

GYÖRFI BEÁTA
(Szeged, Magyarország)

«Умный в гору не пойдет...»: использование электронных корпусов для диахронных исследований

Аннотация: Настоящая статья имеет экспериментальный характер. Через изучение конкретной языковой проблемы – особенности употребления дискурсивных энклитик в двух регистрах древнерусского языка – демонстрируются преимущества и недостатки привлечения электронных корпусов для подобных исследований.

Ключевые слова: корпусная лингвистика, НКРЯ, диахронический корпус, дискурсивные энклитики, древнерусский язык, берестяные грамоты

0. Введение

Корпусная лингвистика является стремительно развивающейся областью языкознания. Исследования, проведенные при помощи электронных корпусов, применяются в разных сферах языкознания: в лексических и фразеологических исследованиях, в составлении грамматик и словарей, в социолингвистических и психолингвистических исследованиях, а также для подтверждения научных гипотез. Хотя корпусная лингвистика на русской почве появилась позже, чем на западе, она стала активно развиваться. Об этом свидетельствует постоянно возрастающее количество русскоязычных корпусов и научных анализов, проведенных на их основе (см. ДБЭРФИ 2021).

Однако использование электронных корпусов до сих пор не распространилось в сфере диахронных исследований, даже на западе мало таких лингвистов, которые опираются на данные корпусов.

Настоящая статья, таким образом, имеет двоякую цель: лингвистическую и методологическую. Через изучение конкретного вопроса – употребления дискурсивных энклитик в памятниках XII–XV вв. – представляются как и преимущества, так и некоторые недостатки работы с корпусом в сфере диахронии. Таким образом, предпринимается попытка побудить коллег к использованию электронных корпусов в ходе своих исследований.

В соответствии с поставленными целями статья построена следующим образом. В первой части дается краткий обзор диахронной корпусной лингвистики и диахронных корпусов русского языка с фокусом на средствах настоящего анализа, т. е. на исторический подкорпус НКРЯ. Во

второй части следует постановка проблемы настоящего исследования: представляется система древнерусских энклитик, уделяется особое внимание дискурсивным энклитикам. В третьей части, опираясь на данные НКРЯ, сопоставляются функциональные особенности использования дискурсивных энклитик в двух регистрах древнерусского языка. В последней части подводятся итоги.

1. Корпусные исследования в сфере диахронной лингвистики

Диахроническая корпусная лингвистика, на первый взгляд, не является чем-то новым, поскольку историческая лингвистика всегда опиралась на корпусы, если под корпусом понимается «собрание текстов или частей текстов, над которым ведется общий языковой анализ» (MEYER, 2002, р. xi). Диахроническая «версия» корпусной лингвистики делает изучение истории, развития, изменения языка быстрым, эффективным и, таким образом, более привлекательным.

В научной литературе различаются два типа диахронных корпусов: доэлектронные, которые создавались перед изобретением компьютера, и электронные, которые уже используют информационные технологии. Далее терминология осложняется тем, что под электронным диахроническим корпусом понимаются, с одной стороны, электронные издания памятников (т. е. сканированные, транскрибированные версии памятников), а с другой – информационно-справочные системы, основанные на собрании древних текстов в электронной форме.

Корпусные исследования могут быть двух типов (и это, конечно, ничем не отличается и в диахронической отрасли): исследования, основанные на корпусных данных (corpus-based), когда данные корпуса используются для подтверждения, опровержения или усовершенствования лингвистической теории, и ориентированные на корпуса (corpus-driven), когда источником лингвистической гипотезы является сам корпус.

На западе, например, в исторической лингвистике английского языка применение электронных корпусов имеет уже богатую и сравнительно долгую традицию. Применение корпусов способствует изучению целого ряда диахронических явлений: 1) языкового развития с контрастивной и компаративной точек зрения, 2) лексических и грамматических изменений, 3) изменений в частотности употребления слов, грамматических конструкций и фразеологизмов, 4) диахронического анализа стилистических и жанровых разновидностей (становление исторической прагматики проходило в 1990 гг.), 5) квантитативного анализа языковых явлений. При помощи корпусов также становится возможным изучение «соперничества» вариативных форм.

