

## Лингводидактические свойства корпусных технологий

**Питирим Юрьевич ЗОЛОТОВ**

ФГБОУ ВО «Тамбовский государственный университет им. Г.Р. Державина»  
392000, Российская Федерация, г. Тамбов, ул. Интернациональная, 33  
ORCID: <https://orcid.org/0000-0002-3886-1952>, e-mail: [pitirim93@gmail.com](mailto:pitirim93@gmail.com)

## Linguodidactic properties of corpus technologies

**Pitirim Y. ZOLOTOV**

Derzhavin Tambov State University  
33 Internatsionalnaya St., Tambov 392000, Russian Federation  
ORCID: <https://orcid.org/0000-0002-3886-1952>, e-mail: [pitirim93@gmail.com](mailto:pitirim93@gmail.com)

**Аннотация.** В последние два десятилетия корпусные технологии, понимаемые как совокупность средств и методов обработки и анализа данных электронных лингвистических корпусов, как вид информационно-коммуникационных технологий, вызывают большой интерес исследователей и преподавателей иностранных языков. Дано объяснение понятиям «корпусная лингвистика», «корпусные технологии», «лингвистический корпус», «конкорданс». Рассмотрены методы изучения корпусных технологий, представляющие из себя аннотацию, абстракцию и анализ. Приведены достоинства лингвистических корпусов. Описана история возникновения и развития лингвистических электронных корпусов от доцифрового до цифрового периода. Приведены минимальные требования к корпусу текстов, включающие репрезентативность, известный объем корпуса, электронную форму, аннотацию и сбалансированность. Рассмотрена типология лингвистических корпусов. По языку текстов различают одноязычные и многоязычные корпуса, которые, в свою очередь, разделяются на смешанные и параллельные. По языковым данным бывают письменные, устные и смешанные. Корпуса бывают неразмеченными и размеченными. Разметка же бывает трех видов: лингвистическая, метатекстовая и экстралингвистическая. По параметру представленности языкового материала корпуса бывают фрагментированные и нефрагментированные. По типу доступа они классифицируются на открытые и ограниченные. По жанровому представлению лингвистические корпуса разнообразны. По объему следует различать представительные, иллюстративные и мониторинговые. Изучены дидактические свойства корпусных технологий в области обучения иностранному языку. Предложено деление лингводидактических свойств корпусных технологий на обязательные и факультативные.

**Ключевые слова:** корпусная лингвистика; корпусные технологии; конкорданс; лингвистический корпус; лингводидактические свойства корпусных технологий

**Для цитирования:** Золотов П.Ю. Лингводидактические свойства корпусных технологий // Вестник Тамбовского университета. Серия: Гуманитарные науки. Тамбов, 2020. Т. 25, № 185. С. 75-82. DOI 10.20310/1810-0201-2020-25-185-75-82

**Abstract.** For the last two decades, corpus technologies, understood as a combination of means and methods of processing and analyzing data of electronic linguistic corpora, as a type of information and communication technology, have attracted great interest of researchers and teachers of foreign languages. We explain the concepts of corpus linguistics, corpus technology, linguistic corpus, concordance. The methods of studying case technologies, which are an annotation, abstraction, and analysis, are considered. The advantages of linguistic corpora are given. The history of the emergence and development of linguistic electronic cases from the pre-digital to digital period is described. Minimum requirements for the corpus of texts are presented. They include representativeness, known volume of the corpus, electronic form, annotation and balance. We consider the typology of linguistic corpora. According to the language of the texts in corpora, there are monolingual and multilingual corpora, which in turn are divided into mixed and parallel ones. Accord-

ing to language data, there are written, oral and mixed corpora. Corpora can be annotated and non-annotated. There are three types of annotation: linguistic, metatextual, and extralinguistic. According to the parameter of representation of the language material of a corpus, there are fragmented and non-fragmented ones. According to the type of access, they are classified as open and restricted. According to the genre representation, linguistic corpora are diverse. The size of a corpus should distinguish between representative, illustrative and monitoring types of corpora. The didactic properties of corpus technologies in the field of teaching a foreign language are studied. The division of the linguodidactic properties of case technologies into mandatory and optional is proposed.

