

ИРГИЗОВА К. В., САФОНКИНА О. С.

**КОРПУС АНГЛИЙСКОГО ЯЗЫКА КАК СРЕДСТВО АНАЛИЗА ОШИБОК
В ПИСЬМЕННОЙ РЕЧИ УЧАЩИХСЯ**

Аннотация. В данной статье рассматривается концепция теории анализа ошибок на основе корпусных ресурсов. Авторы высказывают предположение о том, что исследование природы ошибок учащихся на примерах из учебного корпуса, а также моделирование на их основе корректирующих упражнений, может способствовать уменьшению совершения ошибок в будущем. Предложен ряд практических рекомендаций по работе с корпусными ресурсами для преподавателей английского языка и учащихся.

Ключевые слова: корпусная лингводидактика, теория анализа ошибок, корпусный анализ ошибок, корпус учащегося, навыки академического письма.

IRGIZOVA K. V., SAFONKINA O. S.

**ENGLISH-LANGUAGE CORPORA
FOR ERROR ANALYSIS IN STUDENTS' WRITING**

Abstract. The article provides an overview of the theory of corpus-based error analysis. The authors suggest that the study of the nature of students' errors through examples from learner corpora and modeling relevant corrective exercises may reduce the percentage of subsequent errors. Practical tips for working with corpus resources for English teachers and students are offered.

Keywords: corpus linguistics, error analysis, corpus-based error analysis, learner corpora, academic writing skills.

Цифровая революция принесла в процесс образования ряд новых инструментов. Некоторые из них уже широко известны и активно используются, например, сервисы проверки орфографии письменных текстов. Другие, к примеру, лингвистические корпуса и программы для работы с ними (конкордансеры) используются менее широко. Корпусом называют набор лингвистических данных (письменных текстов или транскриптов устной речи), который может использоваться в качестве отправной точки для лингвистического описания или средства проверки гипотез о языке [9, с. 224]. С помощью корпуса исследователи могут собирать, наблюдать и анализировать лингвистические данные и выяснять сходства и различия между носителями языка и изучающими второй язык, или же между учащимися, имеющими разные уровни владения языком. Построение и функционирование корпусов изучается в рамках корпусной лингвистики, которая, в свою очередь, является результатом междисциплинарного развития информатики и лингвистики, новым образом мышления, методом более глубокого понимания природы языка.

Однако, как отмечают исследователи в этой области, несмотря на популярность корпусной лингвистики, ее применение в области обучения иностранным языкам остается недостаточно популярным. Так П. В. Сысоев утверждает, что для полноценного использования корпуса при работе и преподавателям, и учащимся необходимо владеть информационной компетенцией на достаточно высоком уровне. В свою очередь, А. Ю. Мордовин критикует исключительность использования корпусных технологий в лингводидактике из-за риска возникновения у учащихся «прагматического дискомфорта» при самостоятельной обработке информации во время работы с корпусом [6, с. 115]. В то же время, такие отечественные лингвисты как Н. Б. Гвишиани, В. П. Захаров, В. А. Плунгян, а также западные методисты Д. Синклер и М. Маккарти активно призывают к использованию корпусной лингводидактики в обучении ввиду его аутентичной природы и богатейшей эмпирической базы [2, с. 35].

В настоящий момент использование корпусных элементов в лингводидактике рассматривается с точки зрения выявления и статистического анализа ошибок учащихся путем использования корпусных технологий, а также с точки зрения использования содержимого корпуса в качестве базы примеров для работы над ошибками. Основоположником такого корпусного анализа текстов, позволившего по-новому взглянуть на процесс изучения ошибок в контексте лингводидактики, является С. Грейнджер, сотрудник центра по изучению англоязычных корпусов текстов при Лувенском католическом университете (Бельгия). Главное преимущество корпусной основы, по словам Грейнджер, заключалось в том, что она позволяла осуществлять эмпирический анализ примеров реальной коммуникации с использованием автоматических и интерактивных методов [13].

С точки зрения использования корпусных элементов в лингводидактике, особую пользу представляет влияние корпусных ресурсов на процесс коррекции ошибок учащихся при формировании навыков письменной речи. Причина, по которой для исследования был выбран именно этот вид речевой деятельности, заключается в том, что, во-первых, несмотря на обширную классификацию корпусов различных типов (в частности, устных и письменных) последние представлены наиболее полно. Корпус письменных текстов, являясь уникальной базой данных письменных образцов различных жанров, наиболее отвечает нашим исследовательским интересам. Во-вторых, выбор навыка письменной речи для исследования представлен актуальностью проблемы его формирования. Говоря об обучении языку, стоит сказать, что именно формирование навыка письма представляет наибольшую сложность, равно как для носителей языка, так и для тех, кто изучает иностранный язык. Анализ контента онлайн-платформ Elibrary, ResearchGate, Google Scholar показывает, что для

современных ученых проблема формирования навыков иноязычной письменной речи действительно актуальна, что в принципе неудивительно: письмо играет очень важную роль, поскольку является одновременно и средством, и целью обучения языку. Письменная речь неразрывно связана со всеми остальными видами речевой деятельности, стимулирует их, помогает в сохранении лингвистических знаний и развитии мышления. Помимо прочего, именно письменный материал, согласно исследованиям психологов, запоминается в среднем на 25% лучше, чем услышанный или увиденный [5, с. 108].