Для проведения диахронических исследований редко имеется доступ к сбалансированным данным, которые охватывали бы целое жанровое разнообразие исследуемой языковой эпохи. В изучении такого сложного и многослойного явления, например, как древнерусский язык, результаты

работ часто оказываются недостоверными из-за того, что в проанализированном памятнике преобладают или церковнославянские, или разговорные элементы. Применение объемных и сбалансированных корпусов приводит к более надежным результатам (КҮТӖ 2011: 421).

В соответствии с вышесказанным выходит в свет все больше диахронных корпусов разных языков: для английского Dictionary of Old English Corpus in Electronic Form, для немецкого Mittelhochdeutsche Begriffsdatenbank, Bonner Frühneuhochdeutsches Korpus и DeutschDiachronDigital, для французского Textes de Français Ancien, для испанского Corpus del Español (КҮТӖ 2011).

Из русскоязычных диахронических корпусов стоит отметить Регенсбургский диахронических корпус русского языка, Параллельный корпус переводов «Слова о полку Игореве», корпус «Великие Минеи Четьи» (содержащий документы 16-го века) <http://www.vmc.uni-freiburg.de>, проект «Манускрипт» http://www.manuscripts.ru/?p_lid=1 и Санкт-петербургский корпус агиографических текстов (СКАТ) (<http://project.phil.spbu.ru/scat/page.php?page=project>) (ЗАХАРОВ 2013: URL; МИТРЕНИНА 2014)

1.2. Диахронический подкорпус НКРЯ

Национальный корпус русского языка (<https://ruscorpora.ru/new/index.html>) является одним из наиболее объемных корпусов русского языка, содержащий 990 миллионов словоупотреблений (на 01. 06. 2021). НКРЯ включает в себя ряд подкорпусов, в том числе и исторический. Данный подкорпус состоит из четырех частей: из древнерусского, старорусского, церковнославянского корпусов и из корпуса берестяных грамот.

Церковнославянский корпус (<https://ruscorpora.ru/new/search-orthlib.html>) представляет восточнославянский извод церковнославянского языка. Корпус содержит главным образом литургические тексты. В него входит более 4 миллионов словоупотреблений из 1160 текстов.

Старорусский корпус (https://ruscorpora.ru/new/search-mid_rus.html) содержит 6200 текстов с XIV-го до XVIII-го вв. Он содержит более 8 миллионов словоупотреблений.

Древнерусский корпус (https://ruscorpora.ru/new/search-old_rus.html) содержит 27 документов (летописи, произведения Кирилла Туровского, переводы с греческого языка) и 573 тысячи словоупотреблений.

Корпус берестяных грамот (<https://ruscorpora.ru/new/search-birchbark.html>) включает в себя 885 грамот и 19000 словоупотреблений. Кроме знаменитых новгородских, корпус содержит смоленские, псковские, тверские, витебские и звенигородские грамоты.

Церковнославянский, древнерусский корпуса и Корпус берестяных грамот разрешают проведение лексических и грамматических поисков, учитывая дистанцию между элементами. Также в этих корпусах возможно установить три орфографических варианта правописания: точный, модернизированный и упрощенный. В каждом подкорпусе можно создать

подкорпус для собственных исследований с определением заглавия текста, временного промежутка, жанра или (в Корпусе берестяных грамот) местом создания текста.

2. Дискурсивные клитики древнерусского языка

Клитики являются сложными и спорными элементами, так как они взаимодействуют с разными уровнями языка – фонемным, морфемным, синтаксическим – и прагматикой (подробнее см. ДЬЁРФИ 2020). Исследование клитик попало в центр внимания лингвистов только в последнее 15–20 лет. Большое количество исследований посвящено, например, изучению клитик в романских языках, а также вышли в свет научные работы по теме славянских клитик (FRANKS 1999; ЗАЛИЗНЯК 2008, KOSTA, ZIMMERLING 2013, FRANKS, JUNGHANNS, LAW 2004, ЦИММЕРЛИНГ 2009).

