

Жолобов Олег Феофанович

Казанский федеральный университет,
Россия, 420008, Казань, ул. Кремлевская, 18
ozolobov@mail.ru

Баранов Виктор Аркадьевич

Ижевский государственный технический университет им. М. Т. Калашникова,
Россия, 426069, Ижевск, ул. Студенческая, 7
victor.a.baranov@gmail.ru

Дистрибутивно-квантитативные и семантические характеристики глаголов знания в старославянской и древнерусской письменности*

Для цитирования: Жолобов О. Ф., Баранов В. А. Дистрибутивно-квантитативные и семантические характеристики глаголов знания в старославянской и древнерусской письменности. *Вестник Санкт-Петербургского университета. Язык и литература.* 2021, 18 (1): 56–76. <https://doi.org/10.21638/spbu09.2021.104>

В статье представлен первый опыт дистрибутивно-квантитативного анализа лексико-семантического ряда в древнерусской письменности на материале трех разножанровых подкорпусов исторического корпуса «Манускрипт» (manuscripts.ru) — списков Евангелий, миней и летописей. Предпринято соотнесение программной обработки диахронического корпуса данных и их историко-лингвистического статуса: рассмотрены смысловые отношения в лексико-семантическом ряду глаголов знания *въдѣти*, *въдати*, *знати* в древнерусском языке, на месте которого в русском языке выступает один нейтральный глагол *знати*, так что встает вопрос об эволюционной динамике лексической группы. Установлено, что весь ряд принадлежит исходной лексической системе, притом что глагол *въдати* отсутствовал в старославянских рукописях и получил распространение в древнерусских источниках как разговорного, так и книжного регистров. Выявленные квантитативные биграмм-показатели позволили установить контрастные коллокаты при глаголах *въдѣти* и *въдати*, с одной стороны, и при глаголе *знати* — с другой. Среди коллокатов глагола *знати* не оказалось отвлеченных существительных, а также коллокатов, присоединяющих пропозиции. Обнаружены дистрибутивные участки распространения синонимии в глагольном ряду, которые отразили начало диахронической конкуренции в цепи лексем. В качестве метода оценки близости компонентов биграмм (анализируемых глаголов и их коллокатов) использована статистическая мера T-score, а для нахождения степени соответствия дистрибутивных рядов друг другу — коэффициент корреляции r-Спирмена.

Ключевые слова: старославянская и древнерусская письменность, лексико-семантический ряд глаголов знания, n-граммы, дистрибутивно-квантитативный анализ.

* Исследование выполнено при поддержке Российского научного фонда (РНФ), проект № 20-18-00206.

...Вычленение лексических классов как конструкторов, созданных самим языком в ходе его истории, оказывается описанием его собственного строения, разграничением органических участков языковой системы.

Русский семантический словарь

Введение

Одному современному глаголу *знать* 'иметь сведения; обладать знанием чего-н., иметь о чем-н., ком-н. понятие, представление' (см. [Ожегов 1988: 190]) в древнерусском языке соответствовал лексико-семантический ряд — три лексические единицы: *въдѣти, въдати, знати*¹. В связи с этим требуют ответа следующие вопросы: какое место занимают эти три глагола в исходной системе, каковы их лингвостатистический статус и узואльное распределение в древнерусский период? Ответить на них можно в ходе дистрибутивно-квантитативного анализа большого массива данных.

Одним из источников данных для количественно-статистического анализа средневековых славянских рукописей является исторический корпус «Манускрипт» (manuscripts.ru). Корпус содержит транскрипции славянских текстов X–XV веков общим объемом более 3,5 млн словоупотреблений, имеет мета-, аналитическую и лингвистическую разметку, снабжен корпусным менеджером, обеспечивающим подготовку транскрипций, лемматизацию текстов, формирование подкорпусов, выборку лингвистических единиц и их демонстрацию (о корпусе см., напр., [Baranov 2018; Баранов 2019]). Специализированный модуль n-грамм позволяет извлечь из сформированного подкорпуса сочетание из n компонентов на основе нескольких параметров, включающих, в частности, возможность указания количества компонентов, расстояния между ними, положения компонентов относительно друг друга, количественной или статистической меры оценки n-граммы, буквенной маски компонентов (о модуле см., напр., [Baranov, Gnutikov 2019]).

Извлечение из больших массивов текстов n-грамм и их статистическая оценка являются необходимыми процедурами для решения различных прикладных и исследовательских задач в области искусственного интеллекта, автоматического поиска и извлечения информации и перевода, анализа тональности и модальности текстов, компьютерного аннотирования и реферирования, электронной лексикографии и терминографии, систематизации и кластеризации текстов, автороведческой и судебной экспертизы и др.

Разработано большое количество статистических мер, обеспечивающих нахождение в текстах сочетаний, компоненты которых, тяготея друг к другу, встречаются вместе неслучайно, то есть чаще, чем при случайной (свободной, статистически ожидаемой) сочетаемости [Evert 2005; Manning, Schütze 2000: 151–228], показана их специфика [Браславский, Соколов 2006]; дана классификация методов, подходов,

¹ Глагол *ведати* 'знать' в современном русском языке признается устаревшим (см., напр., [Ожегов 1988: 60]). В словаре это значение иллюстрируется примером евангельского происхождения: *Не ведает, что творит*; ср.: 'знать, отдавать себе отчет' не *въдѣтъ* бо са что *творѣтъ* оу *уа*р *оѣдаси* L 23,34 Zogr Mar As Ostr [SJS I: 375].

приемов извлечения [Бобкова 2015: 94–99]; проанализированы лингвистические основания разграничения свободных конструкций и в разной степени устойчивых сочетаний; регулярно дается характеристика разработанного инструментария и алгоритмов для извлечения коллокаций [Митрофанова, Соколова 2017]; активно применяются грамматические [Захаров, Хохлова 2014], лексико-семантические [Митрофанова и др. 2008], семантические [Kutuzov et al. 2018], тематические [Масевич, Захаров 2016] характеристики компонентов *n*-грамм для решения различных теоретических и прикладных задач и т. д.²

Свойство статистических методов исследования *n*-грамм извлекать из больших массивов текстовых данных сочетания, претендующие на устойчивость, позволяет использовать их и для анализа узуальных характеристик слов и их синтагматического окружения, репрезентирующих семантику лексем.

Целью работы является сопоставительное исследование семантики трех древнерусских синонимичных глаголов *въдѣти*, *въдати*, *знати* на основе сведений об их сочетаемости и количественно-статистических характеристиках. Для этого исследованы сочетаемость глаголов, состав пересекающихся и непересекающихся перечней их лексического окружения и степень статистической близости этих перечней.

Данные извлечены с помощью модуля *n*-грамм из трех разножанровых подкорпусов средневековых рукописей: подкорпуса Евангелий X–XIV веков (13 кодексов и отрывков, 599 438 текстовых форм), подкорпуса служебных миней XI–XIV вв. (17 списков и отрывков, 392 735 текстовых форм) и подкорпуса ранних летописей (три списка, 349 913 текстовых форм) (см. список источников).

В качестве материала использованы сочетания глаголов с левым и правым соседями (коллокатами) (без учета их места по отношению к глаголу)³, статистическая близость глагола и коллоката оценивалась с помощью меры *T-score*, позволяющей оценить степень связанности (близости) компонентов биграмм с анализируемыми глаголами [Evert 2005; Кочеткова 2013; Ягунова, Пивоварова 2013] и одновременно уменьшить влияние на оценку биграмм повторяющихся в первую очередь в списках Евангелий и летописей контекстов:

$$T - score = \frac{F(w_1, w_2) - \frac{F(w_1) \times F(w_2)}{N}}{\sqrt{F(w_1, w_2)}}$$

где $F(w_1)$ — частота первого компонента,
 $F(w_2)$ — частота второго компонента,
 $F(w_1, w_2)$ — частота сочетания $w_1 w_2$,
 N — общее число биграмм в корпусе.

² Следует сказать, что в последнее время появились лингвостатистические работы, выполненные и на материале средневековых славянских текстов (см., напр., [Баранов, Жолобов 2020]).

³ Параметры запросной формы: тип компонентов — словоформа, количество компонентов — два, следование компонентов — свободное, статистическая мера — *T-score*, границы конструкций — не учитываются, маски — регулярные выражения *въдд.**, *въ[д]^[^д] | [смвж].** (форма *ведд* отнесена к формам глагола *ведати*; омонимия формы *ведь* не снята), *з.?нд[^м].?* (извлечены только бесприставочные глаголы), точность коллоката — устранили титла и диакритика, форма вывода: кириллица современная.

Сопоставление перечней биграмм осуществлено в двух направлениях: сравнивались коллокаты одного и того же глагола в разных подкорпусах; коллокаты с разными глаголами в одном подкорпусе. Использовались перечни коллокатов со значением соответствующей биграммы.

Для сопоставления рядов совпадающих коллокатов была использована непараметрическая статистика Спирмена⁴:

$$r_s = 1 - \frac{6 \sum_{i=1}^n (x_i - y_i)^2}{n(n^2 - 1)},$$

где x_i, y_i — значения элементов вариативных (сопоставляемых) рядов (перечней), n — количество сопоставляемых членов рядов [Непараметрические корреляции].

Для приведения биграмм первичной выборки на основе текстовых форм к одному сочетанию использовано два приема: приведение форм коллокатов к леммам в современной транслитерации (буква *ъ* сохраняется), а также вычисление их среднего значения T-score для всех сочетаний, содержащих коллокаты с разными формами одной леммы.