**Keywords:** corpus linguistics; corpus technologies; concordance; linguistic corpus; linguodidactic properties of corpus technologies

**For citation:** Zolotov P.Y. Linguodidakticheskiye svoystva korpusnykh tekhnologiy [Linguodidactic properties of corpus technologies]. *Vestnik Tambovskogo universiteta. Seriya: Gumanitarnye nauki – Tambov University Review. Series: Humanities*, 2020, vol. 25, no. 185, pp. 75-82. DOI 10.20310/1810-0201-2020-25-185-75-82 (In Russian, Abstr. in Engl.)

## ОПРЕДЕЛЕНИЕ ПОНЯТИЙ

С начала двадцатого века корпусным технологиям было уделено особо пристальное внимание со стороны как отечественных ученых, так и зарубежных исследователей [1–5]. Корпусные технологии появились в *корпусной лингвистике*, разделе языкознания, который разрабатывает, создает и использует различные лингвистические текстовые корпуса. Корпусная лингвистика предлагает, что более точный анализ языка более осуществим с корпусами, собранными в естественном контексте, а также с минимальным экспериментальным вмешательством.

Корпусная лингвистика является своего рода инструментом для других разделов языкознания, позволяющим проводить необходимые исследования в более удобной обстановке, обеспечивая более точными полученными результатами. В дополнение к лингвистическим исследованиям собранные корпуса используются для составления словарей и грамматических справочников.

В корпусной лингвистике выделяется набор методов исследования, которые стараются определить то, как корпусная лингвистика прошла путь от экспериментов к теоретическому фундаменту, проследить путь от данных к теории. Учеными-лингвистами С.А. Уоллисом и Г. Нельсоном в 2001 г. была предложена так называемая перспектива трех «А». Здесь имеются в виду три заглавные буквы следующих понятий: аннотация, абстракция и анализ [6].

Первый из этих терминов, *аннотация* (иначе называемая *разметкой*), заключается в том, что к текстам применяется определен-

ная схема, в которой отдельные элементы текста особым образом маркируются. Следующий этап, *абстракция*, предполагает некий переход от терминологической схемы в теоретическую модель. При абстракции зачастую проводится лингвистическое исследование. *Анализ* является этапом, где используются методы статистики, происходит работа с данными.

У лингвистических корпусов есть ряд достоинств. В 1992 г. Я. Свартвик выделил следующие их преимущества:

- данные корпуса более объективны, чем данные, основанные на самоанализе;
- данные корпуса могут быть легко проверены другими исследователями; также они могут совместно использовать одни и те же данные вместо того, чтобы всегда компилировать свои собственные;
- данные корпуса необходимы для изучения различий между диалектами, регистрами и стилями;
- данные корпуса обеспечивают частоту возникновения лингвистических явлений;
- данные корпуса не только служат иллюстративными примерами, но и теоретическим ресурсом;
- данные корпуса дают важную информацию для ряда областей применения, таких как преподавание языка и лингвистические технологии (машинный перевод, синтез речи и т. д.);
- корпорации обеспечивают возможность полной ответственности лингвистических особенности: аналитик должен учитывать все в данных, а не только избранные функции;

– компьютеризированные корпорации предоставляют исследователям во всем мире доступ к данным;

– корпусные данные идеально подходят для не носителей языка [7].

Тем не менее существуют несколько проблем в лингвистике корпусов, которым ученые в данный момент ищут решение. Среди них стоит выделить проблему репрезентативности, проблему разметки и проблему представления результатов. Что касается проблемы репрезентативности, то причины ее возникновения можно узреть вследствие закона Ципфа. Согласно этому закону, в любом лингвистическом корпусе частота любого словоупотребления обратно пропорциональна его рангу в таблице частот. Это означает, что редко используемых слов в языке намного больше, чем часто используемых. Соответственно объем корпуса, а также выборка текстов важны для адекватного представления конкретного языка или подязыка.