Славянские энклитики представляют интерес с разных точек зрения. Что касается синтаксиса, характерной особенностью словоупотребления в славянских языках является наличие грамматических ограничений на расположение фразовых клитик (KOSTA ZIMMERLING 2013: 179). Таким образом, можно сказать, что энклитики представляют собой единственную стабильную часть предложения: несмотря на свободный словоупотребление, они занимают устойчивую позицию (вакернагельскую – вторую позицию после первого слова или первой фразы или V2 – позицию после глагола).

Клитики обращают на себя внимание и отражают изменения языка во времени, поскольку они полностью исчезли в восточнославянских языках, часть энклитик сохранилась в западнославянских языках, также они «процветают» в южнославянских языках.

2.1. Система древнерусских энклитик

Древнерусский язык располагал более развернутой системой клитик, чем современный русский. Их свойства, функция и положение совершенно отличались от современных эквивалентов, поскольку они играли значительную роль в построении фраз.

Исследование древнерусских энклитик представляется сложным по разным причинам. Они изучены мало. В классических трудах по русской исторической лингвистике (СТЕЦЕНКО 1977, ИВАНОВ 1983, КОЛЕСОВ 2005, ЯНОВИЧ 1986, БОРКОВСКИЙ, КУЗНЕЦОВ 1963/2006) их отдельно не описывают или только лишь перечисляют. Если же системы энклитик касаются в научных трудах, то высказываются противоположные взгляды относительно их функции и значения. Например, энклитика *же* по наблюдениям Зализняка имеет противительное, усилительное и причинное значение (ЗАЛИЗНЯК 2008: 29). Мигдальски (2007) и Циммерлинг (2009, 2018) при определении функции *же* подчеркивают уже ее дискурсивную функцию: Мигдальски выражает мнение, что *же*

послужила для выражения иллокутивной силы, а по Циммерлингу она выражает фокус, контраст и эмфатическое значение (ЦИММЕРЛИНГ 2009: 3).

Самый исчерпывающий анализ энклитик встречается у Зализняка (2008). Он составил рейтинг ДР энклитик на основе занимаемой ими позиции в цепочках:

- 1) *же*;
- 2) *ли*;
- 3) *бо*;
- 4) *ти*;
- 5) *бы*;
- 6) дательные местоимения – *ми, ти, си, ны, вы, на, ва*;
- 7) местоимения в винительном падеже – *мя, тя, ся, ны, вы, на, ва, и, ю, е, ъ, я*;
- 8) связки: *есмь (есми), еси, есмъ (есме, есмо, есмы), есте, есвѣ, еста* [2 дв.]; дополнительную группу составляют связки 3-го лица: *есть, суть, еста*.

Как видно из данного перечисления, функционально различались три типа энклитик: дискурсивные, прономинальные и глагольные. Расположение энклитик регулировалось правилом Вакернагеля, однако оно могло быть модифицировано вставлением ритмико-синтаксических барьеров, под влиянием которых энклитики перемещались вправо.

По сфере воздействия энклитики могли быть фразовыми (когда они находились в вакернагельской позиции и относились к целой фразе) и локальными (когда сферой действия являлось только предшествующее слово).

2.2. Дискурсивные энклитики

Дискурсивные энклитики являются «специальным классом наречий», которые способствуют ориентации говорящего управлять структурой разговора (SPENCER, LUIS 2012: 34). Они являются единственным типом клитик, у которых нет других, неклитических аналогов. В некоторых языках они устанавливают ограничения относительно синтаксического и категориального статуса своего главного слова. Изучение дискурсивных энклитик из-за их синтаксических особенностей представляет интерес в ряде языков, как, например, в японском или в финском.