Количественно-статистическая дистрибуция глаголов *въдати*, *въдѣти*, *знати*

Общие количественные данные

Количество биграмм с каждым из глаголов показано в таблице 1⁵.

Таблица 1. Количество биграмм с глаголами *въдати*, *въдѣти*, *знати* в трех подкорпусах

	Евангелия		Миней		Летописи		Всего	
	Абс.	%	Абс.	%	Абс.	%	Абс.	%
<i>въдати</i>	72	0,012	3	0,001	145	0,041	220	0,018
<i>въдѣти</i>	3257	0,544	472	0,120	726	0,208	4455	0,291
<i>знати</i>	274	0,046	15	0,004	9	0,003	298	0,018
Всего	3603	0,201	490	0,042	880	0,084	4973	0,109

Различия между частотностью употребления глаголов в подкорпусах значительны. Наиболее частотным является глагол *въдѣти* в Евангелиях (0,544), единично встретился глагол *въдати* в минеях. В то же время соотношения частотности глаголов в подкорпусах аналогичны: во всех преобладает глагол *въдѣти*, глаголы *въдати* и *знати* значительно уступают ему в употреблении, при этом в Евангелиях чаще встречается глагол *знати*, в летописях — *въдати*.

⁴ Выбор метода обусловлен тем, что значения сопоставляемых рядов коллокатов не во всех случаях соответствуют закону нормального распределения.

⁵ В разделе примеры даются в современной транслитерации.

Соотношение глаголов *вѣдати*, *вѣдѣти*, *знати* между корпусами

Извлеченные биграммы исследовались по следующему плану:

- 1) коллокаты приведены к леммам;
- 2) за значение расстояния коллоката и анализируемого глагола принято значение T-score биграммы;
- 3) вычислены средние значения T-score для каждого коллоката;
- 4) выявлены совпадения перечней коллокатов одного глагола в разных подкорпусах, а также трех анализируемых глаголов в одном подкорпусе;
- 5) для каждого сопоставления вычислена корреляция перечней;
- 6) показаны части перечней коллокатов, различающие дистрибуцию одного глагола в разных подкорпусах и разных глаголов в одном подкорпусе.

Коллокаты глагола *вѣдати*. Лексическое окружение глагола *вѣдати* в Евангелиях включает 26, в минеях — 3, в летописях — 58 знаменательных и незнаменательных слов. Максимальное среднее значение T-score в Евангелиях имеют биграммы с частицей *бо* — 2,222, леммами знаменательных слов являются *господь*, *глаголати*, *въровати*, *дарѣ*, *путь*, *мощи*, *почърпнути*, *дѣяние*, *лѣнивѣ*, *реци*, *таина*, *хръстось*, *богѣ*, *дати*, *пророкъ* (1,823–0,998); в минеях коллокат *единѣ* имеет значение T-score, равное 1,0, *ты* — 2,0, *себе* — ниже 1,0; в летописях максимальное среднее значение имеют биграммы с местоимением *мы* — 1,486, максимальное среднее значение T-score имеют следующие десять коллокатов — знаменательных слов: *богѣ*, *сѣтворити*, *гадати*, *кѣзнь*, *галичанинѣ*, *лѣсть*, *милосѣрдие*, *невѣръствие*, *сѣбѣрати*, *испѣрва* (1,240–1,0).

Совпадающие в подкорпусах коллокаты немногочисленны:

- Евангелия — минеи: *ты*;
- Евангелия — летописи: *азѣ*, *бо*, *богѣ*, *быти*, *дати*, *же*, *и*, *иже*, *реци*, *сѣ*, *ты*, *что*, *яко* — всего 13 совпадений;
- минеи — летописи: *ты*, *себе*.

Корреляция коллокатов глагола *вѣдати* между Евангелиями и летописями представлена в таблице 2⁶.

Таблица 2. Корреляция между коллокатами биграмм с глаголом *вѣдати*

Подкорпус	Евангелия	<i>p-value</i>	Минеи	<i>p-value</i>	Летописи	<i>p-value</i>
Евангелия	1,0	—	—	—	0,478	0,099
Минеи	—	—	1,0	—	—	—
Летописи	0,478	0,099	—	—	1,0	—

Статистическая корреляция тринадцати совпадающих коллокатов Евангелий и летописей — 0,478 (по шкале Чеддока оценивается как средняя); при этом значение *p* несколько выше 0,05, что уменьшает значимость обнаруженной корреляции.

⁶ В связи с единичностью совпадающих коллокатов найти корреляцию между Евангелиями и минеями, а также между летописями и минеями не представляется возможным.

Приведем коллокааты со значением T-score более 1,0, которые составляют уникальную часть лексического окружения глагола *въдѣти* в подкорпусах: Евангелия — *господь, глаголати, въровати, даръ, путь, мощи*; минеи — отсутствуют; летописи — *мы, не, самъ, сътворити, тъ*.

Коллокааты глагола *въдѣти*. Биграаммы подкорпуса Евангелий содержат 148, миней — 189, летописей — 153 уникальных коллокаата, включающих как знаменательные, так и незнаменательные слова. В Евангелиях максимальное среднее значение T-score имеют биграаммы с существительным *заповѣдь* — 3,734; следующими знаменательными словами с наибольшим средним значением T-score являются *ити, господинъ, дати, марфа, мощи, единъ, даръ, законъ, учитель, ненавидѣти* (3,674–1,913). В минеях максимальное среднее значение имеет существительное *дѣтель* — 1,999, следующие девять знаменательных слов с наибольшим значением — *спасение, славити, богъ, реши, богословъ, божьствънъ, съдѣяти, сѣрдьчънъ, устроити* (1,366–1,0). В летописях максимальное среднее значение имеет местоимение *иже* — 3,064; первые десять знаменательных коллокаатов — *приити, изяславъ, владыка, епархъ, святославъ, законъ, истъление, старьць, стрьи, рюриковичъ* (2,410–1,398).

Пересечения перечней коллокаатов следующие:

- Евангелия — минеи: *азъ, аще, бо, богъ, быти, въдѣти, въ, вы, въсь, глаголати, гласъ, господь, да, дело, дѣнь, единъ, же, законъ, заповѣдь, и, и_мест, иже, истина, исусъ, камо, къгда, кѣде, кѣто, ли, лукавъ, мужъ, мы, не, ни, ничьтоже, нъ, о, онъ, отъвъщати, отъкуду, приити, пророкъ, ради, реши, самъ, свои, слово, слышати, съ, сѣрдьце, себе, таина, тъ, ты, убо, хрьстось, црьсарьство, чловькъ, чьто, яко, якоже* — всего 61 совпадающий коллокат (из 148 в Евангелиях и из 189 в минеях);
- Евангелия — летописи: *азъ, аще, бо, богъ, быти, въдѣти, воля, въ, въсь, глаголати, гласъ, дати, единъ, же, жидовьскъ, законъ, и, и_мест, иже, ити, камо, коньць, къ, къгда, кѣде, кѣто, ли, мощи, мужъ, мы, не, ни, никътоже, ничьто, ничьтоже, нъ, о, онъ, отъ, помышление, посълати, приити, пророкъ, реши, самъ, свои, съ, съ, себе, таина, твои, тъ, ты, чьто, яко, якоже* — всего 56 совпадающих коллокаатов (из 148 в Евангелиях и из 153 в летописях);
- минеи — летописи: *а, азъ, аще, бо, богъ, божии, быти, въдѣти, владыка, въ, въсь, глаголати, гласъ, дружина, единъ, же, законъ, и, и_мест, иже, како, камо, къгда, кѣде, кѣнязь, кѣто, ли, мужъ, мы, мысль, не, ни, ничьтоже, нъ, о, онъ, прежде, приити, принести, прияти, пророкъ, реши, самъ, свои, слава, съ, себе, таина, тъ, ты, уже, чьто, яко, якоже* — всего 54 совпадения (из 189 в минеях и из 153 в летописях).

Статистическая корреляция между совпадающими членами рядов коллокаатов глагола *въдѣти* показана в таблице 3.

Таблица 3. Корреляция между коллокаатами биграам с глаголом *въдѣти*

Подкорпус	Евангелия	<i>p-value</i>	Минеи	<i>p-value</i>	Летописи	<i>p-value</i>
Евангелия	1,0	—	0,191	0,141	0,198	0,144
Минеи	0,191	0,141	1,0	—	0,364	0,007
Летописи	0,198	0,144	0,364	0,007	1,0	—

Несмотря на значительное количество совпадений, между коллокатами в подкорпусах миней и летописей существует умеренная (средняя) корреляция (0,364), при этом значение $p = 0,007$ свидетельствует о высокой значимости результата. В подкорпусах Евангелий и миней и Евангелий и летописей — слабая (0,191 и 0,198 соответственно), при значении p выше 0,05.

Приведем перечни коллокатов со значением более 1,0, которые составляют специфику окружения глагола *въдѣти* в каждом из подкорпусов.

