

Беата Дьёрфи

ПОЛОЖЕНИЕ КОРПУСНОЙ ЛИНГВИСТИКИ В РУССКОМ ЯЗЫКОЗНАНИИ
А. А. ШУНЕЙКО: КОРПУСНАЯ ЛИНГВИСТИКА. УЧЕБНИК ДЛЯ ВУЗОВ. 2020.
МОСКВА. ЮРАЙТ ISBN 978-5-534-13603-6

The current status of corpus linguistics in Russian linguistics
Shuneyko, A. A. : Корпусная лингвистика. Учебник для вузов. 2020.
Moscow, Yurayt ISBN 978-5-534-13603-6

Abstract

Corpus linguistics is a relatively new, however rapidly developing field of linguistics. Nevertheless, its methodology and results are scarcely applied in current linguistic research. In the present paper a short overview of the history of corpus linguistics is given. The difficulties of the development and spread of this discipline in Russia are also described. A brief outline of Russian textbooks of corpus linguistics is also provided with special focus on Shuneyko's latest work.

Keywords: *the history of corpus linguistics, corpora of the Russian language, corpus methodology*

Корпусная лингвистика является сравнительно новой, и поэтому довольно малоизученной областью языкознания как в русской, так и в венгерской лингвистике. Несмотря на то, что существуют разные (и национальные, и узкоспециализированные) корпуса обоих языков, труды по корпусной лингвистике ограничиваются вопросами и трудностями создания корпусов или их описанием. Однако новейшие достижения корпусных исследований в таких работах не применяются.

В настоящей статье, помимо представления учебного пособия А.А. Шунейко, дается краткий обзор возникновения корпусной лингвистики. Перечисляются русскоязычные учебные пособия по данной области и сопоставляется положение российских и зарубежных исследований по данной сфере.

Становление корпусной лингвистики

Корпусная лингвистика (КЛ) представляет собой комплексную научную дисциплину. О данном свойстве свидетельствует и тот факт, что она упоминается под разными названиями: называют данную отрасль компьютерной, математической, статистической или квантитативной лингвистикой. В наши дни КЛ окружает нас, и, таким образом является востребованной областью среди лингвистических дисциплин, поскольку такие технические достижения, как автоматический перевод, программа проверки текста, голосовой поисковый запрос на смартфонах, подсказка синонимов и антонимов, генератор рифм, система фильтрации спама работают благодаря существованию языковых корпусов.



КЛ как самостоятельная область лингвистики возникла в 1950-е годы. Ее возникновение и развитие происходило параллельно с развитием компьютерной технологии, на которую она полагается опирается. Поскольку компьютер является удобным средством хранения и обработки информации, корпуса появились сразу же, как только лингвисты получили доступ к нему [ШАРОВ 2003: 14]. Подавляющее большинство исследований началось на материале английского языка. Это объясняется, с одной стороны, бурным развитием компьютерной техники в США, а с другой – интеллектуальным климатом в британской лингвистике в 60-е годы.

Исследования, опирающиеся на использование хранящихся на карточках корпусов, проводились уже до появления электронных корпусов [см. БОА 1940]. Они фокусировались на изучение частотности слов или коллокаций. Предшественниками корпусов можно считать списки частотности Джулиана и Буши [МАКЭНЕРИ–ХАРДИ 2013] для таких языков, как китайский, французский, испанский или арамейский. Приобретенные таким образом данные использовались в первую очередь в целях улучшения методики преподавания иностранных языков.

Корпусный анализ объемных текстов проводился компьютерами, как только они появились. Ранние компьютерные корпусные исследования продолжали изучение частотности и сочетаемости свойств слов. Благодаря развитию методики корпусной лингвистики в 1960-е годы Альфонс Жулянд установил главные принципы работы с электронными корпусами (такие, как сбалансированность, учет статистического разброса). [МАКЭНЕРИ–ХАРДИ 2013, ЖУЛЯНД–ЧАНГ–РОДРИГЕС 1964].