П.В. Сысовым было предложено определение лингвистическому корпусу. Автор интерпретирует его следующим образом: «массив текстов, собранных в единую систему по определенным признакам (языку, жанру, времени создания текста, автору и т. п.) и снабженных поисковой системой» [2, с. 99]. Лингвистическим корпусом, иначе говоря, является собрание текстов в электронном виде, отобранных по определенным признакам и доступных для качественного и количественного анализа.

*Корпусные технологии* – это совокупность средств и методов обработки и анализа данных электронных лингвистических корпусов.

*Конкорданс* – это один из основных инструментов в лингвистике корпусов, который предполагает использование программного обеспечения корпусов для поиска появления определенного слова или фразы. Более подробно этот инструмент будет разобран далее.

## ИСТОРИЯ РАЗВИТИЯ КОРПУСНОЙ ЛИНГВИСТИКИ

Считается, что корпусная лингвистика как раздел языкознания появилась в 60-е гг. XX века. Тем не менее, нельзя не отметить, что корпусная лингвистика имеет долгую предысторию и предпосылки к своему появ-

лению. Конечно, до изобретения вычислительных машин существование электронных корпусов было невозможно, однако в бумажном виде большие объемы текстов существовали задолго до конца второго тысячелетия нашей эры. Доцифровой период в развитии корпусной лингвистики насчитывает несколько веков. Изначально нецифровые корпуса были тесно связаны с религиозными текстами. В качестве примера могут служить симфонии (или конкорданции), являющиеся по своей сути списками слов с указанием стихов из Библии. В дальнейшем появлялись крупные словари, созданные на основе карточек. Данный этап в истории развития корпусной лингвистики можно назвать лексикографическим. Уже на рубеже XX века начинают появляться собрания текстов для лингвистического анализа.

Цифровой этап в развитии корпусной лингвистики начинается с появлением электронных корпусов первого поколения. Самым первым его представителем стал Брауновский корпус (полное оригинальное название – Brown University Standard Corpus of Present-Day American English), авторами которого считаются Н. Френсис и Г. Кучера. Именно этот корпус стал основоположником современной корпусной лингвистики и заложил базовые принципы создания электронных корпусов текстов. Корпус состоит из миллиона слов американских английских текстов, напечатанных в 1961 г. Чтобы сделать корпус сбалансированным, тексты были отобраны в разных пропорциях из 15 различных текстовых категорий: публицистика, тексты бытовой тематики, религиозной, научные тексты, различные типы художественной литературы и т. д.

Составители Брауновского корпуса руководствовались четырьмя критериями для отбора его содержания:

– *происхождение автора и состав текста* (только представители американского варианта английского языка);

– *возможность обработки данных с помощью компьютера*;

– *синхронизация* (создание корпуса началось в 1961 г., поэтому были отобраны тексты из этого года публикации);

– *численное соотношение жанрового разнообразия текстов*.

Сегодня этот корпус считается небольшим и уже устаревшим. Тем не менее, корпус все еще используется. Большая часть его полезности заключается в том, что структура Брауновского корпуса была скопирована другими составителями корпусов. Корпус LOB (Lancaster-Oslo-Bergen; язык – британский английский) и корпус Kolhapur (индийский английский) являются двумя примерами корпусов, выполненных в соответствии с Брауновским корпусом. Оба они состоят из 1 миллиона слов письменного языка (500 текстов по 2000 слов каждый), которые были отобраны в тех же 15 категориях, что и Брауновский корпус.

Доступность корпусов, которые настолько схожи по структуре, является ценным ресурсом, например, для исследователей, заинтересованных в сравнении разных вариантов языка. В течение долгого времени Брауновский корпус и корпус LOB были единственными доступными для компьютерной обработки. Поэтому многие исследования в области корпусной лингвистики были основаны на этих корпусах.

