

УДК 81'32

А.Н. Хромаленкова, В.Ю. Панкова  
Российский университет дружбы народов

**ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ  
ПРЕПОДАВАНИЯ ИНОСТРАННОГО ЯЗЫКА:  
ИНТЕГРАЦИЯ УЧЕБНЫХ КОРПУСОВ В ПРОЦЕСС  
ПРЕПОДАВАНИЯ (на примере REALEC (Russian Error-  
Annotated Learner English Corpus))**

*В статье описываются потенциальные способы интеграции учебных корпусов в преподавание иностранных языков. Автор рассматривает функциональные отличия корпусов ошибок от любых других текстовых корпусов и анализирует их как инструмент, способный помочь не только исследователю, но и преподавателю. На примере корпуса REALEC способы интеграции рассматриваются в условиях «родной язык учащихся – русский, иностранный [изучаемый] язык – английский».*

*Ключевые слова: корпус, компьютерные инструменты, REALEC, преподавание иностранного языка.*

UDC 81'32

A.N. Khromalenkova, V.Yu. Pankova  
Peoples' Friendship University of Russia

**UPGRADING EFFICIENCY OF TEACHING A  
FOREIGN LANGUAGE BY INTEGRATING THE USE OF  
LEARNER CORPORA INTO THE PROCESS OF  
TEACHING (on the material of REALEC (Russian Error-  
Annotated Learner English Corpus))**

*The article deals with potential ways of integrating learner corpora into the process of teaching foreign languages. The*

*author describes functional features distinguishing error-tagged corpora from other ones and analyzes the former as the helping means for both researchers and teachers. Basing on the example of REALEC the ways of integration are shown in case students' native language is Russian and the one taught to them is English.*

*Key words: corpus, computer tools, REALEC, teaching a foreign language.*

### **Введение**

Компьютерные корпусы текстов были созданы в поддержку исследователям-лингвистам еще в 60-е годы XX века. Они представляют собой электронные базы данных, состоящие из текстов, объединенных по некоторому (некоторым) признаку (признакам). Целью создания таких корпусов является получение статистических и/или иллюстративных языковых данных, которые отличались бы представительностью и достоверностью. Именно поэтому современные корпусы, как правило, насчитывают десятки и сотни тысяч текстов и бывают оснащены многообразием компьютерных инструментов, позволяющих производить все необходимые подсчеты, сортировку и анализ автоматически, то есть за исследователя. Именно эти критерии – представительность и компьютеризированность – являются сегодня отличительными чертами надежного корпуса. Но к ним мы еще вернемся.

Для начала кратко, но наглядно представим себе, что такое компьютерный корпус текстов. Для этого назовем другие две его немаловажные особенности. Первая касается способов поиска в нем и сортировки необходимых исследователю лингвистических единиц. Они, как правило, максимально многообразны, а потому удобны (например, искать можно обычно не только конкретные слова или словосочетания, но также заданные при помощи так называемых регулярных выражений языковые элементы. То

есть можно задать, к примеру, некоторые грамматические, семантические или иные параметры искомой словоформы или словоформ – а саму лемму не обязательно задавать вовсе, если необходимо выяснить по корпусу только встречаемость тех или иных параметров). Из этой особенности корпусов логически вытекает вторая – компьютерные корпусы, как правило, бывают размечены. Именно разметка делает возможным большинство дальнейшей обработки: морфологическая, синтаксическая и/или любая другая разметка присваивает (иногда всем) токенам (то есть словам) текста особые ярлыки – так называемые тэги, содержащие информацию о токенах, которую способны распознавать поисковые системы.

Отличие же учебных корпусов, интересующих нас, состоит в том, что именно и только в них используется так называемая разметка ошибок, то есть, по сути, разметка их типов. Поэтому такие корпусы также называют корпусами ошибок. «Учебными» же они называются потому, что составляют их из текстов всегда на определенном языке, не являющимся для их авторов родным. Чаще всего при этом (для удобства и однозначности) корпусы организованы таким образом, что включают в себя только тексты авторства носителей одного и того же родного языка, изучающих данный конкретный (также общий) иностранный язык.

Для чего придумали корпусы ошибок? Как используется в процессе лингвистических исследований их специфика, и как именно это может быть применимо к процессу преподавания? Прежде чем приступить к ответу на последний вопрос, рассмотрим, что какие труды других исследователей могут иметь отношение к нашей проблеме.

#### **Материал исследования**

Одной из наиболее основательных работ в данной области можно считать труд бельгийской

исследовательницы Сильвиан Грейнджер «Error-tagged learner corpora and CALL: a promising synergy» («Учебные корпуса с разметкой ошибок и CALL: успешная совместная работа в перспективе» (Прим.: перевод наш – А.Х.)). В ней исследователь дает учебному корпусу определение как электронной коллекции аутентичной информации по иностранному языку и выделяет уже признанные нами важнейшие его черты: компьютеризованность и представительность. Именно первое, согласно С. Грейнджер, позволяет анализировать информацию об иностранном языке с помощью целого ряда автоматических инструментов, что обеспечивает быстроту и продуктивность данному процессу. Здесь используются поиск, подсчет, самые разные виды сортировки, а также множество программы, снабжающих текстовые данные необходимой лингвистической информацией (например, о синтаксической структуре, грамматических категориях и т.п.). И именно второе [представительность] позволяет учебным корпусам, подобно прочим, быть в своей области надежным источником данных и надежной базой для проводимых исследований и всех находок, заключений и выводов относительно процессов в освоении иностранного языка.