В это же время наблюдается отход лингвистов от корпусной лингвистики. Данное явление относится к периоду «хомскианской революции», т.е. к возникновению генеративной лингвистики. Хомский стремился переориентировать направленность лингвистических исследований: он настаивал на том, что объектом изучения должна быть не перформанция (E-language, по его поздней терминологии), т.е. созданные человеком языковые конструкции, предложения, а скорее компетенция (I-language), т.е. способность порождать нефинитное количество правильных предложений. Он утверждал, что данные перформанции мало или ничего не раскрывают о свойствах компетенции, поскольку количество возможных предложений нефинитное, однако имеющиеся предложения представляют лишь конечное множество языка [ХОМСКИЙ 1965: 8].

Таким образом, объектами исследований стали искусственно придуманные носителями языка правильные (или неправильные) предложения. Однако, взятые из речи, зачастую неправильно оформленные, хаотичные, зависящие от более широких контекстов корпусные высказывания не соответствовали новым целям исследования.

Взгляды Хомского привели к смене взглядов в лингвистике: до тех пор господствующие эмпирические тенденции уступали место рационализму, который опирается на интуицию и самоанализ при обзоре данных.

Эмпирический взгляд сохранился лишь в странах восточной Европы (например, математическая лингвистика в Центральной Европе [ПАПП 1966] и в Великобритании. В британской лингвистике эмпирические тенденции остались сильными. Они предполагали использование реальных примеров для проверки лингвистических гипотез [МАКЭНЕРИ–ХАРДИ 2013:18].

В 70-е годы создавались первые исследовательские центры (как например ICAME – The International Computer Archive of Modern English) для проведения корпусных исследований. Центры сотрудничали друг с другом. В результате такого сотрудничества был создан первый в мире лингвистический корпус – Брауновский корпус. Создатели данного корпуса, В.Н. Френсис и Г. Кучера, обращали внимание на такие вопросы, как сбалансированность и репрезентативность, которые до сих пор интересуют создателей корпусов.

Брауновский корпус оказал огромное влияние на развитие дисциплины: на его основе создано множество исследований по корпусной лингвистике, и своим методом отбора текстов и жанровым распределением текстов он послужил образцом для создания других корпусов.

В 1960-70-е годы корпусные исследования проводились преимущественно в британских университетах. С 1980-х годов КЛ была реабилитирована и корпусные данные дальше не воспринимались недопустимыми. В это же время такие американские университеты, как Northern Arizona University, University of California Santa Barbara, University of Michigan и Brigham Young University стали центрами научных исследований в этой области.

С 70-х годов результаты корпусных исследований начали применяться на практике. Данный процесс шел по двум главным направлениям.

Одним из основоположников в применении результатов корпусных исследований считается Рэндольф Кверк, который впервые использовал корпусные данные в области грамматических описаний. Он был создателем первой описательной грамматики английского языка, которая основывается на корпусных данных [КВЕРК и ДР. 1985]. Данная монография играла ключевую роль в эволюции КЛ.

Другое направление применения КЛ связано с именем Джона Синклера, который поставил слово, коллокации, фразеологию и анализ дискурса в центр своих корпусных исследований. В 1980-е годы его работа над проектом COBUILD революционизировала лексикографию и привела к появлению нового поколения словарей и учебных материалов, основывающихся на корпусных данных.

Создание грамматик английского языка свидетельствует об общем воздействии корпусной лингвистики на другие отрасли языкознания: она способствовала появлению разных подходов к лингвистическим данным.

Со временем стала развиваться и сама методика создания корпусов и практика их аннотирования (данный процесс связан с именем Джеффри Лича). Некоторые лингвисты в сотрудничестве с компьютерными специалистами разрабатывали возможности расширения корпусных аннотаций, развивали новые средства для редактирования и поиска данных (например, программное обес-

печение для частеречной аннотации). В 1990-е годы появились другие, современные методы аннотации.