Также к первому поколению корпусов относится London-Lund Corpus (LLC). Являясь также небольшим по объему, он был первым корпусом разговорного языка. Он состоит из 100 устных текстов около 5000 слов каждый. Тексты подразделяются на разные категории, такие как спонтанный разговор, спонтанный комментарий, спонтанная и подготовленная речь и т. д. Тексты имеют транскрипцию, а также они содержат подробный просодический разбор речи.

Что касается истории отечественной корпусной лингвистики, то была попытка создания так называемого Машинного Фонда русского языка. Работа по созданию корпуса началась в 1985 г. в Институте русского языка АН СССР. К сожалению, создание корпуса не удалось завершить из-за проблем с финансированием в начале 1990-х гг. Однако корпус русского языка все же был создан в то время. Правда, случилось это на территории Швеции в городе Уппсалы. Уппсальский корпус русского языка, созданный в институте славистики г. Уппсалы, содержал 1 миллион словоупотреблений и порядка 600 текстов.

Электронные корпуса второго поколения приходят в 1990-е гг., когда компьютерные

технологии становятся более совершенными. Среди первых представителей этого периода можно отметить корпус The Cobuild Project/The Bank of English (BoE). Это британский мониторинговый корпус, постоянно пополняющий объем словоупотреблений. 25 % корпуса составляет устная речь, а 75 % – письменная.

Одним из самых значимых корпусов, который часто используется исследователями, является Британский Национальный корпус (часто используемая аббревиатура на английском – BNC). В корпус входят 100 миллионов слов. Как и британский мониторинговый корпус содержит как письменный, так и речевой материал, но в отличие от BoE, Британский Национальный корпус является конечным, то есть после завершения его составления не добавляется больше текстов. Тексты Британского Национального корпуса были выбраны в соответствии с тщательно определенными критериями, чтобы сделать корпус сбалансированным. Тексты были закодированы с разметкой, предоставляющей информацию о текстах, авторах, говорящих.

Позднее были созданы и другие национальные корпуса, в том числе и Американский Национальный корпус (American National Corpus), и Национальный корпус русского языка, содержащие миллионы словоупотреблений.

## КРИТЕРИИ И ТИПЫ ЛИНГВИСТИЧЕСКИХ КОРПУСОВ

Современные лингвистические корпуса должны отвечать ряду критериев или параметров. Первым из таких критериев является *репрезентативность*. Это означает, что корпус должен быть достоверно представительным. Это достигается за счет нужного объема и жанрового разнообразия текстов. Другим важным критерием является *сбалансированность*. Сбалансированный корпус имеет равномерное распределение текстов различных категорий. Важно также и знание *точного объема корпуса*. Это необходимо для исследователей, проводящих количественные исследования. *Электронная форма* – четвертый параметр. Электронный формат представления текста значительно облегчил обработку информации. Наконец, пятым и последним критерием является *разметка* (так-

же используется термин *аннотация*). Аннотация – это «введенная автоматически или вручную лингвистическая или метатекстовая информация обо всех выбранных единицах корпуса: тексте, предложении, текстоформе, морфеме, звуке и т. д.» [8, с. 30].

С течением времени появлялись разные типы корпусов для различных целей. Существуют различные типы классификаций корпусов. Во-первых, классификация корпусов *по языку текстов* предлагает деление корпусов на одноязычные, где все тексты одного корпуса написаны на одном языке, и многоязычные, где тексты корпуса написаны на двух и более языках. Многоязычные корпуса принято делить на смешанные и параллельные. Первые включают в себя тексты, не являющиеся переводом друг друга. Параллельные же корпуса содержат тексты-оригиналы и тексты-переводы. Параллельный корпус также обладает свойством выравнивания, что означает, что тексты и их переводы связаны по предложениям и абзацам. Особым типом корпуса в данной классификации является сравнительный корпус, в котором помимо текста-оригинала присутствуют несколько текстов-переводов. Примером такого корпуса может служить разрабатываемый в Университете Мэриленд в США корпус переводов Библии, насчитывающий несколько тысяч переводов.