Функция дискурсивных клитик древнерусского языка резко отличалась от нынешнего состояния: *же* – по всей вероятности выражала контрастивный топик, *ли* – означала альтернативу, *бо* – служила для

выражения причинных отношений, *ти* – носила утвердительное значение, а *бы* послужила для выражения оппозитива.

2.3. Изучение дискурсивных энклитик корпусным методом

В настоящем исследовании рассматривается распространенность и расположение энклитик в двух регистрах языка: в языке летописей и в языке берестяных грамот. Объектом исследования, таким образом, является огромное количество материала, обработка которого была бы невозможна без применения корпусов.

Следуя за стандартными проблемами исследования энклитик, мы ищем ответы на следующие вопросы:

1. Какую позицию занимают дискурсивные энклитики? Вакернагельскую или иную?
2. С какой частотой встречаются отдельные виды дискурсивных энклитик в тексте летописей и берестяных грамот?
3. В каком регистре они используются чаще?
4. Меняется ли частотность их использования со временем? Как?
5. Несмотря на то что в теории энклитики могут присоединяться к любому фонетическому слову, есть ли у них склонность к доминантному главному слову?
6. Насколько распространены цепочки клитик в двух регистрах?

Как уже говорилось, исторический корпус НКРЯ является репрезентативным материалом, так как он включает в себя четыре подкорпуса (церковнославянский, старославянский, древнерусский и корпус берестяных грамот), которые покрывают все регистры древнего состояния языка. Исторический корпус, подобно основному корпусу, является размеченным. Разметка включает в себя лемматизацию, т. е. сведение всех словоформ данной леммы к единой лемме, морфологическую разметку, т. е. перечень грамматических признаков, и дополнительную информацию. В случае переводных текстов разметка включает в себя и греческую форму, и греческую лемму.

Корпус предоставляет пользователю возможность осуществлять поиск необязательно во всех текстах, а задать собственный подкорпус и работать на суженном, выбранном материале. При этом в историческом корпусе можно отобрать тексты по самым разным параметрам: в древнерусском подкорпусе – по заглавиям, в корпусе берестяных грамот – по месту нахождения документа, по номеру грамоты, по жанру, по языку, по материалу и по условной дате возникновения грамоты.

Поисковая система обеспечивает поиск слов или словосочетаний в фиксированной форме, во всех возможных формах, поиск слов с определенной грамматической характеристикой, поиск слова в сочетании с любым словом или с словом с определенной грамматической характеристикой. Система также предоставляет статистическую информацию относительно

употребления данного элемента. В результате поиска изучаемая единица представляется в контекстах со ссылками на источник, а также дается статистическая информация о частоте ее использования о чем.

2.3.1. Создание подкорпусов

В ходе анализа сопоставляются особенности использования дискурсивных энклитик в двух регистрах древнерусского языка: в языке летописей и в берестяных грамотах.

Язык летописей является книжным языком, совмещающим элементы русского извода церковнославянского языка и элементы разговорного языка. Язык берестяных грамот отражает разговорную речь.

Для осуществления анализа в обоих корпусах был задан подкорпус. В древнерусском подкорпусе были отсортированы летописи: «Повесть временных лет» (ПВЛ) (XII), Киевская летопись (КЛ) и Галицкая летопись (ГЛ) (начало XIII в.), Волынская летопись (ВЛ) (конец XIII в.), Суздальская летопись (СЛ) (XIV в.) и Новгородская 1-я летопись, Синодальный список (НЛ) (XV в.). С данными текстами можно работать и совместно и по отдельности.

В этот момент проявляется первый недостаток работы с НКРЯ: при создании подкорпуса у пользователей нет информации о его объеме, таким образом, невозможно проводить частотный и пропорциональный анализ. Можем лишь делать предположения, опираясь на количество результатов. Это, конечно, может привести к ошибочным выводам, поскольку некоторые летописи (например КЛ) являются более объемными, чем другие (например НЛ).

