О. Н. Камшилова

ИССЛЕДОВАТЕЛЬСКИЙ ПОТЕНЦИАЛ КОРПУСА АНГЛИЙСКИХ ТЕКСТОВ ПЕТЕРБУРГСКИХ ШКОЛЬНИКОВ: АНАЛИЗ ИНТЕРЯЗЫКА

«Корпус английских текстов петербургских школьников» — попытка создать электронную оперируемую текстовую базу данных для решения лингвистических задач. С помощью автоматических текстовых инструментов на базе созданного корпуса эффективно решается задача выявления и описания лингвистических характеристик интерязыка — грамматической системы языка обучающихся английскому языку как иностранному.

Ключевые слова: специальный корпус, формат корпуса, корпус текстов обучающихся английскому языку как иностранному (EFL), интерязык, базовые структуры, модели, «структурная бедность».

O. Kamshilova

SPB EFL CORPUS POTENTIAL: A STUDY OF INTERLANGUAGE IN A RUSSIAN EFL CLASSROOM

SPb LC is an attempt to compile a target-specific structure, a text collection in accord with essential corpus design criteria. Operated with reliable free tools, the corpus proved efficient enough in spotting and analyzing the learner language. The SPb LC findings are preliminary observations of learner interlanguage grammar.

Keywords: special corpus, corpus format, EFL learner corpus, interlanguage, basic structures, patterns, “structural poverty”.

Благодаря известным достижениям исследовательской группы, возглавляемой С. Гранже (Université catholique de Louvain, Бельгия), с конца 90-х в арсенал лингвистических баз данных прочно вошли так называемые корпуса текстов обучающихся (Learner Corpora, далее — LC). Это обширные электронные коллекции текстов, как правило, эссе, авторами которых являются молодые люди, изучающие английский язык как иностранный (EFL). Выбор возрастных параметров и языка для LC объясняется вполне понятными причинами — востребованностью английского языка в современной академической среде и необходимостью совершенствовать процесс обучения в соответствии с требованиями времени. Подобные корпуса текстов (*International Corpus of Learner English (Louvain, Бельгия)*, *Hong Kong Academic English (Гонконг)*, *Michigan Corpus of Academic Spoken English (Чикаго, США)*, *English as Lingua Franca in Academic Settings (Тампере, Финляндия)* и др.) представляют собой ценнейший ресурс для выявления и анализа широко распространенных и регулярно повторяющихся ошибок в текстах, порождаемых носителями самых разных родных языков. Масштабы LC, а также современные компьютерные средства обработки текстовых данных, которыми они оснащены, предоставили ученым обширный эмпирический материал, повлиявший на сам подход к разработке методик и к созданию учебной литературы. По свидетельству С. Гранже и М. Ранделла, в учебной лексикографии использование корпусных данных революционно пре-

образовало не только процесс создания словарей, но и сам продукт: учебные словари существенно изменились [6].

В европейской традиции, во многом благодаря успехам бельгийской группы, анализ речевых ошибок в текстах LC и извлечение полезной информации для совершенствования методик обучения английскому как иностранному преимущественно ведется в сфере лексики, грамматические структуры при этом рассматриваются лишь как необходимый контекст для словоупотребления. Иной подход к использованию данных LC сложился в Северной Америке (США, Канада) в рамках теории овладения вторым языком (second language acquisition, или SLA). Последователей этого направления интересует, в частности, процесс усвоения грамматики второго языка через этап, получивший в свое время название *интерязыка (interlanguage)* [9]. Интерязык — это переходная грамматика (transfer grammar), к которой прибегает учащийся, осваивающий второй язык (в нашем случае — иностранный) по пути к полному овладению им (native like performance). Интерязык — это те структуры, которые учащийся предпочтительно выбирает. Они представляют собой системное и осознанное вмешательство в нормативную грамматику иностранного языка (ср.: «systematicity and opportunity for intelligent intervention rather than random error» [2, с. 393]), поскольку их логика обеспечивает учащемуся определенную степень уверенности в общении; при этом, с формальной точки зрения, они не являются ошибками, однако с точки зрения носителя языка, они неуместны, неуклюжи и иногда непонятны.