- Евангелия (из 148): *аминь, блаженъ, въровати, господинъ, грясти, даръ, жена, изити, или, иоанъ, искати, кѣнигы, кьи, льнивѣ, листовие, марфа, мои, нѣкъто, ненавидѣти, ныня, отць, пасха, писание, подобати, почърпнути, прѣдати, прозябнути, путь, рабъ, равви, слуга, слушати, трепетати, тѣже, тѣкъмо, ученикъ, учитель* (38 коллокатов);
- миней (из 189): *дѣтель, съпасение, славити* (3 коллоката),
- летописи (из 153): *болгаринъ, данилъ, епархъ, из, изяславъ, истъление, льсть, мѣстиславъ, никѣто, новѣгородъ, печенъгъ, поити, рюриковичъ, святославъ, сице, старъць, стръи, тѣгда, ярополкъ, яти* (20 коллокатов).

Коллокаты глагола *знати*. Биграмы подкорпуса Евангелий содержат 17, миней — 12, летописей — 9 уникальных коллокатов. В Евангелиях максимальное среднее значение T-score имеют биграмы с частицей *не* — 3,942; знаменательные слова — *чловѣкъ, отць, добръ* (2,190–0,991); в минеях максимальное среднее значение также имеют биграмы с частицей *не* — 1,556, коллокаты — знаменательные слова — *сѣврстѣникъ, чудьнъ, пещера, прѣподобьнъ, пастухъ* (1,0); в летописях — биграмы с существительным *плодъ* — 1,414, две другие знаменательные леммы — *грѣхъ, рана* (1,0).

Совпадения единичны:

- Евангелия — миней: *быти, и, не*;
- Евангелия — летописи: *азъ, быти, и_мест, не*,
- миней — летописи: *быти, не, ни*.

Статистические отношения между несколькими совпадающими членами рядов коллокатов глагола *знати* приведены в таблице 4.

Таблица 4. Корреляция между коллокатами биграмм с глаголом *знати*

Подкорпус	Евангелия	<i>p-value</i>	Миней	<i>p-value</i>	Летописи	<i>p-value</i>
Евангелия	1,0	—	1,000	—	0,2	0,800
Миней	1,000	—	1,0	—	0,5	0,667
Летописи	0,2	0,800	0,5	0,667	1,0	—

Высокие статистические значения корреляции коллокатов глагола *знати* между Евангелиями и минеями (1,0) объясняются тем, что очень частотные незначительные слова и глагол *быти* имеют аналогичную степень тесноты (близости) по отношению к *знати*. Корреляция коллокатов Евангелий и летописей, а также миней и летописей (0,2 и 0,5 соответственно) малоинформативна в связи с высоким значением *p-value*.

Отличия дистрибуции глагола *знати* в подкорпусах незначительны: Евангелия — *вы, мои, отъць, чловѣкъ* (T-score > 1,0), *мы, николиже* (T-score = 1,0), *добрѣ, же, отсель, тѣ, убо, яко* (T-score < 1,0); минеи — *пастухѣ, пещера, прѣподобьнѣ, развѣ, съвѣрствьникѣ, чудьнѣ* (T-score = 1,0), *нѣ, твои* (T-score < 1,0); летописи — *плодѣ* (T-score > 1,0), *грѣхѣ, рана* (T-score = 1,0), *а* (T-score < 1,0).

Дистрибуция глаголов *вѣдати*, *вѣдѣти*, *знати* в каждом из подкорпусов

Покажем корреляцию между коллокатами исследуемых глаголов в каждом из подкорпусов.

Подкорпус Евангелий. Пересечения рядов коллокатов при глаголах *вѣдати*, *вѣдѣти*, *знати* в подкорпусе Евангелий следующие:

- *вѣдати* — *вѣдѣти*: *азѣ, бо, богѣ, быти, вѣровати, глаголати, господѣ, дарѣ, дати, же, и, и_мест, иже, лѣнивѣ, мощи, почѣрпнути, пророкѣ, путь, реци, сѣ, таина, ты, хрѣстосѣ, чѣто, яко* — всего 25 совпадений (из 26 коллокатов глагола *вѣдати* и из 148 глагола *вѣдѣти*);
- *вѣдати* — *знати*: *азѣ, быти, же, и, и_мест, яко* — всего 6 совпадений (из 26 коллокатов глагола *вѣдати* и из 17 коллокатов глагола *знати*);
- *вѣдѣти* — *знати*: *азѣ, быти, вы, же, и, и_мест, мои, мы, не, отъць, тѣ, убо, чловѣкъ, яко* — всего 14 совпадений (из 148 коллокатов глагола *вѣдѣти* и 17 коллокатов глагола *знати*).

Покажем корреляцию дистрибутивного окружения этих глаголов в подкорпусе Евангелий (см. табл. 5).

Таблица 5. Корреляция между коллокатами биграмм с глаголами *вѣдати*, *вѣдѣти*, *знати* в Евангелиях

Глагол	<i>вѣдати</i>	<i>p-value</i>	<i>вѣдѣти</i>	<i>p-value</i>	<i>знати</i>	<i>p-value</i>
<i>вѣдати</i>	1,0	—	0,044	0,835	0,714	0,111
<i>вѣдѣти</i>	0,044	0,835	1,0	—	0,081	0,782
<i>знати</i>	0,714	0,111	0,081	0,782	1,0	—

Сильная корреляция в подкорпусе Евангелий обнаружена между коррелятами глаголов *вѣдати* и *знати* (0,714) (*p-value* = 0,111, что уменьшает значимость найденной корреляции), отсутствие корреляции — между глаголами *вѣдѣти* и *знати* (0,081), *вѣдати* и *вѣдѣти* (0,044) (*p-value* в обоих случаях подтверждает отсутствие корреляции). Обратим внимание на то, что практически все (25 из 26) корреляты глагола *вѣдати* являются и коррелятами глагола *вѣдѣти*, при этом корреляция на основе значения T-score между ними отсутствует, что говорит о том, что, несмотря на полное совпадение, расстояния между каждым конкретным коллокатом и глаголами *вѣдати* и *вѣдѣти* существенно различаются.

Для каждого из глаголов приведем перечни коллокатов со значением T-score > 1,0, которые отличают каждый из глаголов от других:

- *въдати* (из 26) — отсутствуют;
- *въдѣти* (из 148) — *аминь, аще, блаженъ, воля, въсь, гласъ, господинъ, грясти, да, днь, единъ, жена, жидовъскъ, законъ, заповѣдь, изити, или, иоанъ, иска-ти, исусъ, ити, камо, къде, кѣнигы, кѣто, кыи, ли, листвие, марфа, нѣкъто, ненавидѣти, ни, ничѣто, ничѣтоже, нѣ, ныня, онъ, отъвъщати, отъкуду, пасха, писание, подобати, помышление, посълати, прѣдати, прозябнути, рабъ, равви, самъ, свои, слуга, слушати, сърдѣце, трепетати, тѣже, тѣкъмо, ученикъ, учитель, якоже* — всего 60 коллокатов;
- *знати* (из 17) — отсутствуют.

Подкорпус миней. Совпадения коллокатов при глаголах *въдати*, *въдѣти*, *знати* в подкорпусе миней следующие:

- *въдати* — *въдѣти*: *единъ, себе, ты;*
- *въдати* — *знати*: отсутствуют;
- *въдѣти* — *знати*: *быти, и, не, ни, нѣ, пастухъ, развъ* — всего 7 совпадений.

Покажем корреляцию коллокатов двух пар глаголов *въдати* — *въдѣти*, *въдѣти* — *знати* (табл. 6).

Таблица 6. Корреляция между коллокатами биограмм с глаголами *въдати*, *въдѣти*, *знати* в минеях

Глагол	<i>въдати</i>	<i>p-value</i>	<i>въдѣти</i>	<i>p-value</i>	<i>знати</i>	<i>p-value</i>
<i>въдати</i>	1,0	—	0,500	0,666	—	—
<i>въдѣти</i>	0,500	0,666	1,0	—	0,893	0,007
<i>знати</i>	—	—	0,893	0,007	1,0	—

Обнаружена сильная корреляция коллокатов глаголов *въдѣти* и *знати* (0,893) и средняя — *въдати* и *въдѣти* (0,500, *p-value* = 0,666). В первом случае значимость результата подтверждается значением *p* (0,007) и, возможно, объясняется вхождением в совпадающую часть перечней коллокатов в первую очередь служебных слов, местоимений, глагола *быти*, аналогично соотносящихся с анализируемыми глаголами по причине своей частотности.

Покажем перечни коллокатов со значением T-score > 1,0, которые отличают каждый из глаголов от других:

- *въдати* (из 3) — отсутствуют;
- *въдѣти* (из 189) — *азъ, бо, богъ, вы, да, дѣтель, кѣто, ли, мы, рещи, самъ, славити, съпасение, тѣ, чѣто, яко, якоже* — всего 17 коллокатов;
- *знати* (из 12) — отсутствуют.

Подкорпус летописей. Совпадения коллокатов при глаголах *въдати*, *въдѣти*, *знати* в подкорпусе летописей следующие:

- *въдати* — *въдѣти*: *а, азъ, бо, богъ, братъ, быти, владимиръ, въ, въсь, градъ, дати, же, и, иже, како, камо, кѣзнь, кѣто, льсть, милосърдие, мы, мысль, на, не, ни, ничѣто, нѣ, о, рещи, самъ, себе, съ, твои, тѣ, тѣгда, ты, умрети, чѣто, яко* — всего 39 совпадений (из 58 коллокатов глагола *въдати* и из 153 коллокатов глагола *въдѣти*);

- *въдати* — *знати*: *а, азъ, быти, не, ни* — всего 5 совпадений (из 58 коллокатов глагола *въдати* и из 9 коллокатов глагола *знати*);
- *въдѣти* — *знати*: *а, азъ, быти, и мест, не, ни* — всего 6 совпадений (из 153 коллокатов глагола *въдѣти* и из 9 коллокатов глагола *знати*).