Задание подкорпуса в корпусе берестяных грамот происходит иначе: здесь можно выбрать тексты по разным критериям (по жанру, месту нахождения, материалу). Для настоящего анализа были отсортированы грамоты по дате их создания и выбраны тексты того же периода, что и в древнерусском корпусе (т. е. XII–XV вв.). Для заданного подкорпуса здесь представляется информация о количестве содержащихся текстов и словоформ (корпус включает в себя 868 грамот, 18832 словоупотребления).

2.3.2. Расположение энклитик

Энклитики обращали на себя внимание лингвистов относительно своей устойчивой синтаксической позиции. В современных славянских языках обнаружены две позиции энклитик: вакернагельская (2P), когда энклитики следуют за первым фонетическим словом (или за первой фразой, как в сербском), и V2, когда энклитики следуют за глаголом. Созданы и классификации славянских языков по этому параметру (KOSTA, ZIMMERLING 2013).

В ходе развития языка в памятниках отражалось все больше случаев, когда энклитики занимали позицию «ниже», т. е. перемещались направо. Данные энклитики являются или локальными, при этом воздействуют

только на единственное слово, или отражают уже разрушение системы энклитик.

Для проявления расположения энклитик мы изучаем словосочетания, поскольку энклитики присоединяются к фонетическому слову. В первой полосе поиска среди дополнительных признаков надо указать, что первое слово словосочетания (тактовой группы) должно стоять в начале предложения. Во второй полосе надо только уточнить частеречный признак второго элемента (т. е. частицу).

В результате поиска из текста летописей получили 1160 вхождений, а из корпуса берестяных грамот только 17.

(1) Звенигородцемъ же лютѣ воруцимса имъ с ними и не поущаюшимъ ко граду. ни ко встрожнѣмъ вратомъ. внем же стояцимъ вкр(с)тъ града Василкоу же княжцю во Белзѣ. и придоша же ѿ него великии Вачеславъ. Толъстѣи. и Мирославъ и Дьмянъ. и Воротиславъ. инии воаре мнозѣ и вои ѿ Белза. а ѿ Лестка из Лаховъ. Соудиславъ Бернатовичъ. со многими Поланзи. и ѿ Пересопници. приде Мьстиславъ. Нѣмзи со многими вои. Влксандръ с вратомъ [Галицкая летопись // Ипатьевский список (БАН, 16.4.4), первая четверть XV в., по изд.: Полное собрание русских летописей. Т. II. Ипатьевская летопись. М., 2001. С. 715–848]

(2) ...исти въддита же ми... [Берестяная грамота 455 (1180–1200)]

(3) ... а нь вслушаиса [Берестяная грамота 66 (1340–1360)]

(4) мацешини же дети се соуте гордосте: непокорение: прекословее: презоресво: хоула: кевета: зломыслие: глево: вражда: пеанесво: игры не:приазнины и всакаа злобе: а кало есте клевета: хоула: гнево осоужение прекословее: сваро: вои: зависте: вражда злопоминание: непокорение злосердее: злии помысли: смехотворение: и вса игры бесовескыа: та же пакы запоисво: резонимание: гравление: развои: татба: дшеговление потвори поклепо отрава влоуди прелюводеание царотворение [Берестяная грамота Торж. 17 (1180-1200)]

Как видно из примеров, поисковая система не сумела удачно справиться с заданной проблемой. Среди результатов было много примеров с другими частицами (3), часть результатов фрагментирована, не доступен контекст (2) (3).

Также с такой настройкой параметров получили сравнительно мало примеров для 2Р позиции. Данное явление объясняется тем, что древние тексты не расчленены на предложения, а в них различаем лишь клаузы

(ЗАЛИЗНЯК 2008: 12–13). Поисковая система не распознает данные синтаксические единицы.