Грамматические структуры интерязыка — результат интерференции родного языка и собственных стратегий, обеспечивающих на определенной стадии коммуникацию на осваиваемом языке. В дальнейшем, при успешном овладении вторым языком, эти структуры уступают место естественным для английского языка моделям. Исследователей процессов овладения вторым языком интересуют вопросы, связанные с моментом появления и временем использования этой «переходной грамматики», равно как и с механизмами ее порождения. Для этих целей наблюдения в специальных LC оказываются многообещающими [2].

Подобные структуры можно наблюдать и в речевых продуктах изучающих английский язык как иностранный (EFL). Овладение иностранным языком не обязательно предполагает как конечный результат совершенное владение, поэтому «переходная грамматика» в дальнейшем формирует базовый структурный фонд (*ср*. понятие *fossilization* (окаменение) в SLA), включающий максимально простые и максимально изоморфные структурам родного языка модели. Именно такие структуры чаще используются как в спонтанной, так и в подготовленной речи на иностранном языке, приводя к известной «структурной бедности» глобального английского языка (языка межнационального общения) даже при достаточно высоком сертифицированном уровне владения [7; 8]. Достаточно объективные ответы на вопросы о том, какие именно грамматические модели востребованы говорящим на иностранном языке, чем обусловлено их употребление, какова их частота и степень сложности и как они соотносятся с базовыми грамматическими структурами английского языка, можно получить с помощью специального корпуса текстов.

Специальный корпус текстов — это сбалансированный корпус, как правило, небольшой по размеру (несколько тысяч словоупотреблений), подчиненный определенной исследовательской задаче и предназначенный для использования только в целях, соответствующих замыс-

лу составителя. Использование находящегося в свободном доступе достаточно большого количества инструментов обработки текста (корпусных менеджеров) превращает коллекции текстов в электронные продукты, способные накапливать и обрабатывать лингвистическую информацию согласно задачам исследователя. Подобный корпус, основ-ной целью которого было исследование особенностей английских текстов, порождаемых русскими школьниками, был создан на кафедре прикладной лингвистики РГПУ им. А. И. Герцена в рамках магистерской программы*.

Описанные выше LC также представляют собой специальные корпуса, созданные в соответствии с поставленными исследовательскими задачами. Благодаря сотрудничеству, в том числе международному, многие из этих корпусов содержат весьма представительный материал, например, величина бельгийского ICLE — 3 млн словоупотреблений [6, с. 16]. Только он один из упомянутых ранее LC содержит английские тексты носителей русского языка, хотя количество русских студентов, проходящих обучение в Европе и Америке на английском языке, достаточно велико, а в русских школах английский является безусловным лидером среди преподаваемых иностранных языков. Поэтому создание корпуса английских текстов русских студентов и школьников представляется своевременным и полезным. Причем не только в аспекте оптимизации обучения, но и для проведения разного рода исследований, которые могут решать на основе представительного аутентичного материала более общие лингвистические вопросы, например, о природе интерференции, об универсальном и индивидуальном в языках разных типов, о коммуникативном «минимуме» и др. Наконец, реализация программы по созданию специального корпуса подтверждает эффективность применения современных компьютерных технологий, позволяющих формулировать новые лингвистические задачи и определять новые объекты лингвистического анализа.

Создание «Корпуса английских текстов петербургских школьников» (далее — «Корпус») осуществлялось в соответствии с общими принципами корпусного строительства и с учетом критериев, специально разработанных для LC [5, с. 8]. Для обработки и извлечения лингвистической информации использовались доступные и надежные текстовые инструменты**.

Аутентичный текстовый материал был собран в школах Санкт-Петербурга в период с ноября по декабрь 2007 года. Авторами текстов являются 78 учеников 9–11-х классов, предварительно прошедших тестирование. Уровень владения английским языком был определен как средний / intermediate (26%) и выше среднего / upper-intermediate (74%).