Покажем корреляцию коллокатов этих трех глаголов (табл. 7).

Таблица 7. Корреляция между коллокатами биграмм с глаголами *въдати*, *въдѣти*, *знати* в летописях

Глагол	<i>въдати</i>	<i>p-value</i>	<i>въдѣти</i>	<i>p-value</i>	<i>знати</i>	<i>p-value</i>
<i>въдати</i>	1,0	—	0,423	0,007	0,700	0,188
<i>въдѣти</i>	0,423	0,007	1,0	—	0,886	0,019
<i>знати</i>	0,700	0,188	0,886	0,019	1,0	—

Сильная (высокая) корреляция существует между коллокатами глаголов *въдѣти* и *знати* (0,886) и *въдати* и *знати* (0,700), средняя — между коллокатами глаголов *въдати* и *въдѣти* (0,423). При этом для последней пары значимость найденного результата высокая (*p-value* = 0,007), для двух первых — невысокая (0,019 и 0,188 соответственно). Следует учесть, что высокая корреляция установлена на основе 5 и 6 случаев совпадений служебных слов, местоимений, глагола *быти*.

Покажем перечни коллокатов (*T-score* > 1,0), которые отличают каждый из глаголов от других:

- *въдати* (из 58) — *сътворити*;
- *въдѣти* (из 153) — *болгаринъ, владыка, глаголати, данилъ, единъ, епархъ, законъ, из, изяславъ, истъление, къ, къде, кнѣзь, ли, мьстиславъ, никъто, ничьтоже, новъгородъ, онъ, отъ, печенъгъ, поити, посылати, приити, рюриковичь, святославъ, сице, слава, старьць, стрьи, уже, ярополкъ, яти* — всего 33 коллоката;
- *знати* (из 9) — *плодъ*.

Основные итоги количественно-статистического анализа

Абсолютное количество биграмм с этими глаголами в подкорпусах существенно различно: от 3 с *въдати* в минеях до 3257 с *въдѣти* в Евангелиях.

Анализ совпадений в перечнях коллокатов каждого из глаголов в подкорпусах также демонстрирует существенные различия — от единичных случаев до половины и практически всех коллокатов. Так, коллокаты глагола *въдати* в Евангелиях и минеях совпали лишь в одном случае, в Евангелиях и летописях — в 13 (что для Евангелий составляет 50 % всех коллокатов).

Существенным является обнаружение практически полного совпадения коллокатов глагола *въдати* с коллокатами глагола *въдѣти* в Евангелиях (из 26 коллокатов глагола *въдати* 25 встречаются в окружении глагола *въдѣти*), коллокатов глагола *знати* и глагола *въдѣти* (из 17 коллокатов глагола *знати* 14 используются при глаголе *въдѣти*) и др.

В тех случаях, когда выявлено достаточное количество совпадений, вычислено статистическое соответствие (корреляция r_s -Спирмена) между рядами биграмм, которое позволило оценить похожесть и контрастность пересекающейся части коллокатов не только по количеству совпадений, но и по отношению коллоката к глаголу — насколько похожа теснота анализируемых глаголов и их коллокатов в разных подкорпусах.

Следует отметить, что при одинаковом или близком количестве совпадений между рядами коллокатов коэффициенты корреляции между ними достаточно близки. Например, корреляция пар *въдати* — *знати* в Евангелиях, *въдѣти* и *знати* в минеях, *въдѣти* и *знати* в летописях при 6 или 7 пересечениях практически равны — 0,714, 0,893 и 0,886 соответственно. Это позволяет считать и количество пересечений, и значение корреляции в равной степени показательными и надежными величинами.

В то же время обнаружено, что в случае вычислений корреляции на основе единичных совпадений, которые включают в первую очередь служебные слова, величины корреляции положительные и выше или значительно выше, чем при нахождении корреляции между перечнями, в которых совпадает большое количество коллокатов. Это объясняется тем, что величины T-score биграммы, включающей коллокат — частотное служебное слово, аналогичны в случае разных глаголов и в разных подкорпусах.

Опыт лингвистической интерпретации биграмм с глаголами знания

Общая характеристика глаголов

Глаголы *въдѣти* и *знати* имеют индоевропейское происхождение и являются общеславянскими (см. [Фасмер, т. 1: 283; т. 2: 100–101]).

Оба глагола представлены в «Старославянском словаре», опирающемся только на собственно старославянские рукописи X–XI вв.: *въдѣти* насчитывает свыше 800 употреблений, а *знати* — более 100 (см. [Старославянский словарь 1999: 164, 238–239]). Перечни греческих глаголов, которые переводятся славянскими лексемами, имеют пересечения: εἰδέναι, γι(γ)νώσκειν, притом что оба греческих глагола, как и славянские глаголы, являются продолжением индоевропейских лексем и соотносительны с индоевропейскими корнями **woid-* и **ǵnō-*. Славянская глагольная пара не воспроизводит полностью соотношение двух названных греческих глаголов, поскольку обе греческих лексемы могут передаваться каждым из двух славянских глаголов, а это доказывает по крайней мере латентное смысловое тождество глаголов. Тем не менее взаимная мена глаголов, свидетельствующая о синонимии, в старославянском выражена, хотя в евангельских чтениях такая мена представлена в словарных материалах лишь в одном случае: и овца по немъ идѣтъ · ѿко знаѣтъ гласъ его оѿдаси J 10,4 As (въдатъ Zogr Mar Ostr) [SJS, т. I: 680].

Особая архаичность глагола *въдѣти* находит выражение в атематическом спряжении: 1 л. ед. ч. *въмь*, 2 л. ед. ч. *въси*, 3 л. ед. ч. *въ(ств)ъ*⁷, 1 л. дв. ч. *въвъ*, 2–3 л.

⁷ См. реликты данной формы в выражении *Бог весть*, а также во фразеологизмах *невесь какой*, *невесь сколько*, *невесь что*.

въста, 1 л. мн. ч. *въмъ*, 2 л. мн. ч. *въсте*, 3 л. мн. ч. *въда(ть)*⁸. Еще более ярко архаический характер глагола *въдѣти* проявляется в сохранении индоевропейской формы медиального перфекта *въдѣ* ‘я видел’ > ‘я знаю’ [Фасмер, т. 1: 284–285], которая в древнерусском языке устойчиво использовалась вместо ожидаемой атематической формы *въмь*⁹. Исходный средний залог формы *въдѣ* актуализирует семантику внутреннего состояния агенса, поскольку категория медиальности замыкает действие на самом субъекте. Глагол *въдѣти* характеризует субъект действия с внутренней стороны его осведомленности о той или иной ситуации. Этот атематический глагол известен берестяным грамотам, где форма 1 л. ед. ч. наблюдается только в виде архаизма *въдѣ*: *а тобѣ вѣдѣ ако естъ не годьнъ* — «А тебе, я знаю, что не любо» (НГБ 752¹⁰, 1080–1100-е гг.); *а а вѣде [т. е. вѣдѣ] ожь · жь васъ есте тъвдръ ольскы(нъ)* — «Я же знаю, что у вас товар Олексы» (НГБ 548, сер. 1150-х — нач. 1210-х гг.); *а про села сама веси* — «а про села сама знаешь» (НГБ 944, 1160–1210-е гг.).

Действие, обозначаемое глаголом *знати*, характеризует субъект действия иначе — в его активной деятельности и внешней направленности на узнавание и опознание объекта по присущим ему признакам: *како ма знаеши* (Ин. 1, 49 [SJS, т. I: 680]) (см. показательную модификацию этого значения — ‘быть в интимных отношениях’: *и не знааше ея · дондеже роди сѣбъ свои прѣвѣнець* Мт 1, 25 Ас Сав Боян [Старославянский словарь 1999: 238]).

Приставочные образования *по-въдѣти* и *по-знати* обнаруживают значимое смысловое отличие исходных глаголов: если в первом случае значение производящего существенно видоизменяется, выражая коммуницирующее и информирующее направления деятельности (‘рассказать, сообщить’, ‘объявить’, ‘показать, засвидетельствовать’), то во втором — незначительно модифицируется, получая в том числе акспекуальный характер (‘познать, постичь’, ‘узнать, опознать’). Таким образом, глагол *въдѣти*, обозначая изначально свидетельское знание ‘видеть и знать’ (= ‘быть свидетелем’), получил естественное развитие в актуализации значения передачи информации при образовании глагола *повѣдѣти*. Ср. также существительное *въсть* < **wěd-tv* ‘весть, известие’, связанное с тем же корнем.

В старославянском языке даже при расширенном понимании состава его рукописных источников отсутствует отвлеченное существительное при глаголе *знати*, в отличие от глагола *въдѣти*, где такое производное имеется: *знание* ‘знакомый, знакомые, друзья’ vs *въдѣние* ‘знание, познание, ведение’ [SJS, т. I: 374, 680]. См. также: *знакъ* ‘знакомый’ [Срезневский 1893: 987]¹¹ vs *въдѣ* ‘знание, ведение, познание’ [SJS, т. I: 374]. Производные глагола *знати* обозначают объект узнавания, а не сам процесс.