Если же для поиска определяем только грамматический признак «частица» и «вручную» анализируем результаты, то становится ясным, что почти все примеры отражают 2Р позицию энклитик. К сожалению, нельзя подтвердить статистическими данными это предположение.

2.3.3. Встречаемость дискурсивных энклитик

Для решения данной проблемы в полосу поиска недостаточно вписать конкретные клитики, надо также определить их грамматические признаки (т. е. их частеречную принадлежность), поскольку некоторые клитики омонимичные с другими формами (напр.: *же* может быть и союзом, *ти* – местоимением). Осуществив поиск, мы получили следующие результаты в корпусе древнерусских текстов:

	же	ли	бо	ти	бы
ПВЛ	1069	113	380	5	7
КЛ	1949	38	350	44	1
ГЛ	1094	19	199	4	6
ВЛ	535	17	74	5	6
ЦЛ	880	23	219	3	-
НЛ	245	21	79	-	-

Результаты из корпуса берестяных грамот:

	же	ли	бо	ти	бы
Грамоты XII вв.	57	12	-	83	4
Грамоты XIII вв.	25	8	-	14	-
Грамоты XIV вв.	2	13	-	5	16
Грамоты XV вв.	7	-	-	1	18

Хотя нет возможностей для проведения пропорционального анализа (поскольку электронный корпус не предоставляет нам информации о количестве слов в отдельных текстах), содержащиеся в таблицах данные указывают на то, что энклитики в языке летописей, по всей вероятности, были более распространены, чем в берестяных грамотах. Конечно, из-за нехватки информации об объеме летописного корпуса невозможно делать выводы относительно частоты их употребления.

Что касается различия в распространенности отдельных энклитик: в текстах летописей самой широкоупотребляемой энклитикой является *же*, а в грамотах – *ти*. Частотность данных энклитик, по всей вероятности, возводится к жанровой специфике памятников. Текст летописей следует строгим логическим правилам, таким образом, обозначение новой или известной информации в них представляется важным. Видимо, в грамотах первичным является утверждение какой-то информации.

Сведения таблиц способствуют рассмотрению изменения частотности энклитик во времени. Данные обоих корпусов указывают на тенденцию исчезновения дискурсивных энклитик.

2.3.4. Свойства главного слова

Энклитики могут присоединяться к любому фонетическому слову, несмотря на его дистрибутивные свойства. Однако в некоторых языках можно выделить доминирующий тип главного слова при энклитиках. Применение корпусов открывает возможность изучать и данное явление.

В обоих подкорпусах в первой полосе поиска был установлен частеречный признак главного слова, во второй полосе вписаны энклитики. Запрос дал следующие результаты в древнерусском корпусе:

	же	бо	ли	ти	бы
S	2252	171	20	9	1
Adj	183	60	14	-	-
Num	26	6	-	-	-
V	747	528	45	3	-
Adv	364	79	11	6	6
Pron	119	37	7	6	2

Как показывают результаты этого количественного анализа, можно указать доминирующее слово по частеречным признакам для каждой энклитики. *же* в подавляющем большинстве случаев присоединяется к существительным, *бо* и *ли* – к глаголам. Мы имеем мало примеров с энклитиками *ти* и *бы*, но на основе имеющихся данных можно предположить, что *ти* выступает преимущественно с существительными, а *бы* – с наречиями.

Результаты на основе корпуса берестяных грамот отражают другие тенденции:

	же	ли	бо	ти	бы
S	12	2	-	12	14
Adj	7	6	-	9	2
Num	-	-	-	1	-
V	35	13	-	17	3
Adv	10	3	-	9	-
Pron	28	2	-	21	3

В языке берестяных грамот энклитика *же* отдает предпочтение глаголам, *ли*, подобно языку летописей, – глаголам, *ти* – местоимениям и *бы* – существительным.

2.3.5. Цепочки клитик

В исследованиях по славянским клитикам ведущее место занимает изучение цепочек клитик и правил следования клитик внутри цепочек. Корпусный анализ способствует и изучению сочетаемостных свойств слов, таким образом он делает возможным изучение и этой проблемы.