Размер «Корпуса» на данный момент составляет 38 122 словоупотребления.

Регламентация текстовых параметров включала тип коммуникации (устная и письменная); жанровую специфику; объединяющую тему, поскольку она определяет выбор лексики, и степень подготовленности речевого продукта, фиксированного в тексте (подготовленная / неподготовленная речь).

«Корпус» включает два подкорпуса письменных текстов (эссе и личное письмо) и два подкорпуса скриптов устных текстов (монолог и диалог). Жанры и темы выбирались в соответствии со школьной программой и с форматом ЕГЭ по английскому языку. Формулировки заданий заимствованы из подготовительных материалов к ЕГЭ по английскому языку [10; 13] Критерий степени подготовленности текстов, таким образом, принципиально отличается от того, что был предложен С. Гранже: поскольку

информанты были достаточно знакомы с темами, время на выполнение задания было ограничено. При этом использование справочных материалов (словарей, учебников и т. п.) не допускалось.

Информация, касающаяся авторов текстов, собиралась в анкеты и позднее была положена в основу паспорта текста. Стандартизованный паспорт имеет следующий вид:

Text Type: Essay
ID: Black dog
Age: 16
Sex: Male
School: 631
Form: 9
Mother-tongue: Russian
L2: German
Additional education: Access course
Language Level: Intermediate
Topic: Many people keep pets nowadays. However, other people think that keeping a pet is troublesome
Wordcount: 276 words
Date: 09.12.2007

Тексты были размечены с помощью доступных корпусных менеджеров и обработаны с помощью парсеров и программ, составляющих частотные словари, конкордансы, списки коллокаций (см. примечания). В результате каждый текст представлен в корпусе в трех модификациях — в оригинальной, с синтаксической разметкой (parsed) и с разметкой ошибок (error-tagged).

Выбор инструментов для корпусной разметки осуществлялся с учетом материала и исследовательской задачи [11]. Аутентичные тексты обучающихся содержат большое количество нестандартных форм из-за множества орфографических и грамматических ошибок. По этой причине следует с большой осторожностью выбирать программное обеспечение, рассчитанное на работу со стандартизованными текстами [4, с. 15]. Относительно малый размер «Корпуса английских текстов петербургских школьников» в этом случае оказался скорее достоинством, чем недостатком, поскольку позволял проводить ручную коррекцию. «Корпус» оснащен программой, позволяющей пополнять текстовую базу.

Важным отличием «Корпуса» является средний (intermediate) уровень владения английским языком его информантов. Следует отметить, что LC, включающие тексты уровня intermediate, как в канадском корпусе [2], очень редки. Авторы текстов большинства известных LC имеют подтвержденный продвинутый (advanced) уровень владения иностранным языком, кроме того, их средний возраст — выше: это студенты вузов и колледжей. Сложившиеся индивидуальные стили обучения, осознанные цели (мотивация) обучения в совокупности с программами курсов Advanced, уделяющими особое внимание работе с письменными текстами, обращение к справочным материалам делают уровень порождаемого текста приближенным к уровню текста носителя языка. Поэтому о полуподготовленных (темы письменных текстов и устных высказываний были хорошо знакомы информантам из прошлой практики) речевых произведениях информантов петербургского «Корпуса» есть больше оснований говорить как о текстах, фиксирующих грамматику интер-языка. Поскольку петербургские школьники были ограничены во времени и процесс порождения речи был максимально приближен к спонтанному, зафиксированные в тексте вокабуляр и грамматические модели предположительно должны отражать тот языковой запас, к которому обучающиеся реально прибегают при общении на иностранном языке.

Рассмотрим первые результаты работы с материалами «Корпуса». В качестве

основной ***гипотезы*** исследования на начальном этапе было принято предположение о том, что наиболее характерные и частотные грамматические структуры и модели, образующие грамматическую базу глобального английского языка, закрепляются уже на уровне intermediate. Эти модели определенно ориентированы на родной язык обучающихся и только частично соотносятся с базовой грамматикой литературного (standard) английского языка.

