пермского университета 2 (16), *str.* 40–47. (Серия «История»).
- Баранов, В. А. 2012а. Электронные коллекции древнейших и средневековых славянских рукописей на портале «Манускрипт»: функциональные возможности. Синайский кодекс и памятники древней христианской письменности: традиции и инновации в современных исследованиях. Санкт-Петербург, 2012, *str.* 169–182.
- Баранов, В. А. 2012б. Лингвистические, методические и технологические вопросы создания и использования корпуса средневековых славянских текстов. Русистика: язык, культура, перевод. София: Изток-Запад, 2012, *str.* 404–414.
- Баранов, В. А., Р. А. Аникина. 2012а. Параллельный корпус русских летописей в Интернете: цели, задачи, технологическая основа, использование. Информационные технологии и письменное наследие. Петрозаводск–Ижевск, 2012, *str.* 12–18.
- Баранов, В. А., Р. А. Аникина. 2012б. Параллельный корпус русских летописей XIII–XV вв. в Интернете: инструментарий и методика лингвотекстологического анализа средневекового текста. Интеллектуальные системы в производстве 2 (20), *str.* 157–162.
- Баранов, В. А., С. В. Дубовцев. 2012. Модуль статистики информационно-аналитической системы «Манускрипт»: функции и демонстрация данных. Информационные технологии и письменное наследие. Петрозаводск–Ижевск, 2012, *str.* 23–26.
- Баранов, В. А., О. В. Зуга. 2012. Лингвистические и компьютерные основы создания корпуса Евангелий XI–XIII вв. Актуальные пробле-

- мы теории и методологии науки о языке. Санкт-Петербург: ЛГУ им. А. С. Пушкина, 2011, str. 136–138.
- Баранов, В. А. 2014. Организация поиска и демонстрации коллокаций в корпусе «Манускрипт». Проблемы истории, филологии, культуры 3 (45). Москва—Магнитогорск—Новосибирск, 2014, str. 275–277.
- Баранов, В. А. 2015. Исторический корпус как цель и инструмент корпусной палеославистики. *Scripta & e-Scripta* 14–15. Sofia: «Boyan Penev» Publishing Center; Institute of Literature, BAS, 2015, str. 39–62. <https://drive.google.com/file/d/0BwBejXXryRcROVQ4TnlpZFh6am8/view?usp=sharing>, дата обращения 15.И.2017.
- Баранов, В. А. 2016а. Модуль n-грамм исторического корпуса «Манускрипт»: структурные и лингвистические параметры. Научное наследие В. А. Богородицкого и современный вектор исследований Казанской лингвистической школы 1. Казань: Изд-во Казан. ун-та, 2016, str. 50–61. <https://drive.google.com/file/d/0BwBejXXryRcROHc5T3YxTENhLUk/view?usp=sharing>, дата обращения 15.И.2017.
- Баранов, В. А. 2016б. Опыт создания модуля n-грамм системы «Манускрипт» и оценки эффективности его использования для поиска коллокаций в корпусе М. В. Ломоносова. Интеллектуальные системы в производстве 4 (31), str. 124–131. <https://drive.google.com/file/d/0BwBejXXryRcRV3ZpemZKcC1CMTQ/view?usp=sharing>, дата обращения 15.И.2017.
- Кочеткова, Н. А. Статистические языковые методы. Коллокации и коллигации. *Cyberleninka.ru*. <http://cyberleninka.ru/article/n/statisticheskiye-yazykovye-metody-kollokatsii-i-kolligatsii>, дата обращения 15.И.2017.
- Пентус, М., Пиперски, А., А. Сорокин. Математические модели в лингвистике. Коллокации и их автоматическое определение: лекции. <https://goo.gl/NALvX4>, дата обращения 15.И.2017.
- Ягунова, Е. В., Л. М. Пивоварова. 2013. От коллокаций к конструкциям. Русский язык: конструкционные и лексико-семантические подходы. Санкт-Петербург, 2013. (*Acta Linguistica petropolitana: Труды Института лингвистических исследований РАН*). <https://goo.gl/tHeoR>, дата обращения 15.И.2017.
- Baranov, Victor A. 2010. Machine-Readable Linguistic Internet Resources as a Basis for Historical-Philological Studies. *Journal of Applied Mathematics, Statistics and Informatics* 6, № 2. Trnava: The University of SS. Cyril and Methodius, Faculty of Natural Sciences, 2010, str. 63–79.
- Evert, S. Association Measures. *Computational Approaches to Collocations*. <http://collocations.de/AM/index.html>, дата обращения 15.И.2017.

Alati i metode pretraživanja kolokacija u povijesnom korpusu  
»Manuskript«  
(na primjeru glagoljskih rukopisa)

Sažetak

U radu se prikazuje postupak pronalaženja kolokacija u povijesnom korpusu *Manuskript* (*manuscripts.ru*). N-gramski modul koji korisniku pruža velik broj parametara pretraživanja, statističku analizu i izdvajanje višerječnih sveza, dopušta eksperimentiranje s ciljem otkrivanja semantički i/ili gramatički utvrđenih oblika riječi u potkorpusima. Postupak pripreme upita i raščlamba rezultata primjenjuje se na potkorpusu glagoljskih rukopisa. Utvrđeno je postojanje brojnih višerječnih sveza s različitim stupnjem okamenjenosti.

Tools and methods of searching for collocations in the historical corpus "Manuscript"  
(on the example of Glagolitic manuscripts)

Abstract

The article deals with the procedure of searching for collocations in the historical corpus "Manuscript" (*manuscripts.ru*). The n-gram module of the corpus that provides the user with a great number of parameters of searching, statistical analysis and output of many-component combinations permits experiments with the aim of revelation of semantically and/or grammatically established complexes of word forms in sub-corpora. The procedure of preparation of queries and analysis of results is applied to the sub-corpus of Glagolitic written records. Many-component combinations with various degrees of stability are revealed.

Ключевые слова: исторический текстовый корпус, коллокации, глаголические рукописи, n-граммы

Ključne riječi: povijesni tekstovni korpus, kolokacije, glagoljski rukopisi, n-gramski modul

Keywords: historical text corpus, collocation, Glagolitic manuscript, n-gram module<sup>26</sup>

<sup>26</sup> Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (РФФИ) в рамках проекта «Лингвостатистический анализ однокомпонентных и многокомпонентных лексических единиц исторического корпуса "Манускрипт"» (проект № 18/012/00463).