Результаты применения корпусных методов привели к новым достижениям, новым открытиям и критической переоценке прежних работ. Применяя корпусные данные, даже удалось доказать, что представление Хомского относительно интуиции носителей языка в некоторых случаях является ошибочным [ПУЛЛУМ–ШОЛЦ: 2002]. Корпусная методика оказалась особенно плодотворной в таких сферах, как функциональная лингвистика, когнитивная лингвистика, социолингвистика, контрастивная лингвистика и психолингвистика, и в таких областях, как семантика, анализ дискурса и стилистика.

Становление корпусной лингвистики в России

Согласно данным корпусов английского языка термин *corpus linguistics* был впервые использован в 1977 году. Сама КЛ как особое направление сложилась к началу 1990-х годов. В России новый термин стал известен, по-видимому, в 1996 году благодаря лекциям одного из создателей знаменитого Международного корпуса английского языка (International Corpus of English) Сидни Гринбаума. Первая фиксация данного термина на русском языке относится к 1999 году [КОПОТЕВ 2014].

Наука о корпусах в России из-за историко-идеологических причин начала развиваться на 40 лет позже, чем в западных странах. История корпусной лингвистики тесно связана со становлением кибернетики, так как современные корпуса создаются на компьютерах.

Как известно, первый компьютер был создан в Пенсильванском университете в 1946 году. В 1948 году Норберт Винер опубликовал книгу «Кибернетика, или управление и связь в животном и машине», которая заложила основу новой науки. Однако партийное руководство СССР сделало вывод, что «некая система рассматривает в качестве сходных систем машины и животных, а в каком-то смысле также людей, и старается установить общие законы управления ими» [ШУНЕЙКО 2020: 56]. Поскольку данный взгляд шел вразрез с официальной марксистско-ленинской идеологией, новая наука попала в опалу. Положение изменилось только после смерти Сталина в 1953 г., когда выровнялось отношение государства к технологии и кибернетике. С 1960-х гг. кибернетика и связанные с ней науки начали распространяться. Это включило СССР в процесс информатизации мира, однако существенных результатов добиться не удалось [ШУНЕЙКО 2020: 57].

Появлению первых электронных русских корпусов предшествовал долгий доцифровой период, когда создавались неэлектронные корпуса уже с применением корпусных методов.

Например, еще в 1960–1970-е гг. был создан «Частотный словарь русского языка» (под рук. Л.Н. Засориной), построенный на основе примитивных текстовых файлов объемом в 1 млн словоупотреблений, включивший в себя лексику 4 жанров [ЗАХАРОВ 2013].



Основным толчком к созданию первого русского корпуса было появление «*Машинного фонда русского языка*» в 1985 г. по инициативе академика А.П. Ершова. Это был грандиозный проект с участием более 40 организаций. В задачи фонда входило накопление на машинных носителях и в базах данных текстовых, лексикографических и грамматических источников, необходимых для научного изучения русского языка и для осуществления прикладных разработок. Одновременно велось создание программных средств для проведения лингвистических исследований. Однако после 1991 г. в новых экономических условиях работы по созданию фонда постепенно стали сокращаться и наконец совсем прекратились [ЗАХАРОВ 2013].

Интересно, что первый русскоязычный корпус был создан не в России, а в 1980-е годы в Университете Уппсалы, в Швеции, с целью представить русский литературный язык. Объем «*Уппсальского корпуса русского языка*» составляет 1 млн словоупотреблений на основе 600 текстов. Уппсальский корпус сейчас входит в так называемые «*Тюбингенские корпуса русских текстов*», созданные в рамках работы специального научно-исследовательского сектора SFB 441 Тюбингенского университета в 1990–2000-е гг. с возможностью поиска онлайн [ЗАХАРОВ 2013].

Ключевым моментом в развитии корпусной лингвистики на русской почве является 2001 г., когда группа московских, петербургских и воронежских лингвистов начала работу над проектом под названием Национальный корпус русского языка (НКРЯ). Корпус содержит и письменные, и устные тексты. Кроме основного корпуса в него входят и несколько подкорпусов (диалектный, поэтический, исторический, параллельный и др.). Объем корпуса на 22 декабря 2020 г. превышал 961 млн словоупотреблений. Тексты, входящие в корпус снабжены метаразметкой, словоформы – морфологической и семантической разметкой. Спецификой корпуса является наличие семантической и синтаксической разметки и мультимедийного подкорпуса. О востребованности корпуса свидетельствуют многочисленные публикации, подготовленные на корпусе [ЗАХАРОВ 2013].