Одним из основных приемов работы с корпусными данными является *анализ частотных списков слов*. Предварительный сравнительный анализ частотных списков трех рабочих LC — French learner writing corpus [4, с. 17], Quebec advanced LC [2] и «Корпуса английских текстов петербургских школьников» — позволяет сделать несколько интересных наблюдений. В таблице приведены первые 10 слов частотных списков по анализируемым корпусам.

Отметим, что французский и канадский корпуса представлены только письменными текстами (эссе), а уровень владения английским языком их авторов выше (advanced). Тем не менее, сравнение первой десятки наиболее частотных слов достаточно информативно.

Первое место личного местоимения *I* в частотных списках петербургского «Корпуса» объясняется преимущественным выбором ориентированных на говорящего тем устных высказываний (*Give a two-minute talk on yourself, Give a two-minute talk on friendship, You and your friend think about a leisure time activity*) и одного из письменных форматов (письмо другу). Однако состав первой десятки в подкорпусе эссе приближается по составу к спискам французского и канадского корпусов: ранги «to» и «the» по трем корпусам представляются соотносимыми, что определенно является следствием специфического для данного жанра текста форматом.

Особый интерес вызывает высокая частота союза *and*, глагольной формы *is* во всех сопоставляемых корпусах и специфически высокий для петербургского «Корпуса» ранг личных глагольных форм *like* и *have*. Данные словаря-конкорданса свидетельствуют, что союз *and* используется для связи многочисленных коротких клауз, однородных членов, а также в начале предложения и в заполнении пауз при обдумывании реплики, что весьма характерно для спонтанной, полуподготовленной устной речи. Появление в десятке самых частотных слов глаголов *like* и *have*, которые за исключением редких случаев используются как знаменательные (например, в подкорпусе эссе обнаружены 109 случаев употребления *have*, в число которых входят только четыре вспомогательных и 14 модальных значений), предполагает, что петербургские школь-

Слова	**French LC**	**Quebec LC**	**SPb LC all texts**	**SPb LC Essay**	**SPb LC Letter**	**SPb LC Monologue**	**SPb LC Dialogue**
1	the	the	I	to	I	I	I
2	of	to	to	the	to	and	you
3	to	I	and	and	you	to	to
4	a	a	you	I	and	my	the
5	and	of	the	is	in	a	like
6	is	and	a	a	the	have	and
7	in	in	is	it	my	is	we
8	that	that	it	you	a	like	go
9	it	is	in	people	very	in	it
10	be	it	like	have	is	very	think

ники довольно активно прибегают к использованию модели I / WE <HAVE> Y и I / WE <LIKE> C.

Высокая частота *think* обусловлена выбором стандартного оборота эпистемической оценки *I think* предпочтительно в начале высказывания, что отмечается как одна из типичных черт речи обучающихся [6, с. 17]. Единственный глагол, попавший в первую десятку частотного списка сопоставляемых корпусов, — это глагол *be* / *is*. Без доступа к текстовой базе данных этих корпусов невозможно установить их функциональную принадлежность (знаменательный, связочный, служебный или модальный глагол), однако высокий ранг инфинитивной формы, скорее всего, предполагает ее использование в составе аналитических форм. Это дает основание прогнозировать большее разнообразие глагольных моделей и, как следствие, грамматических структур в текстах обучающихся продвинутого (Advanced) уровня.

В петербургском «Корпусе» глагольная форма *is* употреблена 413 раз. Автоматический анализ словосочетаний (collocation list) отмечает 22 употребления в функции знаменательного глагола в бытийном значении (*there is, is in*), четыре модальных употребления (*is to*) и 11 случаев в функции вспомогательного глагола (три — в пассивной глагольной форме — *is called / fried* и восемь — в длительной — *is working / keeping / making etc*.). Таким образом, подавляющее число употреблений этой глагольной формы (376) приходится на функцию связки в модели X IS Y. Именно эта модель считается характерной для той формы языка, которую принято называть сегодня глобальным английским.