Глагол *въдѣти* занимает ведущие позиции в древнерусской письменности. В картотеке тезаурусного «Словаря древнерусского языка (XI–XIV вв.)» зафиксировано свыше 1000 употреблений различных форм глагола *въдѣти* (см. [СДРЯ, т. II:

⁸ О вариативности форм 3 л. см. [Жолобов 2014].

⁹ О вариативности форм 1 л. ед. ч. глагола *въдѣти* см. [Žolobov 2016: 308–310]. Реликтом формы 1 л. ед. ч. *въдѣ* в русском языке является частица *ведь*. Ср. также императив *вижь* по атематическому образцу *въжь*.

¹⁰ Здесь и далее НГБ (+ номер грамоты) — Новгородские грамоты на бересте, цит. по [Зализняк 2004].

¹¹ Единственный пример приводится по древнерусскому списку XI–XII вв. древнеболгарского перевода Оглашений Кирилла Иерусалимского.

290–291]), в то время как у глагола *знати* насчитывается 202 примера (см. [СДРЯ, т. III: 399–400]). В рассмотренных трех подкорпусах содержится, как показано было в квантитативном описании, 4455 биграмм с глаголом *въдѣти* и 298 с глаголом *знати*¹² (см. первый раздел настоящей работы).

Анализ двух десятков наиболее типичных контекстов в ряде древнерусских оригинальных источников показал, что глагол *въдѣти* обозначал, как правило, знание ситуации, которое выражалось в предпочтительном управлении пропозицией, конструкциями с двойным винительным или отвлеченными существительными, в то время как глагол *знати* обычно обозначал действие, направленное на конкретный объект для его опознания по внешним признакам [Птенцова 2008]. Схождение в типах употребления глаголов знания в этих источниках оказалось чрезвычайно редким¹³.

Распространение глагола въдати в древнерусской письменности

В классическом старославянском отсутствует глагол *въдати* (см. [Старославянский словарь 1999]). В пражском словаре старославянского языка, включающем данные поздних списков старославянских переводов, этот глагол приводится (см. [SJS, т. I: 374]). Все три контекста с данным глаголом, однако, принадлежат двум позднедревнерусским и одному среднерусскому спискам древнечешских церковнославянских текстов, поэтому могут оцениваться как древнерусская инновация¹⁴.

Распространение глагола *въдати* по модели *знати* на месте *въдѣти* обусловлено морфологическим переходом в продуктивный третий класс: 1 л. ед. ч. *въдаю*, 2 л. ед. ч. *въдаеши*, 3 л. ед. ч. *въдае(ть)*, 1 л. дв. ч. *въдаевъ*, 2–3 л. дв. ч. *въдаета*, 1 л. мн. ч. *въдаемъ*, 2 л. мн. ч. *въдаете*, 3 л. мн. ч. *въдаю(ть)*. Проанализированный выше материал подкорпусов доказывает, что распространение глагола *въдати* является древнерусской особенностью. Так, если в старославянских Евангелиях этот глагол отсутствует, то в проанализированном подкорпусе древнерусских списков Евангелия встречается в составе 72 биграмм с 26 уникальными коллокатами, что составляет существенную величину (см. первый раздел настоящей работы). Глагол *въдати* — полный синоним глагола *въдѣти*, поскольку является лишь его морфологической модификацией. Анализ биграмм подкорпуса Евангелий подтверждает этот факт: коллокаты глагола *въдати* полностью совпадают с коллокатами глагола *въдѣти*, что в данном случае указывает на замену в сочетаниях атематического глагола на тематический. См., например, мену глаголов в биграммах с коллокатом *бо: въдаше бо · яко зависти ради предаша юго* [АЕ, л. 106 об.¹⁵] vs *въдаше бо · яко льсти*

¹² Омонимия субстантивной и глагольной форм *въсть* в некоторой степени повлияла при подсчетах на общие количественные характеристики, вместе с тем являясь характерным структурно-семантическим признаком глагола.

¹³ В описание были включены отдельные контексты из Сказания о Борисе и Глебе, Жития Феодосия Печерского, берестяных грамот и двух летописей — Лаврентьевской и Ипатьевской.

¹⁴ В старославянском вместе с тем существовала аналогичная пара производных глаголов с приставкой *по-*, выражавшая видовое противопоставление: *повѣдѣти* ‘рассказать’, ‘разгласить’, ‘показать’ vs *повѣдати* ‘рассказывать, сообщать’, ‘провозглашать’, ‘показывать, свидетельствовать’ (см. SJS, т. II: 74–76). В то же время в самом старославянском глагол *повѣдати* является новым: он лишь один раз встречается в Евангелиях, однако в Мариинском четвероевангелии его нет — только *повѣдѣти*.

¹⁵ Где 106 — номер листа, 2 — об. листа.

ради предаша него [АЕ, л. 117]; с коллокатом *яко*: вѣвѣ ꙗко съ естъ снѣ наю [ПЕ, л. 23] vs вѣдѣвѣ ꙗко съ естъ снѣ наю [ТЕ, л. 154] и под.

В подкорпус миней эти замены почти не проникли: здесь отмечено лишь три случая, и все они имеют коллокаты, пересекающиеся с коллокатами глагола *въдѣти* (см. первый раздел настоящей работы).

В летописях количество употреблений глагола *въдати* стремительно растет (145 биграмм). Более чем вдвое по сравнению с евангелиями увеличивается и количество уникальных коллокатов — их 58. Из этих коллокатов довольно много — 39 — те же, что и у глагола *въдѣти*. Такие количественные данные свидетельствуют, с одной стороны, о полной синонимии глаголов, а с другой — о независимом функционировании глагола *въдати*. В целом ряде случаев вместо глагола *въдѣти* в Лаврентьевской летописи выступает глагол *въдати* в Ипатьевской летописи. См., например, мену глаголов с коллокатом *како*: *вѣна же рекоста вѣ вѣвѣ ꙗко естъ члѣвкъ створенъ* [ЛЛ, л. 59 об.] vs *вѣна же рекоста ꙗко естъ створенъ члѣвкъ* [ИЛ, л. 65 об.]; с коллокатами *бо* и *милосърдие*: *мъи во вѣмъи мърдые ѿрополче* [ЛЛ, л. 101 об.] vs *мъи во вѣдѣемъ мърдие ѿрополче* [ИЛ, л. 112] и под.

В материалах берестяных грамот по изданию А. А. Зализняка [Зализняк 2004: 723] именно глагол *въдати* полностью господствует: здесь зафиксировано 13 случаев употребления разных форм глагола *въдати*, а также пять случаев с производным *въдатиса* ‘рассчитываться’ при трех примерах с глаголом *въдѣти* и одним — *знати*. Количественное соотношение глаголов *въдѣти* и *въдати*, таким образом, в книжном и разговорном языке было противоположным, а вот глагол *знати* занимал одинаковое периферийное положение. Функционирование глаголов *въдѣти*, *въдати* и *знати* не связано с оппозицией сакрального и профанного, вопреки бытующим предположениям (об этом см. также [Птенцова 2008: 265–267])¹⁶.

Семантико-синтаксические особенности глаголов вѣдѣти, вѣдати и знати

По данным СДРЯ [СДРЯ, т. II: 288, 290–291], семантические структуры глаголов *въдѣти* и *въдати* почти идентичны, отличаясь одним — третьим — компонентом: 1) ‘знать что-л.’ (с оттенками: а) ‘знать кого-л.’, б) ‘признавать кого-л., что-л.’); 2) ‘ведать, распоряжаться, управлять чем-л.’; 3) в сочетаниях с инфинитивом ‘знать, мочь, быть в состоянии, производить, произвести действие, выраженное инфинитивом’¹⁷.

¹⁶ В статье А. Л. Соломоновской [Соломоновская 2014: 121–122] рассматривается распределение синонимичных глаголов *вѣдѣти*, *разѣмѣти*, *знати* в сербско-церковнославянском переводе Ареопagitик. Глагол *знати* имеет лишь единичное употребление (4 примера), из чего заключается, что он может быть связан с обозначением профанного знания. Разграничение семантики глаголов *вѣдѣти*, *разѣмѣти* в переводе греч. ἴσχυωστω автор соотносит с противопоставлением статического и динамического аспектов знания. Любопытно, что первый глагол предпочитается, когда сохраняется связь с его этимологической природой, отсылающей к значению знания-видения, т. е. свидетельского знания.

¹⁷ Для глагола *въдѣти*, кроме того, в словаре отмечаются отдельные случаи употребления вместо глагола с родственным корнем — *видѣти*. Такое же явление отмечено для этого глагола и в старославянском (см. [SJS, т. I: 375]). Паронимическая близость данных глаголов питала тенденцию замещения глагола *въдѣти* синонимами.

Глагол *въдѣти* не просто преобладает количественно в подкорпусах Евангелий, миней и летописей, но и имеет наиболее разнообразные синтагматические связи и наибольшее число коллокатов, словарный состав которых был впервые установлен (см. первый раздел нашей работы). Вместе с тем количественно-статистический анализ дистрибуции показал, что в подкорпусах имеется довольно объемное ядро общих коллокатов, состав которого теперь стал известен: для Евангелий и летописей — 56 коллокатов, для Евангелий и миней — 61 коллокат, для миней и летописей — 54 коллоката. Коллокаты, как можно убедиться, отражают не только лексические валентности глагола *въдѣти*, но и его типовые синтаксические позиции, а также ассоциативные связи, когда коллокаты отражают лишь контекстуальную рядоположенность.