Здесь следует заметить, что кроме НКРЯ с начала нового тысячелетия появился ряд русскоязычных корпусов, как узкоспециальные, так и мегакорпус. Не претендуя на полноту, привожу некоторые из них: Хельсинкский аннотированный корпус (ХАНКО), Мультимедийный русский корпус (МУРКО), Русскоязычный эмоциональный корпус (РЕС), Санкт-Петербургский корпус агиографических текстов XV–XVII вв. (СКАТ), Корпус русского литературного языка, и мегакорпус объемом более 20 млрд. словоупотреблений – Генеральный интернет-корпус русского языка (ГИКРЯ), а также в Иркутском государственном лингвистическом университете идет работа по созданию Учебного мультимодального корпуса (УМКО).

Статус КЛ в России до сих пор оспаривается. Возникнув относительно недавно, некоторые лингвисты думают, что КЛ не имеет собственной теоретической базы, она «является улучшенной методикой сбора и обработки материала, новым информационным ресурсом» и таким образом не воспринимают КЛ самостоятельной

дисциплиной [МАЙОРОВА 2017: 42]. Данный взгляд подтверждает и тот факт, что КЛ не имеет собственного УДК-классификатора [ИЗОТОВ 2015: 85].

С развитием дисциплины начинают выходить в свет учебные пособия по корпусной лингвистике. основоположником среди учебных материалов можно считать «Введение в прикладную лингвистику» А.Н. Баранова (2001). Учебник нацелен на студентов прикладной лингвистики, поэтому в нем трактуются проблемы машинного перевода, количественной и компьютерной лингвистики, с акцентом на информатической и математической стороне дисциплины. Подобным учебным пособием, в котором рассматривается прикладная сторона области, является «Введение в прикладную лингвистику» Е.П. Сосниной (2012).

Специальные русскоязычные пособия по корпусной лингвистике появились с 2013 года. Классическим трудом считается «Корпусная лингвистика» В.П. Захарова, С.Ю. Богдановой (2013). В этом же году было издано «Введение в корпусную лингвистику» К.К. Боярского (2013).

Данные пособия определяют место КЛ среди других областей языкознания, знакомят студентов с историей возникновения и основными понятиями КЛ (аннотация, токенизация, стемминг, лемматизация, парсинг и т. д.), дают краткое представление значительных международных и русских корпусов, предлагают типологическую классификацию корпусов и также дают возможность студентам освоить навыки работы с корпусами. Конечно, каждое пособие имеет свою специфику: например, в монографии Захарова и Богдановой (2013) приведены примеры конкретных исследований в нескольких областях лингвистики (в лексикографии, грамматике и анализе дискурса). Последователем этих работ по своей структуре и тематике является учебник Б.Б. Базаровой (2016).

Инновационной и самой популярной работой является электронный учебник доцента Хельсинкского университета, М. Копотева «Введение в корпусную лингвистику» (2014). Сам формат книги хорошо подходит к корпусной тематике и способствует его известности. В данном пособии простым языком, наглядно, через множество примеров излагаются вышеприведенные традиционные вопросы. Вдобавок представляется и процедура составления собственного корпуса. Книга также снабжена практическими заданиями.

Последнее издание по корпусной тематике, вышедшее в 2020 г., монография А.А. Шунейко «Корпусная лингвистика. Учебник для вузов». Автор пособия – доцент Комсомольского-на-Амуре университета, он имеет широкий спектр научных исследований от масонской символики до изучения коммуникативных стратегий. Учебник издан в электронной и в бумажной версиях. Он логично построен в том смысле, что в первом разделе излагаются теоретические вопросы корпусной грамматики, а второй раздел посвящен практике.