Предпочтения обучающихся в выборе грамматических структур и моделей, спрогнозированные с помощью частотного списка, списка коллокаций и словаря-конкорданса, были проанализированы на основе синтаксического представления текстов «Корпуса», осуществленного с помощью автоматического парсера Connexor Natural Knowledge (см. примечания). Этот инструмент оказался достаточно точным при анализе текста, не подвергавшегося нормализации. Кроме того, графическое синтаксическое представление синтаксических моделей в этой программе достаточно удобно для дальнейшей обработки данных. Согласно выдвинутой гипотезе, корпусные данные дают возможность акцентировать внимание на самых распространенных грамматических структурах, которые систематически и намеренно выбирают обучающиеся и которые составляют ядро интерязыка. Закономерно встает вопрос о том, насколько эти структуры соответствуют структурам базовой английской грамматики, а если такое соответствие обнаруживается, чем именно определяется «структурная бедность» и «простота без элегантности» [7] текстов обучающихся.

При всей конвенциональности термина «базовая грамматика» (Basic Grammar) существует не одна попытка формального описания структур, которые обычно соотносят с данным термином (от известных моделей дескриптивной и структурной грамматики до учебных моделей в практике преподавания английского языка как иностранного). Список грамматических структур и их представление, принятые в настоящей работе, заимствован из Longman Grammar of Spoken and written English [1] — одного из самых авторитетных современных изданий:

- S V A (Mary is in the house);
- S V C (Mary is kind / a nurse);
- S V Od (Somebody caught the ball);
- S V Od A (I put the plate on the table);
- S V Oi Od (She gives me presents);
- S V (The child laughed).

Предварительное исследование в синтаксически размеченном корпусе убедительно свидетельствует об использовании обучающимися этих базовых грамматических структур при порождении текстов. Каждая из перечисленных структур была обнаружена в текстах «Корпуса». Однако «иностранный акцент» в этих текстах создается тем, какие из данных структур наиболее востребованы и какое именно наполнение получают элементы этих структур.

Общеизвестным фактом является предпочтительное использование иностранцами, говорящими / пишущими на английском языке, структур SVC и SVA, преимущественно со связкой *be* [7; 8]. Данные петербургского «Корпуса» демонстрируют ту же тенденцию: относительная частота предикатов с *be*-связкой в «Корпусе» — 0,33.

Репертуар глагольного компонента (V) в S V Od, S V Od A, S V Oi Od и SV структурах указывает на другую черту актуальной грамматики обучающихся, а именно намеренное избегание перфектных и длительных глагольных форм. Замена видовых форм простыми несет в себе потерю информации для носителя языка, для которого нюансирование аспекта протекания события / ситуации является важной составляющей сообщения, но для коммуникации на уровне «переходной грамматики» это не является осложняющим обстоятельством. Эта тенденция сохраняется и в «глобальных английских» текстах. Из всех модальных операторов глагольного компонента авторы текстов «Корпуса» отдают предпочтение *can* (абсолютная частота в корпусе — 235, ранг в частотном словаре — 20). Таким образом, указанные базовые структуры английской грамматики реализуются в текстах «Корпуса» преимущественно в виде моделей S Vsimp Od, S Vsimp Od A, S Vsimp Oi Od, SVsimp и S can V Od, S can V Od A, S can V Oi Od, S can V.

Анализ частотного списка указал на преимущественное употребление глаголов *have* в подкорпусе эссе и *like* в подкорпусе монологических текстов. Есть основания полагать, что эти предпочтения также являются следствием преимущественного использования простых структур. Обращение к синтаксически размеченному корпусу позволило обнаружить 362 употребления *have* по всему «Корпусу» и 131 в подкорпусе эссе. В подавляющем большинстве случаев *have* был отмечен как знаменательный глагол: 323 («Корпус») и 114 (подкорпус эссе) контекстов, модальное значение обнаружено в 35 и 17 контекстах соответственно. В функции вспомогательного глагола *have* был использован очень редко (четыре контекста) и некорректно:

[*I “have been falling in love” with “Counter-strike” for 6 years*]***

Абсолютное предпочтение знаменательному *have* может быть следствием частого использования модели S <HAVE> Od. Данные словаря коллокаций подтверждают это предположение и определяют в качестве самых предпочтительных реализаций этой модели контексты типа

[*I also have a cat and fish*].