В рассмотренной базе данных глаголы *въдѣти* и *въдати* фиксируются в первом значении, указанном выше. Отличия биграмм данных глаголов от биграмм глагола *знати* демонстрируют существенные различия в семантико-синтаксических характеристиках.

Тождество значений глаголов *въдѣти* и *знати* отчетливее всего выступает в Евангелии, где оба глагола представлены в одном контексте с одинаковым управлением существительным *гласъ*: и овца по нѣмь идѹть · яко вѣдати глѣ него · по чюжемъ же не идѣть · нъ вѣжати ѿ него · яко не знають чюждиихъ гласа [ТЕ, л. 155]. В тождественных контекстах в двух разных чтениях — [Ин. 8, 19] и [Ин. 14, 7] — находятся глаголы *въдѣти* и *знати*: ѿвѣща їсѣ ни мене вѣсте ни ѿца моего · аще ма бзисте вѣдали · и ѿца моего оубо бзисте вѣдали [СЕ, л. 12] vs никтоже придѣть къ ѿцю тѣкмо мною · аще ма бзисте знали · и ѿца моего оубо бзисте знали [СЕ, л. 19]. Семантическая равноценность глаголов проявляется в замене глагола *знати* на глагол *въдати*, который, в отличие от глагола *знати*, не имеет дистрибутивных ограничений в употреблении: никтоже придѣть къ ѿцю тѣкмо мною аще ма бзисте вѣдали и ѿца моего знали бзи (sic!) оубо [МЕ, л. 26]. Глаголы *въдѣти* и *знати* варьируются в тождественных по смыслу контекстах с одинаковыми коллокатами, в которых, однако, с глаголом *въдѣти* коррелируют дополнительные предикации: и пакы ѿвѣржеса съ клатвою · яко не знаю чловѣка [ТЕ, л. 55] vs Онъ же начатъ ротитиса и клатиса · яко не вѣмъ члѣка сеѿо негоже глѣте [ТЕ, л. 89]. Однако в кратком апракосе расширяющая предикация отсутствует и эти последние чтения полностью уравниваются: тѣгда начатъ ротитиса и клати · ако не знаю члѣка [ПогЕ, л. 159] vs и пакы ѿвѣржеса съ клатвою · ако не вѣдѣ члѣка [ПогЕ, л. 158 об.]. Безусловно, эти дистрибутивные пересечения способствовали закреплению семантических схождений глаголов, поскольку постоянно воспроизводились в чтениях на богослужении.

Анализ показал, что 14 коллокатов из 17 у глагола *знати* в подкорпусе Евангелий являются общими для глагола *въдѣти* (раздел 1.3.1), притом что последний связан со 148 коллокатами. Это подтверждает абсолютное господство глагола *въдѣти*. Именно он не имеет дистрибутивных ограничений и способен выражать всю гамму оттенков в контекстах с глаголами знания. Разграничение между глаголами наиболее ярко проявляется не только в количестве коллокатов, но и в их качественном составе: среди коллокатов глагола *знати* отсутствуют постпозиционные подчинительные союзы и союзные слова, вводящие зависимые предикации. См. примеры подобного употребления глаголов *въдѣти*, *въдати* в дополнение к летописным примерам, приведенным в [Птенцова 2008]: с коллокатом *яко*: вѣмъ яко ѿ

бѣ пришьлъ ѳси [МЕ, л. 4]; с коллокатом *иже*: си вѣдѣть ѳже рекоуѣ азъ [МЕ, л. 150 об.]; с коллокатом *откуду*: мене вѣсте и вѣсте ѳкоуду ѳсмь [МЕ, л. 15 об.]; с коллокатом *камо*: не вѣмъ камо идеши [МЕ, л. 144 об.]; с коллокатом *что*: не вѣстѣса что просѣца [МЕ, л. 40]; с коллокатом *кто*: никтоже не вѣсть кто ѳсть сѣъ тѣкмо оцѣ [МЕ, л. 75 об.]; с коллокатом *кѣи*: не вѣстѣса конего дѣа ѳста взи [МЕ, л. 73 об.]; с коллокатом *кѣда*: не вѣсте кѣда гѣ въ домъ придетъ [СЕ, л. 101]; с коллокатом *кѣде*: не вѣдѣ кѣдѣ и положиша [СЕ, л. 166 об.] и мн. др. Отсутствует у глагола *знати* правосторонний коллокат *бо* — энклитика, указывающая на первое место глагола во фразе (см.: вѣдаше во искони ѳъ кто соуть невѣроуѣцеи [МЕ, л. 14] и под.).

В отличие от глаголов *вѣдѣти*, *вѣдати*, коллокаты, называющие объекты при глаголе *знати*, целиком ограничены обозначениями определенных персонажей, отвлеченные и даже собственно предметные существительные здесь отсутствуют: николиже знахъ васъ · отидѣте ѳ мене · дѣлающеи беззаконие [ТЕ, л. 14]; Тѣгда начѣтъ рѣтитисѣ и клатисѣ ѳко не знаю чѣка · и авие коуръ възгласи [ТЕ, л. 55]; ѳвѣржесѣ глѣа · не знаю него жено [ТЕ, л. 132]; не съ ли ѳсть ѳ · сѣъ ѳсифовъ · ѳемѣже мзи знаемъ оца и мѣръ [ТЕ, л. 148] и т. п. Однако за стандартным употреблением глагола *знати* в целом ряде случаев скрывается совершенно отличный от привычного смысл, поскольку типичное распределение коллокатов сопровождается контексты богословского, вероисповедного содержания: Вся мнѣ рѣдана сѣтъ оцѣмъ моимъ · и никтоже не знаеть сѣа тѣкмо оцѣ · Ни оца кто знаеть тѣкмо сѣъ · и ѳемѣже волить сѣъ ѳкрѣтити [ТЕ, л. 21 об.]; да знають тебе ѳдиногѣ истиннагѣ бѣ · и негоже посѣла ѳ хѣ [ТЕ, л. 165 об.] и др. В подобного рода контекстах, таким образом, глагол *знати* освобождается от исходной привязки к обозначению действия ‘узнавать, опознавать кого-л., приобретая отвлеченный характер.

Периферийное положение глагола *знати* в летописях, выразившееся в его низкой частотности, в биграмах имеет стандартные пересекающиеся коллокаты с глаголами *вѣдѣти*, *вѣдати*: общими здесь стали более половины коллокатов — 6 и 5 из 9 (см. первый раздел настоящей работы). В летописных текстах изредка наблюдаются нестандартные коллокаты при глаголе *знати*, воспринятые из книжных источников: ти во суть волци а не овца · ѳже достоинъ ѳ плода знати ѳа · хранитисѣ ихъ [ЛЛ, л. 9 об.]; и рещи баше давѣдскѣ азъ знаю грѣ мои предо мною ѳсть вончѣ [ЛЛ 83, л. об.]. В последнем примере глагол *знати* приобретает значение ‘сознавать’. См. также расширенный семантический объем глагола *знати* в контексте без новых коллокатов: тѣи во ѳси вѣ нѣ развѣе тебе иногѣ не знаѣ [ЛЛ, л. 155].

Расширение семантического объема глагола *знати* в Евангелиях, постоянно бывших на слуху, при сохранении типичной коллокации и углубление синонимии с глаголами *вѣдѣти* и *вѣдати* создавало почву для конкурентных преимуществ первого глагола в диахронической перспективе. В СДРЯ [СДРЯ, т. III: 399] в двух списках XIV в. фиксируется отвлеченное значение у отглагольного существительного *знание* ‘представление о чем-л., знание’, которое отсутствовало в старославянском.

Выводы

В ходе предпринятого анализа и описания были выявлены коллокаты лексико-семантического ряда глаголов знания в древнерусском языке — *вѣдѣти*, *вѣдати*,

знати — на базе трех объемных подкорпусов — Евангелий, миней и летописей. Глаголы первоначально семантически расходились. Если глагол *въдѣти* обозначал внутреннюю характеристику агенса — его свидетельскую осведомленность о ситуации или событии, то глаголом *знати* агенс характеризуется с внешней стороны в его направленности на узнавание и опознание объекта по внешним признакам. Выражая знание о ситуации, глагол *въдѣти* зачастую управляет пропозицией, тогда как при глаголе *знати* употребляются обозначения конкретного лица (включая анафорические замены), а подчинение новых предикаций не наблюдается. Связь глагола *въдѣти* со значением свидетельского знания проявилась в словообразовании в образовании глагола свидетельствования и сообщения *повѣдѣти*.

Установлено, что сам факт распространения глагола *въдати* в подкорпусе Евангелий является древнерусской инновацией, а подкорпус летописей фиксирует рост частотности и расширение базы коллокатов при этом глаголе. Единство базы коллокатов при глаголах *въдѣти* и *въдати* в подкорпусе Евангелий указывает на вторичность лексических замен и семантическое тождество глаголов.