Следует упомянуть также работу Д.П. Дутры и Т.Б. Сардины «Referential expressions in English learner argumentative writing» («Референциальные выражения в аргументированных эссе изучающих английский язык» (Прим.: перевод наш – А.Х.)). Ее авторы (и они, разумеется, не единственные) решили перейти от теории к практике: ими было проведено конкретное лингвистическое исследование при помощи компьютерного корпуса. Целью авторов было сравнить использование различных референциальных выражений англоязычными и неанглоязычными студентами. Для этого была извлечена

статистика из LOCNESS (Native English Essay corpus – корпуса текстов, написанных носителями английского языка) и двух учебных корпусов также английского языка. Все три корпуса в итоге помогли исследователям выяснить, какие именно подкатегории референциальных выражений используются изучающими английский язык чаще, а какие – реже (в отличие от носителей данного языка). Таким образом, учебные корпуса выступили в роли источника всей необходимой статистической информации, подкрепленной примерами, получение которой, к тому же, не потребовало много времени и усилий. Компьютерные корпуса также помогли авторам изначально конкретизировать тему своего исследования: по извлеченной статистике, именно референциальные выражения оказались наиболее частотными сочетаниями от 3 до 5 слов во всех трех корпусах английских текстов.

Как можно видеть, обе приведенные работы соотносятся с проблематикой данной статьи не абсолютно. Однако это объяснимо, обоснованно и понятно: вопрос, которым задается автор, скорее всего, не беспрецедентен, тем не менее, среди опубликованных на данный момент работ на него можно найти лишь косвенные ответы – анализируя контекст близких по тематике работ, например, таких, какие упомянуты выше. В целом же актуальность данного анализа заключается именно в том, что автор статьи, вероятно, первая предпринимает попытку дать четкий ответ на вопрос «как применимы учебные корпуса в процессе преподавания иностранного языка?», сделать некоторые предположения и заключения именно внутри этой, достаточно четко очерченной узкой области.

#### **Методы исследования**

В заключение данной части скажем несколько слов о методологии. Она заключается не только в изучении названных выше трудов, но и в непосредственной работе с

заявленным в начале корпусом REALEC, изучении его функционала как при помощи мануалов (инструкций), так и на практике (то есть путем его использования). В заключение укажем, что все сказанное выше было тщательно проанализировано в контексте потенциальной интеграции в преподавание.

### **Результаты и обсуждение**

Итак, для начала рассмотрим функционал корпуса англоязычных эссе русскоговорящих студентов REALEC чуть подробнее, а затем на основе данного описания опишем предполагаемые перспективы для интеграции его (и ему подобных) в процесс преподавания иностранных языков.

Именно REALEC (Russian Error-Annotated Learner English Corpus) был избран в качестве образцового корпуса ошибок для нашей статьи неслучайно: он является первым в своем роде корпусом, составленным на основе работ носителей русского языка, писавших свои тексты на английском. На сегодняшний день корпус необычайно представительен, включает в себя около 1200 работ студентов – в большинстве своем эссе, написанных во время подготовки к сдаче экзамена IELTS или на самом экзамене. REALEC часто используется преподавателями в процессе проверки студенческих работ: в НИУ ВШЭ, где этот корпус и был создан, существует активная практика занесения в него таких эссе с последующей разметкой в них ошибок. Таким образом, первый способ применения такого корпуса очевиден: он может оказаться незаменимым для упрощения обратной связи между учителем и учеником и сведения ее к удаленной (то есть при помощи интернета, при наличии у ученика доступа на сайт). Преподаватель способен предоставить четкую и подробную информацию о результатах проверки любой работы, а учащиеся способны обратиться к ней в любое удобное для них время и

просмотреть все в достаточно удобной форме.

Как организован REALEC? Каждый «ярлык» ошибки (то есть помета, которую присваивает проверяющий слову, словосочетанию или части предложения, которая(ое) содержит, на его взгляд, ошибку) состоит из 5 информационных пунктов. Таковыми являются:

- тип ошибки
- предполагаемая причина ошибки
- величина ошибки с точки зрения языка (насколько «грубой» можно признать данную ошибку)
- степень влияния этой ошибки на понимание текста
- правильный вариант, предлагаемый взамен ошибочного.

В иных корпусах представление информации об ошибке может совпадать с подходом REALEC, а может варьироваться; однако неизменным остается, как правило, наличие, как минимум, первого пункта – «тип ошибки» - и о нем мы скажем еще несколько слов, так как он для нас особенно важен.