Анализ корпусных данных позволил обнаружить специфическую для русских школьников особенность в заполнении позиции Od отмеченной модели: авторы текстов петербургского корпуса предпочитают использовать базовую английскую структуру S <HAVE> Od вместо другой базовой структуры S <BE> C, заполняя позицию дополнения номинализованными формами (существительными) адъективного предикатива C или глагольного сказуемого в структуре SV (prepO): S <HAVE> *allergy / freedom / independence / communication* вместо естественных для английского языка S <BE> *allergic / free / independent* и S <COMMUNICATE> with O. Отмеченное

расширение использования (overuse) структуры S <HAVE> Od с очевидностью можно отнести к фактам «переходной грамматики», то есть к интерязыку:

[*I can understand when people don't keep any animals because they have an allergy*]

[*If you leave your childhood house you'll have your own life*]

[*If you have your own accomodation you also have a freedom*]

[*I enjoy to have communication with interesting people from different countryes*].

Другим высокочастотным словом в корпусе является *like* (359 употреблений). Поскольку больше трети его употреблений приходится на подкорпус монологических текстов (139), где оно входит в десятку самых частотных слов, обратимся к анализу данных именно этого подкорпуса.

Автоматический парсер Connexor выявил 117 контекстов в этом подкорпусе, где глагол *like* используется в функции простого глагольного сказуемого @+FMAINV. Модели типа *I like studying* отмечаются этим парсером как комплементированнные структуры @CS %CS (два употребления). Еще 20 случаев употребления были распознаны парсером как предлоги. Парсер ошибся в 10 случаях из 117 вследствие нестандартного заполнения (ошибки) комплементированной структуры: [*I very like my family; I very like studying English; I very like to swim there*].

Отмеченные 117 контекстов с *like* в роли простого глагольного сказуемого соответствуют модели S <LIKE> C (имеющей лишь поверхностное сходство с базовой структурой SVOd). При этом позиция C заполняется исключительно инфинитивом, что говорит об игнорировании русскими школьниками нюансов употребления инфинитива и глагольной формы *-ing*. Модель с инфинитивом регулярно избирается авторами текстов «Корпуса», что можно объяснить ее внешним сходством с базовой структурой SVOd и параллельной русской моделью *Я люблю читать*. В некоторых монологах эта модель повторяется много раз, налицо — сознательный стратегический выбор и, как следствие, «структурная бедность» текста:

[*I'm 15 years old. I was born in Krasnoyarsk and we moved to St. Petersburg when I was 5 years. I live with my mother, father, sister and brother. My mother is a doctor, father is an architect, my brother is 25 and he is a musician, he has a group. It's very cool to have a big brother my sister is 10 and she goes to school. We have a cat and a dog. I have a lot of friends, my best friend is in my class, and we are emm we like to do something together in our free time. My friends are funny; we have fun together, make parties, and do some crazy things. We meet with our company every weekend.* ***I like skating*** *and I have a lot of friends skaters. In winter I want to learn... to learn snowboard, but there is no snow.* ***I like to draw*** *and I want to learn to draw graffiti. I'm emo,* ***I like this style*** *and a lot of my friends are emo too.* ***I like piercing****, I have one in my tongue, I want to do one in my... right here, under my lip, but* ***my mother doesn't like it****.* ***I like rock music*** *and my favorite group is My chemical Romance.* ***I don't like pop music****. Sometimes I write my own poems and songs and stories. We often travel with my family; we sometimes go to Finland by car.* ***I like traveling*** *and I dream to visit Germany —* ***I like their language and culture****.* ***I also like cartoons and different films and cinema****.* ***I like coffee and candies.***]

Подводя предварительные итоги анализа интерязыка на основе данных созданного исследовательского корпуса текстов EFL, можно сделать следующие в ы в о д ы .