Количественные характеристики биграмм с глаголом *знати* фиксируют его периферийное положение в древнеславянском узусе. Качественный состав биграмм с глаголами *въдѣти*, *въдати* и *знати*, наряду с общими коллокатами, имеет контрастные черты. Так, в биграмах с глаголом *знати* среди правосторонних коллокатов полностью отсутствуют союзы и союзные слова, которые характеризуют биграмы с глаголами *въдѣти* и *въдати*. Качественный состав биграмм, таким образом, непосредственно отражает указанные выше семантические различия в глагольном ряду *въдѣти*, *въдати* и *знати*. Тем не менее обнаружен заметный уровень общности коллокации при глаголе *знати* и глаголах *въдѣти* и *въдати*, свидетельствующий о динамике их семантического сближения в подкорпусе Евангелий. Кроме того, установлено, что в Евангелиях типичные биграмы с глаголом *знати* претерпевают семантическую трансформацию, поскольку глагол *знати* в вероисповедных контекстах освобождается от исходного значения узнавания или опознания кого-либо по внешним признакам, сближаясь в этом отношении с глаголом *въдѣти*, не имевшим дистрибутивных ограничений и обладавшим более широким семантическим диапазоном. Евангельские тексты, которые были постоянно на слуху, сыграли значительную роль в закреплении и углублении синонимии глаголов *въдѣти*, *въдати* и *знати*.

Помимо впервые установленного словаря самих биграмм и коллокатов при глаголах *въдѣти*, *въдати* и *знати* и его контрастных характеристик, были выявлены статистически высокоранговые коллокаты с показателем T-score > 1,0, которые фиксируют образование устойчивых синтагматических единств и синтаксических позиций.

Источники

Славянские Евангелия — Славянские Евангелия корпуса «Манускрипт». <http://manuscripts.ru/mns/portal.main?p1=9> (дата обращения: 01.08.2020): Ватиканская библиотека, Slav. 3 Glag.; РГАДА. Ф. 381. № 1, 14; РНБ. Ф. п.1.5, 1.13, 1.58; Погод. 11, Соф. 1; РГБ. М. 1666; Рум. 104, 105; ГИМ. Син. 1203; ЦБАН ЛитССР, F19-1.

АЕ — Евангелие апракос краткий («Архангельское евангелие»), 1092 г. (РГБ. М. 1666). 178 л.

ПЕ — Евангелие апракос полный («Пантелеймоново евангелие»), XII–XIII в. (РНБ. Соф. 1). 224 л.

ТЕ — Евангелие тетр («Типографское евангелие»), XII в. (РГАДА. Ф. 381 (Син. тип.). № 1). 193 л.
СЕ — Евангелие апракос полный («Симоновское евангелие»), 1270 г. (РГБ. Рум. 105). 167 л.
МЕ — Евангелие апракос полный («Музейное евангелие»), кон. XII — нач. XIII (?) в. (РГБ. Рум. 104). 158 л.
ПоГЕ — Евангелие апракос краткий («Погодинское евангелие»), XI–XII или XII–XIII в. (РНБ. Погод. 11). 264 л.

Славянские минеи — Славянские минеи корпуса «Манускрипт». http://manuscripts.ru/mns/portal.main?p1=10&p_lid=1 (дата обращения: 01.08.2020): НБВ. Собр. Зам. № 21; Люб. Cod. Кор. 2; РНБ. Q. п.72; РГАДА. Ф. 381. № 84, 89, 91; MSC, Cod. Sin. slav. 4/N; РНБ. Соф. 199, 202, 203, 204, F. п. 1.36, Q. п. 1.25, Q. п. 1.40, ОЛДП, Q. 180; РГАДА. Ф. 381. № 110; ГИМ. Син. 166, 895; БАН. Алекс.-Свирск. 37.

Летописи корпуса «Манускрипт» — Летописи корпуса «Манускрипт». http://manuscripts.ru/mns/portal.main?p1=12&p_lid=1 (дата обращения: 01.08.2020): РНБ. F. п. IV. 2; БАН. 16.4.4; 34.5.30.

ЛЛ — Лаврентьевская летопись, 1377 г. (РНБ. F. п. IV. 2). 173 л.

ИЛ — Ипатьевская летопись, пер. пол. (сер.?) XV в. (Ипатьевский список. БАН. 16.4.4). 307 л.

НГБ (+ номер грамоты) — Новгородские грамоты на бересте [Зализняк 2004].

Словари

Ожегов 1988 — Ожегов С. И. *Словарь русского языка*. Под ред. Н. Ю. Шведовой. М.: Русский язык, 1988. 750 с.

РСС 1998 — *Русский семантический словарь. Толковый словарь, систематизированный по классам слов и значений*. Т. 1. М.: Азбуковник, 1998. XXV, 807 с.

СДРЯ — *Словарь древнерусского языка (XI–XIV вв.)*. Т. I–XII. М., 1988–2019.

Срезневский 1893 — Срезневский И. И. *Материалы для словаря древнерусского языка*. Т. I. СПб.: Отд.-ние рус. яз. и словесности Имп. акад. наук, 1893.

Старославянский словарь 1999 — *Старославянский словарь (по рукописям X–XI веков)*. М.: Русский язык, 1999. 842 с.

Фасмер 1986 — Фасмер М. *Этимологический словарь русского языка*. В 4 т. Трубочев О. Н. (пер. с нем., доп.), Ларин Б. А. (ред., предисл.). Т. 1: А–Д. М.: Прогресс, 1986. 573 с. Т. 2: (Е–Муж). 671, [1] с.

SJS I–IV — *Slovník jazyka staroslověnského*. Т. I–IV. Praha: Nakladatelství Československé Akademie věd, 1956–1997.

Литература

Баранов 2019 — Баранов В. А. Создание и использование исторических корпусов славянских письменных памятников. *Scripta & e-Scripta*. 2019, 19: 33–57.

Баранов, Жолобов 2020 — Баранов В. А., Жолобов О. Ф. Лингвостатистическое исследование частотных слов в Словах Кирилла Туровского (по рукописи РНБ, F. п. I. 39). *Slověne*. 2020, 9 (1): 29–80. <https://doi.org/10.31168/2305-6754.2020.9.1.2>.

Бобкова 2015 — Бобкова Т. Извлечение коллокаций из корпуса украинских текстов. *Research Journal Studies about Languages*. 2015, (27): 93–105.

Браславский, Соколов 2006 — Браславский П., Соколов Е. Сравнение четырех методов автоматического извлечения двухсловных терминов из текста. В кн.: *Компьютерная лингвистика и интеллектуальные технологии*. М.: Изд.-во Рос. гос. гумм. ун-та, 2006. <https://clck.ru/RzcG6> (дата обращения: 01.08.2020).

Жолобов 2014 — Жолобов О. О рефлексх *ti-* и *t-* форм глаголов в древнерусском языке. *Russian Linguistics*. 2014, 38 (1): 121–163.

Зализняк 2004 — Зализняк А. А. *Древненовгородский диалект*. М.: Языки славянской культуры, 2004. 872 с.

- Захаров, Хохлова 2014 — Захаров В. П., Хохлова М. В. Автоматическое выявление терминологических словосочетаний. *Структурная и прикладная лингвистика*. 2014, (10): 182–200.
- Кочеткова 2013 — Кочеткова Н. А. Статистические языковые методы. Коллокации и коллигации. *Новые информационные технологии в автоматизированных системах*. 2013, (16): 301–305. <https://clck.ru/RzcNe> (дата обращения: 13.06.2020).
- Масевич, Захаров 2016 — Масевич А. Ц., Захаров В. П. Методы корпусной лингвистики в исторических и культурологических исследованиях. В сб.: *Компьютерная лингвистика и вычислительные онтологии*. СПб., 2016. С. 24–43. <https://clck.ru/RzcWC> (дата обращения: 01.08.2020).
- Митрофанова и др. 2008 — Митрофанова О. А., Белик В. В., Кадина В. В. Корпусное исследование сочетаемостных предпочтений частотных лексем русского языка. В сб.: *Компьютерная лингвистика и интеллектуальные технологии*. М., 2008. <https://clck.ru/RzcYr> (дата обращения: 01.08.2020).
- Митрофанова, Соколова 2017 — Митрофанова О. А., Соколова Е. В. Автоматическое извлечение ключевых слов и словосочетаний из русскоязычных текстов с помощью алгоритма KEA. В сб.: *Компьютерная лингвистика и вычислительные онтологии*. Вып. 1. СПб., 2017. С. 157–165. <https://clck.ru/RzcaY> (дата обращения: 01.08.2020).
- Непараметрические корреляции — Непараметрические корреляции. *StatSoft*. <https://clck.ru/RzcdD> (дата обращения: 30.10.2020).
- Птенцова 2008 — Птенцова А. В. Семантическая оппозиция глаголов знати и вѣдѣти на материале русских оригинальных памятников XI–XVI вв. *Die Welt der Slaven*. 2008, (LIII): 265–278.
- Соломоновская 2014 — Соломоновская А. Л. Семантическое поле «интеллектуальная деятельность» в диалекте средневекового переводчика. *Вестник славянских культур*. 2014, (31): 118–127.
- Ягунова, Пивоварова 2013 — Ягунова Е. В., Пивоварова Л. М. От коллокаций к конструкциям. В сб.: *Русский язык: конструкционные и лексико-семантические подходы*. СПб., 2013. <https://bit.ly/2OWkAmC> (дата обращения: 13.06.2020).
- Baranov 2018 — Baranov V. A. Text Corpus of Medieval Manuscripts as a Goal and a Tool for Linguistic Research. In: *Editing Mediaeval Texts from a Different Angle: Slavonic and Multilingual Traditions*. Paris, Bristol, Connecticut: Peeters Leuven, 2018. P. 283–308.
- Baranov, Gnutikov 2019 — Baranov V. A., Gnutikov R. M. The statistics and n-gram modules of the historical corpus “Manuscript”. In: *Digital and Analytical Approaches to the Written Heritage*. Sofia: Gutenberg Publishing House, 2019. P. 9–28.
- Evert 2005 — Evert S. *The statistics of word cooccurences word pairs and collocations*. PhD thesis. Stuttgart: Universität Stuttgart, 2005. 353 p. <https://clck.ru/Rzcob> (accessed 15.07.2020).
- Kutuzov et al. 2018 — Kutuzov A., Øvrelid L., Szymanski T., Velldal E. Diachronic word embeddings and semantic shifts: a survey. In: *Proceedings of the 27th International Conference on Computational Linguistic*. Santa Fe, 2018. P. 1384–1397. <https://clck.ru/Rzcpk> (accessed 15.07.2020).
- Manning, Schütze 2000 — Manning C., Schütze H. *Foundations of Statistical Natural Language Processing*, 2000. 680 p. <https://clck.ru/RzcrE> (accessed 15.07.2020).
- Žolobov 2016 — Žolobov O. Present tense forms variability in the Paroemiaron Zacharianum d. 1271 (to the parchment internet-edition). *Zeitschrift für Slawistik*. 2016, 61 (2): 305–321.