По типу языковых данных можно различить три вида лингвистических корпусов. В первую очередь это устные, где содержатся записи и транскрипты устной речи. Наибольшее число корпусов причисляется к письменным. Также существуют смешанные корпуса, которые включают как письменную, так и устную речь, причем, как правило, письменной речи отведена большая доля, чем устной.

Корпусы могут быть как размеченными, так и неаннотированными (то есть без разметки). Характер разметки также является классифицирующим параметром. Разметка бывает (1) метатекстовая, содержащая паспорт текста (информация об авторе и тексте), (2) лингвистическая, которая бывает синтаксической, семантической, морфологической и т. п., и (3) экстралингвистическая, которая содержит информацию о жестах и иных сопровождающих невербальных знаках.

Лингвистические корпуса также могут различаться по жанровому разнообразию. Они могут охватывать весь спектр жанров от разговорных до научных. Также корпуса различаются по объему на представительные (иначе национальные), мониторинговые и иллюстративные. По типу доступа лингвистические корпуса следует разделять на открытые и те, где доступ ограничен. Как правило, последние являются платными.

По представленности языкового материала лингвистические корпуса разделяются на полнотекстовые и n-граммные, или фрагментированные. Суть последних заключается в том, что текст в таких корпусах разбит на небольшие участки, называемыми граммами с целью удобства работы с ними.

Некоторые инструменты, предоставляемые корпусами, широко используются в практике обучения иностранным языкам. Наиболее распространенным программным обеспечением для анализа корпусов является *конкорданс*. Этот инструмент извлекает примеры слов или тегов (или последовательности слов/тегов) и представляет их пользователю.

Конкорданс является основным инструментом в лингвистике корпусов, который предполагает использование программного обеспечения корпусов для поиска появления определенного слова или фразы. Эта идея не является новой, и многие ученые на протяжении многих лет вручную делали, например, конкордансы христианской Библии, кропотливо находили и записывали каждый пример определенных слов (симфонии). С помощью компьютера мы теперь можем искать миллионы слов за считанные секунды. Поисковое слово или фраза часто называется «узлом», а линии конкорданса обычно представлены узловым словом/фразой в центре строки с семью или восемью словами, представленными с обеих сторон. Они известны как дисплеи Key-Word-In-Context (KWIC). Линии конкорданса обычно сканируются вертикально, то есть просматриваются сверху вниз или снизу вверх, ориентируясь на узловое слово или фразу, которые расположены по центру. Первоначально это может показаться неудобным, потому что мы привыкли читать слева направо. Линии конкорданса позволяют нам читать совершенно по-новому, вертикально или даже от центра в стороны в обоих направлениях.

## ДИДАКТИЧЕСКИЕ СВОЙСТВА КОРПУСНЫХ ТЕХНОЛОГИЙ

Чтобы рассмотреть дидактические свойства корпусных технологий, обратимся к предложенной Ю.И. Семич группе свойств [4]. Ю.И. Семич, отталкиваясь от общих дидактических свойств ИКТ, выделенных П.В. Сысоевым, а также дидактических свойств параллельных корпусов А.А. Кокоревой, предлагает дидактические свойства, характерные конкретно корпусным технологиям.

П.В. Сысоев предлагает десять общих дидактических свойств компьютерных технологий [9]. Первым свойством является многоуровневость информационных интернет-ресурсов. Вторым свойством являются интернет-ресурсы, которые обладают большим разнообразием функциональных типов. Третьим неотъемлемым свойством корпусных технологий является мультимедийность. Под четвертым свойством предлагается то, что документы в компьютерных технологиях могут обладать гипертекстовой структурой. Пятое свойство – это возможность сооружения некой персональной зоны пользователя технологий. Шестым свойством является то, что в Интернете допустимо совершать коммуникацию как синхронно, так и асинхронно. Седьмым дидактическим свойством компьютерных технологий автор указывает возможность того, что управление учебной деятельностью учеников и информационно-методическое обеспечение могут быть автоматизированы. Восьмым свойством является возможность следования персональной образовательной траектории. Девятым свойством провозглашается «кооперация» в качестве педагогической технологии. И последним свойством является достижение относительной автономии обучающихся в учебной деятельности. (П.В. Сысоев) [9].