Продуцирование полуподготовленной письменной и устной речи на английском

языке (EFL) в режиме заданного времени демонстрирует ограниченный репертуар достаточно простых конструкций и моделей, к которым чаще всего прибегают обучающиеся. Выбор этих конструкций обусловливается внешним сходством с самыми простыми базовыми структурами английского языка и параллельными структурами родного языка. Систематическое предпочтение этих структур развернутым и более естественным моделям стандартного английского языка, которые отрабатываются в процессе обучения и знание которых подтверждено предварительным тестированием, приводит к так называемой «структурной бедности» речевых произведений *не*носителей языка. В репертуаре грамматических структур, обнаруженных в «Корпусе английских текстов петербургских школьников», есть такие, которые представляют собой случаи «переходной грамматики» (интерязыка) и составляют, наряду с максимально простыми базовыми структурами, ядро грамматики EFL, которое не совпадает с базовыми грамматическими структурами литературного английского языка. Поскольку подобные модели отмечаются в «глобальных английских» текстах в разных сферах международного общения, можно с большой долей уверенности говорить о том, что складывающиеся нормы «глобального английского» во многом опираются на «окаменевшие» модели интерязыка.

Эффективным методом выявления и анализа подобных языковых объектов являются специальные исследовательские корпуса текстов. Оборудованные автоматическими инструментами обработки и поиска лингвистической информации, корпуса текстов предоставляют обширный эмпирический материал и изменяют сам характер описания языкового объекта. Многоаспектный подход к анализу языкового объекта, опору на представительный материал можно считать несомненным преимуществом корпусного метода. Анализ интерязыка далеко не единственная лингвистическая задача, которую можно решать с помощью созданного «Корпуса». Перспективными представляются исследования словаря корпуса, специфики текстовых категорий. Методический потенциал размеченного подкорпуса ошибок также очевиден. «Корпус английских текстов петербургских школьников» — открытый проект, текстовая база которого может пополняться.

ПРИМЕЧАНИЯ

* Магистерские исследования Е. А. Николаевой «Принципы и методы формирования и поддержки корпуса текстов обучающихся (на примере корпуса английских текстов петербургских школьников)» и М. В. Колиной «Исследование базовых английских грамматических структур (Basic Grammar) в корпусе текстов обучающихся (Learner Corpus)», 2009 г.

** При работе с текстовым материалом использовались следующие программные инструменты: AntConc [http://www.antlab.sci.waseda.ac.jp], AntConc 3.2.1 [http://www.antlab.sci.waseda.ac.jp/software.html], Connexor Natural Knowledge (ENGCG) [http://www.connexor.eu/technology/ machinese/demo/syntax], The Compleat Lexical Tutor for Data Driven Language Learning on the Web [http://www.lextutor.ca/], Find and replace [http://www.abacre.com/afr.zip].

*** Здесь и далее в квадратных скобках представлен оригинальный текстовый материал корпуса.

СПИСОК ЛИТЕРАТУРЫ

1. *Biber D., Johansson S., Leech G., Conrad S., Finegan E.* Longman Grammar of Spoken and Written English. Harlow: Longman, 1999.

2. *Cobb T.* Analyzing late interlanguage with learner corpora: Quebec replications of three European studies // Canadian Modern Language Review. 59(3). February, 2003. P. 393–423.

3. *Gaskell D., Cobb T.* Can learners use concordance feedback for writing errors? Submitted to System, November, 2003, Revision April 12, 2004: [Электронный ресурс]. Режим доступа: http://www.lextutor.ca/concordancers/concord_e.html.

4. *Granger S.* Computer learner corpora, second language acquisition and foreign language teaching. Amsterdam: J. Benjamins Pub. Comp., 2002.

5. *Granger S.* The computer learner Corpus: a versatile new source of data for SLA research // Learner English on Computer. Granger S. (ed). London: Longman, 1998. P. 3–18.