Цепочки в теории содержат не исключительно дискурсивные, но и местоименные и связочные энклитики.

Во-первых, в обоих полосах поиска отмечен частеречный признак «частица». В результате получились примеры с отрицательными частицами «не» и «ни» в обоих подкорпусах.

(5) и приѣхаша к нему вси на єдино мѣсто. всеволодъ же не хотя того. вже са бра(т)ѣа съвзкоупила въдиноу мзисль. посла къ двдвчема рка има [Киевская летопись // Ипатьевский список]

Далее для сужения результатов в полосах поиска были вписаны дискурсивные энклитики. Таким образом, в обоих корпусах было получено несколько цепочек клитик, однако они состояли из двух энклитик:

(6) и глѣху ему и нашъ еси поклонилъса еси нашему старѣшины и намъ. внъ же глѣ имъ. вашъ старѣшина естъ антихрѣсть. а вы вѣси есте. и перекрѣстиса. и тако ищезнаху. ввогда же ли пакы в ноци прихожаху к нему [Повесть временных лет. Ипатьевский список]

(7) се бо wskверниша и пожгоша стѣи домъ твои. манастырь матере [т]воеа и трупье рабъ твоихъ оубиша бо ѿ братьа нашего нѣколко

вружьемъ· безбожьнии снѣе измалеви· пущении на казнь хрестыаномъ·
аще ди бо си суть ѿ пустына етривьскиа· mezi възтокомъ и сѣверомъ·
ицьло жь есть· ихъ колѣнъ· [Повесть временных лет. Ипатьевский
список]

Найден лишь один пример с тремя клитиками:

(8) (ѿ) [и]вана ко роусиле выправи ми дъвр... ---ю арине възьми оу
кзньзюа грѣноу оу неж[а]... ал[и] ти не въздасть а възьми (в)ъ третъ а
а ти... ...къ да же ти ми боудете дъвр[о] (а) присълоу ти... оже хъчеши
пъ -зне а присъли ----[е] челъв[е](к)... чи ти поченете (ч)етъ
лести(ти)... а не мъзи четь (м)ълзвити [Берестяная грамота Смол. 12
(1100-1200)]

Следует также отметить, что в обоих регистрах обнаружено мало таких цепочек. Следовательно, можно предположить, что цепочки не были характерны для данного языкового состояния.

3. Итоги

В статье рассмотрены достоинства и недостатки применения электронных корпусов в области диахронических исследований. Изучены особенности употребления энклитик в двух регистрах языка. Результаты корпусного анализа мы подтвердили некоторые уже известные явления: например, 2Р положение энклитик и тенденцию к их исчезновению к XVI в. Новизной исследования являются количественные показатели частотности употребления энклитик в летописных текстах и в тексте берестяных грамот, частотности отдельных энклитик и частеречных свойств главного слова при них.

Хотя поисковая система НКРЯ не справилась со всеми заданными запросами, главным достоинством электронных корпусов является то, что составление и анализ объемных текстов стали возможными и более эффективными.

Литература

- БОРКОВСКИЙ В.И., КУЗНЕЦОВ П.С. (1963/2006) Историческая грамматика русского языка. Москва, 1963/2006.
- ДЪЕРФИ Б. (2021) Положение корпусной лингвистики в русском языкознании. А.А. Шунейко: Корпусная лингвистика. Учебник для вузов. 2020. Москва. Юрайт ISBN 978-5-534-13603-6 // Slavica (Annales Instituti Slavici Universitatis Debreceniensis), 50. 286-296.
- ЗАЛИЗНЯК А.А. (2008) Древнерусские энклитики. Москва, 2008.
- ЗАХАРОВ В.П. (2013) Корпусная лингвистика в России. Доклад на конференции: IV международный научный симпозиум Retro-2013. Ретроспектива филологии