В начале работы дается определение объекта учебного пособия – корпусной лингвистики. Автор оригинальным образом определяет место дисциплины среди других наук: он опирается на различие между субстанциональной и реляционной моделями познания. В лингвистике – подобно другим наукам – используются обе модели: а) субстанциональная, когда анализируется формальная

сторона языковой единицы, ее сочетаемостные характеристики, формальные изменения или частотность употребления, б) реляционная – при описании значения, функциональной стороны единицы [ШУНЕЙКО 2020: 15]. До сих пор не удалось совместить результаты, полученные в рамках отдельных моделей, между собой. Корпусная лингвистика ассоциируется, в первую очередь, с субстанциональной моделью, однако ее применение может снять противоречие между данными познавательными моделями.

В третьей части рассматривается понятие корпуса и его структура. Объясняются функция корпусного менеджера и процедура поиска.

В четвертой части излагается процедура создания корпусов от стадии отбора текстов до конвертирования выверенного текста в структуру информационно-поисковой системы. Поскольку наличие разметки отличает корпус от общего собрания текстов, особое внимание здесь уделяется объяснению таких процедур, как токенизация, лемматизация, стемминг и парсинг. Также перечисляются разные виды разметки.

В пятой части дается подробная классификация корпусов по разным основаниям.

Шестая часть посвящена одним из основных законов КЛ – закону рангового распределения. Закон был сформулирован задолго до возникновения КЛ и применяется и в других областях науки. Относительно лингвистики закон устанавливает, что «если возьмем слова текста, и располагаем их по убыванию частоты их использования в данном тексте, то окажется, что частота каждого слова в списке приблизительно обратно пропорциональна его порядковому номеру (рангу) на линии убывания частот» [ШУНЕЙКО 2020: 36]. На практике это означает, что второе слово списка встречается примерно в два раза реже, чем первое, четвертое слово – в 4 раза реже и т. д.

Любой текст содержит элементы структуры и элементы информации. Элементов структуры в общем инвентаре языка меньше, а элементы информации в разных текстах меняются. Однако элементы структуры повторяются, их частотность во многом раз превышает частотность элементов информации. Закон является важным инструментом количественного исследования языка. Действие закона приводит к тому, что частотность слов в языке крайне неравномерна: одни слова превосходят в частотности другие. Самые частотные слова образуют ядро/центр текстов. Заслуживает внимания и то интересное следствие закона, что 10 тысяч самых частотных лемм русского языка покрывают приблизительно 85% всех текстов, однако они составляют всего 20% общего объема слов. [ШУНЕЙКО 2020: 129]

Следующие две части освещают историю возникновения корпусной лингвистики в мире и в России.

В результате возникновения корпусной лингвистики были созданы первые корпуса (Брауновский корпус, Ланкастерский корпус английского языка или корпус LOB) Развитие технологии хранения и извлечения текстов с 90-ых гг. XX века сделало возможным появление второго поколения корпусов. Они отличаются от корпусов первого поколения своим объемом (содержат уже более

100 млн словоупотреблений) и поисковыми возможностями. Сюда относятся The Longman Corpus, British National Corpus. В наше время развивается третье поколение корпусов, которые представляют собой мегакорпуса. Их объем превышает миллиард словоупотреблений, и они предлагают уникальные возможности поиска. Такими корпусами являются iWeb – The Intelligent Web-based Corpus, или The Collins Corpus.

В традиционном понимании корпуса служат для хранения и исследования текстов, однако в наши дни корпуса стали использоваться в самых различных областях жизни. В десятой части монографии перечисляются основные сферы их применения: кроме таких общепринятых областей, какими являются научные исследования, статистика языка и речи или машинный перевод, здесь описываются и такие практические цели их употребления, как дешифровка текстов, генераторы текстов, анализ общественного мнения или же выявление плагиата.

Одиннадцатая часть работы посвящена терминологии корпусной лингвистики. Здесь перечислены термины методологического, технического и аналитического плана в алфавитном порядке, с английскими эквивалентами.