А.А. Кокорева выделила три дидактические свойства параллельных корпусов [3]. Первое из них – это многоязычие, то есть в параллельных корпусах присутствуют тексты как минимум на двух разных языках. Второе свойство – это контекстовая обусловленность результатов поиска. И последним свойством является то, что область применения языка можно лимитировать.

Таким образом, обобщая вышеуказанное, Ю.И. Семич выделяет семь дидактических свойств корпусных технологий [4]. Разберем каждое дидактическое свойство по очереди, определив обязательность и факультативность каждого из свойств, так как, по нашему мнению, не все выделенные дидактические свойства корпусов. Многоязычие предполагает, что некоторые электронные лингвистические корпуса включают параллельные тексты на разных языках. Однако такое свойство характерно только для параллельных корпусов. Большинство национальных электронных корпусов представлены лишь на одном языке, то есть являются моноязычными. Поэтому стоит признать данное свойство факультативным, но не безоговорочно обязательным.

Что касается контекстности, то это свойство является одним из основополагающих в корпусных технологиях. Поиск слов, представление их в контексте (в узком и широком смысле) – все это является неотъемлемой составляющей работы с корпусами.

Динамичность также не является обязательным дидактическим свойством корпусных технологий. В некоторых случаях динамичность корпуса является даже минусом, а не положительной его чертой. Постоянное обновление текстовых баз корпуса характерно, в частности, мониторинговым корпусам, таким как The Cobuild Project. Зачастую в подобных корпусах тексты неадекватно представляют конкретный язык, что создает несбалансированность корпуса. Один из наиболее значимых на данный момент корпусов, к примеру, Британский Национальный Корпус (BNC), является статичным.

Возможность осуществления поиска в корпусе по заданному шаблону (шаблонный поиск) является необходимым дидактическим свойством. Возможность выбора функциональных типов текстов корпуса можно объединить с предыдущим и последующим свойством, так как они дублируют друг друга во многом. Сортировка результатов поиска, действительно, удобная функция корпусных технологий. Сетевой доступ характерен для большинства информационных технологий. Таким образом, табл. 1 представляет дидактические свойства корпусных технологий, разбитые на основные и факультативные.

Таблица 1

## Дидактические свойства корпусных технологий

Дидактические свойства	
Основные	Факультативные
1. Контекстность	1. Многоязычность
2. Шаблонный поиск	2. Динамичность
3. Сортировка результатов поиска	
4. Сетевой доступ	

## ЗАКЛЮЧЕНИЕ

Таким образом, мы рассмотрели определения корпусной лингвистики, электронного лингвистического корпуса, корпусных технологий. Привели краткую историческую справку о развитии корпусов и корпусной

лингвистики. Кроме этого, мы рассмотрели свойства, а также типы лингвистических корпусов. Далее были изучены дидактические свойства корпусных технологий. Было предложено разделение дидактических свойств на обязательные и факультативные.

## Список литературы

1. Рыков В.В. Прагматически ориентированный корпус текстов // Диалог-99: тр. Междунар. семинара. Москва; Таруса, 1999. С. 211-220.
2. Сысоев П.В. Лингвистический корпус в методике обучения иностранным языкам // Язык и культура. 2010. № 1. С. 99-111.
3. Кокорева А.А. Методика обучения студентов профессиональной лексики на основе корпуса параллельных текстов (английский язык, направление подготовки «Экономика»): автореф. дис. ... канд. пед. наук. М.: МГГУ им. М.А. Шолохова, 2013.
4. Семич Ю.И. Дидактические свойства и методические функции корпусных технологий // Вестник Тамбовского университета. Серия: Гуманитарные науки. Тамбов, 2018. Т. 23. № 177. С. 39-49. DOI 10.20310/1810-0201-2018-23-177-39-49
5. McEney T., Xiao R., Tono Y. Corpus based Language Studies. N. Y.: Routledge, 2007.
6. Wallis S.A., Nelson G. Knowledge discovery in grammatically analysed corpora // Data Mining and Knowledge Discovery. 2001. Vol. 5(4). P. 305-336.
7. Directions in Corpus Linguistics. Proceedings of Nobel Symposium / by ed. J. Svartvik. 1991. Berlin: Mouton de Gruyter, 1992. 487 p.
8. Кононев М.В. Введение в корпусную лингвистику. Прага: Animedia, 2014.
9. Сысоев П.В. Современные информационные и коммуникационные технологии: дидактические свойства и функции // Язык и культура. 2012. № 1 (17). С. 120-133.