Итак, как правило, типы ошибок бывают классифицированы (как и в REALEC) иерархически, то есть многоуровнево. Система ярлыков, которые можно «навесить» на слово или область предложения, которую сочли ошибочной, можно изобразить, таким образом, древообразно. Ближе к вершине, в таком случае, располагаются более общие категории ошибок; ниже – их деления на подкатегории; они, в свою очередь, также при надобности делятся на еще более мелкие подкатегории - и так до тех пор, пока каждая «веточка» «дерева» не будет оканчиваться максимально конкретизированным типом ошибки.

Именно отсюда – из способности корпуса, подобного REALEC, предоставлять информацию об ошибках

изучающих иностранный язык в таком унифицированном, четком, легко компоуемом виде – и следует основная (на наш взгляд) перспектива использования таких корпусов в преподавании иностранного языка. Итак, обратим внимание на следующий момент: именно работа с учебными корпусами сегодня делает доступной почти любую статистическую обработку собранных данных об ошибках, не говоря уже о мгновенном сборе этой статистики как таковой (для сравнения: при проведении таких процедур вручную и на ту, и на другую стадию обычно уходит время, исчисляемое часами и/или днями; иногда это делает исследования и вовсе невозможным или приводит к его сокращению в случае ограниченности сроков и/или нехватки легкодоступных ресурсов в большом количестве). Полученная же таким современным, автоматизированным способом информация сразу становится доступной для интеграции в любую сферу языкознания, где она может быть потенциально полезной. В нашем случае статистика, которая складывается из конкретных примеров, на которые также можно в любой момент взглянуть, оказывается полезна для преподавателей иностранного языка со следующей стороны.

Нам открывается абсолютно точная картина: где именно возникают трудности у учащихся с данным родным языком чаще (а где реже). Из этого следует, что учитель яснее (и точнее) способен представить, на каких темах следует сделать акцент при обучении данной конкретной группы или в похожих случаях (или же вообще - в принципе, в преподавательской практике как таковой). Однако главное, пожалуй, – ошибки какого рода допускают при изучении данного, конкретного иностранного языка ученики (то есть не только где они ошибаются, но и как они это делают). Данная статистика весьма богато проиллюстрирована примерами, что лишний раз помогает преподавателю.



Известно, что варианты неправильных употреблений и конструкций в осваиваемом языке у носителей разных языков отличаются. В данном случае есть возможность проследить, какие именно неправильные элементы учащиеся с конкретным родным языком – допустим, русским, в случае REALEC – привносят в производимый текст на данном иностранном (в данном случае, английском). Это может дать преподавателям возможность понять, в какую именно сторону необходимо было бы скорректировать преподавание для повышения его эффективности, как именно мыслят ученики при освоении той или иной темы, и с какой стороны можно было бы подойти к этой теме так, чтобы уменьшить количество ошибок подобного характера. В целом, на основе данной информации возможна разработка потенциально неограниченного количества новых методик (или их элементов).

Итак, статистика и иллюстративная база, доступные к извлечению из учебных корпусов, подобных REALEC, настолько разнообразны и представительны, что открывают перед исследователями вообще и в области преподавания в частности достаточный простор для творчества. Мы постарались перечислить лишь самые очевидные из способов их применения в данной сфере.

### **Выводы**

Конечно, в данном способе сбора данных, работе с ними и последующем применении можно усмотреть ряд минусов. Достаточно очевидным является, например, тот факт, что разметка ошибок, в отличие от некоторых других (например, частей речи), в силу своей специфики чаще всего производится экспертами вручную, а значит, является в некотором роде субъективной и недостаточно унифицированной. К примеру, одну и ту же ошибку разные разметчики способны пометить разными ярлыками

согласно своему собственному взгляду и оценке. Однако данный минус потенциально возможно свести к минимуму при помощи сопровождения каждого корпуса четкой аннотацией, какие именно ошибки каким именно способом следует рассматривать, и к какому типу в рамках каждой конкретной классификации относить. Также можно проследить, чтобы разметка производилась только профессионалами в области языка. Однако даже если данный недостаток не удастся исключить в ближайшее время полностью, ему противопоставлено достаточно солидное количество плюсов использования данных корпусов; а надежность полученных данных, хоть и не является полной и абсолютной, но, тем не менее, не уступает надежности, характеризующей сбор подобной информации вручную.

Следовательно, учебные корпуса, на наш взгляд, вполне заслуживают уделяемого им сегодня внимания и действительно способны внести свой вклад в развитие методики преподавания иностранных языков и многие другие сферы языкознания.

#### REFERENCES

1. Dutra, D.P., & Sardinha, T.B. (2013). Referential expressions in English learner argumentative writing. *Twenty years of learner corpus research: looking back, moving ahead*, 117-128.
2. Granger, S. (1994). The learner corpus: a revolution in applied linguistics. *English Today*, 39, Vol. 10, No. 3.
3. Granger, S. (2003). Error-tagged learner corpora and CALL: a promising synergy. *CALICO Journal*, 20 (3).