В шестой части уже рассматривался один из основных законов корпусной грамматики, который был сформулирован до появления первых корпусов – закон рангового распределения. В двенадцатой части перечисляются такие правила, или сильные утверждения, которые возникли после появления корпусов, и которые связаны с функционированием информации: закон Парето (или закон 20/80), реализацией которого является и закон рангового распределения, закон Бенфорда (закон рассеивания информации), который породил индекс научного цитирования, или Индексы Гринберга, которые являются средствами для установления формальных различий между языками благодаря подсчету соотношения между определенными лингвистическими единицами [ШУНЕЙКО 2020: 134] и коэффициент точности и вольности, предложенный М. Л. Гаспаровым и В. В. Настопкене, которые применяются в сфере машинного перевода.

Отношение традиционной и корпусной лингвистики трактуется в тринадцатой части пособия. Они сосуществуют и развиваются параллельно. Традиционная лингвистика опирается на теорию и часто ограничивается небольшим количеством речевого материала, корпусная лингвистика работает с огромным объемом данных. Однако имеются сферы (диалектология, лексикография, лексикология, статистика), где две области пересекаются. Применение корпусных материалов позволяет выявлять самые разные факты бытования языка, и таким образом доказать или опровергнуть теоретические предположения.

В последней части описываются проблемные области и ближайшие перспективы развития корпусной лингвистики. Кроме долгое время сопровождающих развитие науки проблем (как проблема разметки, снятия омонимии, сбалансированности, возрастающего объема) здесь подняты такие проблемы, как разрыв между теоретической и прикладной лингвистикой (т. е. большинство теоретических исследований до сих пор является corpus based, а не corpus driven), или проблема интерпретации корпусных данных.

Второй раздел монографии содержит методические и практические рекомендации и материалы для преподавания корпусной лингвистики, которые могут пригодиться преподавателям вузов: рабочую программу дисциплины, пример контрольной работы, планы практических занятий, творческие задания, список образовательных видеоматериалов и список литературы.

В учебнике собраны результаты последних исследований в изучаемой сфере. Вопросы корпусной лингвистики, даже сложные теоретические, математические или статистические проблемы изложены на простом, доступном для студентов языке.

Книга по своей структуре соответствует установленным образовательным целям, так как в конце каждой части первого раздела имеется блок «Проверь себя» с доступом к онлайн-тестам с двумя попытками решения. Тесты дают быструю обратную связь с оценкой успешности усвоения материала.

Очень полезными (особенно в режиме онлайн-обучения) оказываются видеоматериалы, которые хорошо дополняют и «сопровождают» блоки учебника.

Непросто обнаружить недостатки учебника, но следует отметить, что полезно было бы включить больше наглядных и практических примеров для представленных явлений, а также среди заданий книги отсутствуют задания на поиск.

О сегодняшнем положении КЛ в России можно сказать, что, хотя наука о корпусах на Западе начала развиваться на 40 лет раньше, чем в России, развитие КЛ происходит по аналогичным принципам и стратегиям [ИРГИЗОВА 2019]. В последнее время активные исследования по русской корпусной грамматике введутся в Институте русского языка им. В. В. Виноградова РАН (см. на сайте <http://rusgram.ru/about>). Целью их проекта является описание русской грамматики на основе данных НКРЯ. С 2016 г. издается серия публикаций под названием «Материалы к корпусной грамматике русского языка» (см. <https://publications.hse.ru/books/199980981>, <http://www.ruslang.ru/node/1384>, <http://ruslang.ru/node/1623>), где каждый выпуск посвящен отдельным сферам грамматики русского языка (таким, как частеречная проблематика, глагол, синтаксические конструкции, и т. д.).

Разница бросается в глаза, если посмотрим на современную учебную литературу по КЛ: на Западе ее составляют узкоспециальные сборники статей, в то время, как в России пока издаются только учебные материалы и отдельные статьи по данной области. Однако возрастающее количество и разнообразие современных русских корпусов дают повод для оптимизма.