Статья поступила в редакцию 29 августа 2020 г.

Статья рекомендована в печать 3 декабря 2020 г.

Oleg F. Zholobov

Kazan Federal University,
18, Kremlevskaia ul., Kazan, 420008, Russia
ozolobov@mail.ru

Victor A. Baranov

Izhevsk State Technical University,
7, Studencheskaia ul., Izhevsk, 426069, Russia
victor.a.baranov@gmail.ru

Verbs meaning ‘know’ in Old Church Slavonic and Old Russian writing: Distributive, quantitative, and semantic properties*

For citation: Zholobov O. F., Baranov V. A. Verbs meaning ‘know’ in Old Church Slavonic and Old Russian writing: Distributive, quantitative, and semantic properties. *Vestnik of Saint Petersburg University. Language and Literature*. 2021, 18 (1): 56–76. <https://doi.org/10.21638/spbu09.2021.104> (In Russian)

The article presents the first attempt of a distributive and quantitative analysis of the lexico-semantic series in the Old Russian language based on three multi-genre subcorpora of the historical corpus “Manuscript” (manuscripts.ru): lists of the Gospels, menaia, and chronicles. The authors made a correlation between the software processing of the diachronic corpus of data and their historical and linguistic status. The semantic relations between the verbs with the general meaning ‘know’ in the Old Russian language are considered: *věděti*, *vdati*, *znati*. Their substitution in the modern Russian standard by only the verb *znat’* raises the question about this lexical group’s evolutionary dynamics. The authors established that the entire series belongs to the original lexical system, although the verb *vědati* was not found in Old Slavonic manuscripts and became widespread in Old Russian sources, both colloquial and literary. The analysis proves that the verb *vědati* in the Old Russian written sources acts as a substitute for the athematic verb *věděti*. The revealed quantitative bigram-indicators made it possible to establish contrasting collocations with the verbs *věděti* and *vědati*, on the one hand, and with the verb *znati*, on the other. Among the collocates of the verb *znati*, there were neither abstract nouns nor collocates that attach propositions. Distributive areas of the synonymy distribution in the verb row were found, which reflected the start of diachronic competition in the chain of lexemes. As a method for assessing the proximity of the bigrams’ components (analyzed verbs and their collocates), the authors used the statistical measure T-score and the r-Pearson correlation coefficient to determine the distribution’s degree series mutual correspondence.

Keywords: Old Church Slavonic and Old Russian written sources, verbs from the semantic cluster ‘know’, n-grams, distributive and quantitative analysis.

References

- Баранов 2019 — Baranov V. A. Creation and use of historical corpuses of Slavic written monuments. *Scripta & e-Scripta* 2019, (19): 33–57. (In Russian)
- Баранов, Жолобов 2020 — Baranov V. A., Zholobov O. F. Statistical and linguistic research of frequent words in the Words of Kirill Turovsky (based on the manuscript of the National Library of Russia, F. p. I. 39). *Slověne*. 2020, 9 (1): 29–80. <https://doi.org/10.31168/2305-6754.2020.9.1.2>. (In Russian)
- Бобкова 2015 — Bobkova T. Extraction of collocations from the corpus of Ukrainian texts. *Research Journal Studies about Languages*. 2015, (27): 93–105. (In Russian)
- Браславский, Соколов 2006 — Braslavskii P., Sokolov E. Comparison of four methods for automatically extracting two-word terms from text. In: *Komp’iuternaia lingvistika i intellectual’nye tekhnologii*. Mos-

* The study was supported by the Russian Science Foundation (RSF), project No. 20-18-00206.

- cow: Rossiiskii gosudarstvennyi gumanitarnyi universitet Publ., 2006. <https://clck.ru/RzcG6> (accessed 15.07.2020). (In Russian)
- Жолобов 2014 — Zholobov O. On reflexes of *ti-* and *t-*forms of verbs in Ancient Russian. *Russian Linguistics*. 2014, 38 (1): 121–163. (In Russian)
- Зализняк 2004 — Zalizniak A. A. *Old Novgorod dialect*. Moscow: Iazyki slavianskoi kul'tury Publ., 2004. 872 p. (In Russian)
- Захаров, Хохлова 2014 — Zakharov V. P., Khokhlova M. V. Automatic detection of terminological phrases. *Strukturalnaia i prikladnaia lingvistika*. 2014, (10): 182–200. (In Russian)
- Кочеткова 2013 — Kochetkova N. A. Statistical language methods. Collocations and colligations. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh*. 2013, (16): 301–305. <https://clck.ru/RzcNe> (accessed 15.07.2020). (In Russian)
- Масевич, Захаров 2016 — Masevich A. Ts., Zakharov V. P. Methods of corpus linguistics in historical and cultural studies. In: *Komp'iuternaia lingvistika i vychislitel'nye ontologii*. St. Petersburg, 2016. P. 24–43. <https://clck.ru/RzcWC> (accessed 15.07.2020). (In Russian)
- Митрофанова и др. 2008 — Mitrofanova O. A., Belik V. V., Kadina V. V. Corpus research of combinable preferences of the frequency lexemes of the Russian language. In: *Komp'iuternaia lingvistika i intellektual'nye tekhnologii*. Moscow, 2008. <https://clck.ru/RzcYr> (accessed 15.07.2020). (In Russian)
- Митрофанова, Соколова 2017 — Mitrofanova O. A., Sokolova E. V. Automatic extraction of keywords and phrases from Russian-language texts using the KEA algorithm. In: *Komp'iuternaia lingvistika i vychislitel'nye ontologii*. Issue 1. St. Petersburg, 2017. P. 157–165. <https://clck.ru/RzcaY> (accessed 15.07.2020). (In Russian)
- Непараметрические корреляции — Nonparametric correlations. In: *StatSoft*. <https://clck.ru/RzcdD> (accessed 30.10.2020). (In Russian)
- Птенцова 2008 — Ptentsova A. V. Semantic opposition of the verbs of знати and вѣдѣти on the material of Russian original monuments of the 11th–16th centuries. *Die Welt der Slaven*. 2008, (LIII): 265–278.
- Соломоновская 2014 — Solomonovskaia A. L. Semantic field “intellectual activity” in the idiolect of a medieval translator. *Vestnik slavianskikh kul'tur*. 2014, (31): 118–127.
- Ягунова, Пивоварова 2013 — Iagunova E. V., Pivovarova L. M. From collocations to structures. In: *Russkii iazyk: konstruksionnye i leksiko-semanticheskie podkhody*. St. Petersburg, 2013. <https://bit.ly/2OWkAmC> (accessed 15.07.2020). (In Russian)
- Baranov 2018 — Baranov V. A Text Corpus of Medieval Manuscripts as a Goal and a Tool for Linguistic Research. In: *Editing Mediaeval Texts from a Different Angle: Slavonic and Multilingual Traditions*. Paris, Bristol, Connecticut: Peeters Leuven, 2018. P. 283–308.
- Baranov, Gnutikov 2019 — Baranov V. A., Gnutikov R. M. The statistics and n-gram modules of the historical corpus “Manuscript”. In: *Digital and Analytical Approaches to the Written Heritage*. Sofia: Gutenberg Publishing House, 2019. P. 9–28.
- Evert 2005 — Evert S. *The statistics of word cooccurrences word pairs and collocations*. PhD thesis. Stuttgart: Universität Stuttgart, 2005. 353 p. <https://clck.ru/RzcoB> (accessed 15.07.2020).
- Kutuzov et al. 2018 — Kutuzov A., Øvrelid L., Szymanski T., Velldal E. Diachronic word embeddings and semantic shifts: a survey. In: *Proceedings of the 27th International Conference on Computational Linguistic*. Santa Fe, 2018. P. 1384–1397. <https://clck.ru/Rzcpk> (accessed 15.07.2020).
- Manning, Schütze 2000 — Manning C., Schütze H. *Foundations of Statistical Natural Language Processing*, 2000. 680 p. <https://clck.ru/RzcrE> (accessed 15.07.2020).
- Žolobov 2016 — Žolobov O. Present tense forms variability in the Paroemiaron Zacharianum d. 1271 (to the parchment internet-edition). *Zeitschrift für Slawistik*. 2016, 61 (2): 305–321.

Received: August 29, 2020
Accepted: December 3, 2020