## References

1. Rykov V.V. Pragmatically oriented corpus of texts [Pragmatically oriented corpus of texts]. *Trudy Mezhdunarodnogo seminar «Dialog-99»* [Proceedings of International Seminar "Dialogue-99"]. Moscow, Tarusa, 1999, no. 1, pp. 89-96. (In Russian).
2. Sysoyev P.V. Linguistic corpus in teaching foreign languages. *Yazyk i kul'tura – Language and Culture*, 2010, no. 1 (9), pp. 99-111. (In Russian).
3. Kokoreva A.A. *Metodika obucheniya studentov professional'noy leksike na osnove korpusa parallel'nykh tekstov (angliyskiy yazyk, napravleniye podgotovki «Ekonomika»): avtoref. dis. ... kand. ped. nauk* [Methods of Teaching Professional Lexicon to Students on the Basis of Parallel Texts Corpus (the English Language, "Economics" Programme). Cand. ped. sci. diss. abstr.]. Moscow, Sholokhov Moscow State University for Humanities, 2013. (In Russian).
4. Semich Y.I. Didakticheskiye svoystva i metodicheskiye funktsii korpusnykh tekhnologiy [Didactic properties and methodical functions of corpus technologies]. *Vestnik Tambovskogo universiteta. Seriya: Gumanitarnye nauki – Tambov University Review. Series: Humanities*, 2018, vol. 23, no. 177, pp. 39-49. DOI 10.20310/1810-0201-2018-23-177-39-49. (In Russian).

5. McEnery T., Xiao R., Tono Y. *Corpus Based Language Studies*. New York, Routledge Publ., 2007.
6. Wallis S.A., Nelson G. Knowledge discovery in grammatically analysed corpora. *Data Mining and Knowledge Discovery*, 2001, vol. 5(4), pp. 305-336.
7. Jan Svartvik (ed.). Directions in Corpus Linguistics. *Proceedings of Nobel Symposium 82 Stockholm, 4–8 August 1991*. Berlin, Mouton de Gruyter Publ., 1992, 487 p.
8. Kopotev M.V. *Vvedeniye v korpusnuyu lingvistiku* [Introduction to Corpus Linguistics]. Praha, Animedia Publ., 2014.
9. Sysoyev P.V. Sovremennyye informatsionnyye i kommunikatsionnyye tekhnologii: didakticheskiye svoystva i funktsii [Modern information and communication technologies: didactic characteristics and functions]. *Yazyk i kul'tura – Language and Culture*, 2012, no. 1 (17), pp. 120-133. (In Russian).

#### Информация об авторе

**Золотов Питирим Юрьевич**, ассистент кафедры лингвистики и гуманитарно-педагогического образования. Тамбовский государственный университет им. Г.Р. Державина, г. Тамбов, Российская Федерация.  
E-mail: pitirim93@gmail.com

**ORCID:** <https://orcid.org/0000-0002-3886-1952>

Поступила в редакцию 13.02.2020 г.  
Поступила после рецензирования 10.03.2020 г.  
Принята к публикации 20.03.2020 г.

#### Information about the author

**Pitirim Y. Zolotov**, Assistant of Linguistics and Humanitarian Pedagogic Education Department. Derzhavin Tambov State University, Tambov, Russian Federation.  
E-mail: pitirim93@gmail.com

**ORCID:** <https://orcid.org/0000-0002-3886-1952>

Received 13 February 2020  
Reviewed 10 March 2020  
Accepted for press 20 March 2020