Библиография

- БАРАНОВ, А.Н. [BARANOV, A.N.] 2001: Введение в прикладную лингвистику. Москва: Эдиториал УРСС [Introduction to Applied Linguistics. Moscow: Editorial URSS].
- ЗАХАРОВ, В.П.–БОГДАНОВА, С.Ю. [ZHAROV, V.P.–BOGDANOVA, S.Û.] 2013: Корпусная лингвистика Санкт-Петербург: СПбГУ РИО Филологический Факультет [Corpus linguistics St. Petersburg: St. Petersburg State University RIO Faculty of Philology].

- ЗАХАРОВ, В.П. [ZHAROV, V.P.] 2013: Корпусна 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.) [Corpus linguistics in Russia. Report at the conference: IV International Scientific Symposium Retro-2013. Retrospective of philology in the information society of knowledge (Ukraine, Crimea, Peschanoe, June 27 – July 3, 2013)] // 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).
- ЖУЛЯНД, А.–ЧАНГ-РОДРИГЕЗ, Е. [JUILLAND, A.–CHANG-RODRIGUEZ, E.] 1964: A Frequency Dictionary of Spanish Words. The Hague: Mouton.
- ИЗОТОВ, А.И. [IZOTOV, A.I.] 2015: Новые направления славянского языкознания: корпусная лингвистика // Язык, сознание, коммуникация. Москва [New directions of Slavic linguistics: corpus linguistics // Language, co-knowledge, communication. Moscow], 82–93.
- ИРГИЗОВА, К.В. [IRGIZOVA, K.V.] 2019: Корпусная лингвистика в отечественном и зарубежном языкознании на современном этапе. // Журнал Огарёв [Corpus linguistics in Russian and foreign linguistics at the modern stage. // Magazine Ogarev] – Online (дата обращения: 13.03.2021).
- КВЕРК И ДР. [QUIRK ET AL.] 1985: Quirk, R., Greenbaum, S., Leech, G., Svartvik J. A Comprehensive Grammar of the English Language. London: Longman.
- КОПОТЕВ, М. [KOPOTEV, M.] 2014: Введение в корпусную лингвистику. Прага: Animedia Company [Introduction to Corpus Linguistics. Prague: Animedia Company].
- КУЧЕРА, Х.–ФРЕНСИС, В.Н. [KUČERA, H., FRANCIS, W.N.] 1967: Computational Analysis of Present Day American English. Providence: Brown University Press.
- МАЙОРОВА, А.Д. [MAJOROVA, A.D.] 2017: Корпусная лингвистика: исторический и лингводидактический аспекты // Международный научно-исследовательский журнал, Екатеринбург [Corpus linguistics: historical and linguodidactic aspects // International scientific research journal, Ekaterinburg], №. 5 (59), 42–46.
- МКЭНЕРИ, Т.–ХАРДИ, А. [MCENERY, T. – HARDIE, A.] 2013: The History of Corpus Linguistics // The Oxford Handbook of the History of Linguistics Ed. Allan, K. Oxford: OUP.
- ПАПП, Ф. [PAPP, F.] 1966: Mathematical Linguistics in the Soviet Union. // Janua Linguarum, Series Minor, XL. The Hague: Mouton.
- ПУЛЛУМ, Г.К.–ШОЛЦ, Б.Ц. [PULLUM, G.K. – SCHOLZ, B.C.] 2002: Empirical assessment of stimulus poverty arguments. // Linguistic Review 19: 9–50.
- ШАРОВ, С.А. [ŠAROV, S.A.] 2003: Представительный корпус русского языка в контексте мирового опыта // Научно-техническая информация. – Сер. 2.– № 6: 12–16 [Representative corpus of the Russian language in the context of world experience // Scientific and technical information. – Ser. 2.– No. 6: 12–16].
- ХОМСКИЙ, Н. [CHOMSKY, N.] 1965: Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.

Беата ДЬЕРФИ
Сегедский университет, Институт славистики
Сегед, Венгрия
blazsenyuka@yahoo.com