6. *Granger S., Rundell M.* From corpora to confidence // English Teaching Professional, Issue 50, May 2007. P. 15–18: [Электронный ресурс]. Режим доступа: www.etprofessional.com.

7. *Hinkel E.* Simplicity without elegance: Features of sentences in L1 and L2 Academic texts // TESOL Quarterly. Vol. 37. 2003. № 2. P. 275–301.

8. *Mauranen A.* The Corpus of English as *Lingua franca* in Academic Settings // TESOL Quarterly. Vol. 37. 2003. № 3. P. 515–527.

9. *Selinker L.* Interlanguage // International Review of Applied Linguistics 10. 1972. P. 209–31.

10. *Вербицкая М. В., Махмурян К. С.* ЕГЭ 2009. Английский язык. Тренировочные задания. М.: Эксмо, 2008.

11. *Камшилова О. Н., Колина М. В., Николаева Е. А.* Разработка корпуса текстов петербургских школьников: задачи и перспективы // Прикладная лингвистика в науке и образовании. СПб., 2008. С. 92–98.

12. *Разумова В. В.* Выбор синтаксического анализатора для анализа сложных предложений в представительном корпусе текстов // Прикладная лингвистика в науке и образовании. СПб., 2008. С. 150–154.

13. *Северинова Е. Ю.* ЕГЭ. Английский язык: раздаточный материал тренировочных тестов. СПб.: Тригон, 2008.

REFERENCES

1. *Biber D., Johansson S., Leech G., Conrad S., Finegan E.* Longman Grammar of Spoken and Written English. Harlow: Longman, 1999.

2. *Cobb T.* Analyzing late interlanguage with learner corpora: Quebec replications of three European studies // Canadian Modern Language Review. 59(3). February, 2003. P. 393–423.

3. *Gaskell D., Cobb T.* Can learners use concordance feedback for writing errors? Submitted to System, November, 2003, Revision April 12, 2004: [Elektronnyj resurs]. Rezhim dostupa: http://www.lextutor.ca/concordancers/concord_e.html.

4. *Granger S.* Computer learner corpora, second language acquisition and foreign language teaching. Amsterdam: J. Benjamins Pub. Comp., 2002.

5. *Granger S.* The computer learner Corpus: a versatile new source of data for SLA research // Learner English on Computer. Granger S. (ed). London: Longman, 1998. P. 3–18.

6. *Granger S., Rundell M.* From corpora to confidence // English Teaching Professional, Issue 50, May 2007. P. 15–18: [Elektronnyj resurs]. Rezhim dostupa: www.etprofessional.com.

7. *Hinkel E.* Simplicity without elegance: Features of sentences in L1 and L2 Academic texts // TESOL Quarterly. Vol. 37. 2003. № 2. P. 275–301.

8. *Mauranen A.* The Corpus of English as Lingua franca in Academic Settings // TESOL Quarterly. Vol. 37. 2003. № 3. P. 515–527.

9. *Selinker L.* Interlanguage // International Review of Applied Linguistics 10. 1972. P. 209–31.

10. *Verbickaja M. V., Mahmurjan K. S.* EGJE 2009. Anglijskij jazyk. Trenirovochnye zadanija. M.: Eksmo, 2008.

11. *Kamshilova O. N., Kolina M. V., Nikolaeva E. A.* Razrabotka korpusa tekstov peterburgskih shkol'nikov: zadachi i perspektivy // Prikladnaja lingvistika v nauke i obrazovanii. SPb., 2008. S. 92–98.

12. *Razumova V. V.* Vybor sintaksicheskogo analizatora dlja analiza slozhnyh predlozhenij v predstavitel'nom korpuse tekstov // Prikladnaja lingvistika v nauke i obrazovanii. SPb., 2008. S. 150–154.

13. *Severinova E. Ju.* EGE. Anglijskij jazyk: razdatochnyj material trenirovochnyh testov. SPb.: Trigon, 2008.