- в информационном обществе знаний (Украина, Крым, Песчаное, 27 июня – 3 июля 2013 г.) https://www.academia.edu/7911977/%D0%9A%D0%BE%D1%80%D0%BF%D1%83%D1%81%D0%BD%D0%B0%D1%8F_%D0%BB%D0%B8%D0%BD%D0%B3%D0%B2%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0_%D0%B2_%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D0%B8_Corpus_linguistics_in_Russia (дата обращения: 03.03.2021.)
- ИВАНОВ В.В. (1983) Историческая грамматика русского языка. Москва, 1983.
- КОЛЕСОВ В.В. (2005) История русского языка. Москва – Санкт-Петербург, 2005.
- СТЕЦЕНКО А.Н. (1977) Исторический синтаксис русского языка. Москва, 1977.
- ЦИММЕРЛИНГ А.В. (2009) Элементы предложения и синтаксические позиции в языках современной Европы // Язык и речевая деятельность, 8. 2009. 36–88.
- ЯНОВИЧ Е.И. (1986) Историческая грамматика русского языка. Минск, 1986.
- FRANKS S. (1999) Clitics in Slavic // The Slavic and East European Language Resource Center. <http://seelrc.org/glossos/>
- FRANKS S., JUNGHANNS U., LAW P. (2004) Pronominal Clitics in Slavic // Journal of Slavic Linguistics, vol.12. 2004. pp. 3–36.
- GYÖRFI B. (2020) A big change starts small: pronominal clitics in 12–15th century Old Russian Chronicles // Slavica (Annales Instituti Slavici Universitatis Debreceniensis), 49. Debrecen, 2020. 8–21. DOI: [10.31034/049.2020.01](https://doi.org/10.31034/049.2020.01)
- KOSTA P., ZIMMERLING A. (2013) Slavic Clitics: A typology // STUF, Akademie Verlag, 66/ 2, 2013. 178–214. DOI: [10.1524/stuf.2013.0009](https://doi.org/10.1524/stuf.2013.0009)
- KYTÖ M. (2011) Corpora and historical linguistics // Revista Brasileira de Linguística Aplicada, v. 11, n. 2. 2011. 417–457. DOI: [10.1590/S1984-63982011000200007](https://doi.org/10.1590/S1984-63982011000200007)
- MEYER C.F. (2002) English Corpus Linguistics: An Introduction. Cambridge, 2002. DOI: [10.1017/CBO9780511606311](https://doi.org/10.1017/CBO9780511606311)
- MIGDALSKI, K. (2007) On the emergence of the second position cliticization in Slavic // Formal Description of Slavic Languages FDSL-7. Leipzig, 69–71.
- MITRENINA O.V. (2014) The Corpora of Old and Middle Russian Texts as an Advanced Tool for Exploring an Extinguished Language // Scrinium, Volume 10. 2014. 455–461. DOI: [10.1163/18177565-90000109](https://doi.org/10.1163/18177565-90000109)
- SPENCER A., LUIS A.R. (2012) Clitics: An Introduction. Cambridge, 2012. DOI: [10.1017/CBO9781139033763](https://doi.org/10.1017/CBO9781139033763)
- ZIMMERLING A.V. (2018) ŽE LI or LI ŽE: discourse particles and clitic clusters in Late Old Russian // The XIII Annual Meeting of the Slavic Linguistics Society. Eugene, 28–29 September 2018 https://www.academia.edu/37454252/%C5%BDE_LI_or_LI_%C5%BDE_discourse_particles_and_clitic_clusters_in_Late_Old_Russian)

“Work smart, not hard”: the use of electronic corpora for diachronic research. The present article is meant to be an experiment. Through the study of a specific linguistic problem – the use of discourse enclitics in two registers of the Old Russian language – the advantages and disadvantages of the use of electronic corpora in similar investigations is demonstrated in the research.

Keywords: corpus linguistics, Russian National Corpus, diachronic corpus, discourse clitic, Old Russian, birchbark documents