Однако все чаще наблюдается, что именно письменные навыки развиты у учащихся наиболее слабо. Об этом говорят, например, низкие результаты сдачи ЕГЭ за раздел «Письмо» в сравнении с другими разделами [1, с. 50]. В системе высшего образования также поднимается вопрос о повышении академической грамотности студентов посредством отработки навыков письма [7, с. 187]. В связи с этим многие образовательные учреждения начали вводить в учебные планы дисциплины, направленные на изучение прикладных и практических аспектов формирования компонента академического письма, однако проблема все еще остается актуальной [14, с. 5]. Даже имея богатый словарный запас и обширные знания о грамматике английского языка, учащиеся все равно допускают большое количество языковых ошибок ввиду слабо развитой коммуникативной компетенции, в данном случае – письменной.

Стоит, однако, сказать, что изучение иностранного языка, всегда характеризуется неизбежным появлением ошибок, поэтому, когда мы говорим об исследовании процесса обучения, необходимо сделать на фокус на анализе природы ошибок. Мы полагаем, что при условии выявления основных проблем при изучении языка и моделировании соответствующих корректирующих методов, студенты, несомненно, совершали бы меньше ошибок при письме. По этой причине необходимо обратиться к анализу основных типов языковых ошибок с применением элементов общей теории анализа ошибок (error analysis) и корпусного анализа ошибок (corpus-based error analysis).

Теория анализа ошибок является разделом прикладной лингвистики, направленным на изучение и анализ ошибок, допущенных учащимися в процессе изучения второго и иного иностранного языка. В рамках данной теории основной идеей является не только выяснение уровня владения иностранным языком, но и получение информации о трудностях, которые возникают у учащегося при изучении языка. В дальнейшем эта информация используется для корректировки деятельности преподавателя или модернизации учебных материалов. Анализ ошибок рассматривает освоение языка с точки зрения процесса, требующего активного участия учащихся. Теория была основана на элементах когнитивной психологии, рассматривающей ошибки как «ключ к разгадке того, что происходит в сознании учащихся»

[11]. При таком подходе ошибки рассматриваются как естественное явление, происходящее во время изучения языка до полного усвоения правил. Большая часть работы в данной области приписывается британскому лингвисту С. П. Кордеру, который в 1967 году выдвинул гипотезу, что наилучшее понимание процесса изучения языка возникает в результате глубокого систематического исследования ошибок учащихся путем выявления «их собственной, изначальной заложенной учебной стратегии» (built-in syllabus). Таким образом, последующие десятилетия лингвисты и специалисты в области лингводидактики проводили значительные исследования «естественных аспектов» изучения иностранного языка [11].

Теория анализа ошибок также получила новое развитие на материале лингвистического корпуса, благодаря появлению новых технических возможностей. Для иллюстрации данного процесса нами был выбран учебный корпус, состоящий из работ студентов, изучающих английский язык в качестве второго или иностранного языка. Письменные тексты в составе данного корпуса не подвергаются редактированию, что и подтверждает его ценность для исследования, так как система обнаружения и исправления ошибок наиболее надежна в реальных условиях. Работа проводилась на ресурсах учебного корпуса Cambridge Learner Corpus (CLC), который состоит из эссе, написанных в рамках экзаменов Кембриджского университета по английскому языку для носителей других языков (ESOL). Данный корпус содержит более 30 миллионов слов, взятых из студенческих работ. В число студентов входят представители 130 с разными родными языками. Все ошибки в корпусе подлежат аннотированию с возможностью не более чем одного исправления.

Ошибки, обнаруженные в рамках настоящего исследования, были классифицированы в соответствии с лингвистическим описанием, содержанием и организацией письменных работ. В основу классификации ошибок была положена таксономия Г. Дюлей, М. Бёрта и С. Д. Крэшена [12], в связи с чем, ошибки были классифицированы по трем уровням: лингвистические, дискурсивные и прагматические.

Под лингвистическими ошибками понимаются грамматические ошибки, которые включают в себя морфологические, лексические, синтаксические, а также ошибки в области логико-семантической, грамматической и стилистической соотнесённости и целостности текста [3, с. 116]. Морфологические ошибки в целом связаны с орфографией, неправильным использованием форм множественного числа, опущением окончания –s третьего лица единственного числа и пунктуацией. Лексические ошибки в данном исследовании связаны с неправильным образованием форм слова и коллокационными ошибками. В число синтаксических ошибок входят ошибки в использовании артиклей, предлогов, союзов, вспомогательных глаголов и местоимений, а также ошибки в употреблении времен,

наклонений и залогов. Последней категорией лингвистических ошибок в области целостности текста, являются ошибки в выборе средств когезии и когерентности. К ним относится использование повторов (лексических и частичных), синтаксического параллелизма, перифраза, эллипсисов, слов-связок (linkers) и т.д. Примером лингвистической ошибки может служить следующее предложение с ошибкой в употреблении местоимения: *Perhaps this is the reason why **them** were the first to start the war with France.*

К следующей категории ошибок ученые отнесли дискурсивные ошибки [12]. По сравнению с лингвистическими, которые относительно легко идентифицировать, дискурсивные ошибки являются скрытыми: их нелегко обнаружить, поскольку они обычно отражаются в организации текста. Такие ошибки в основном характеризуются нарушением связности идей и упорядоченности информации. В качестве примера рассмотрим фразу *Then I want to go to Ireland to attend the Dublin summer school **that** I understand English language much better*, в котором две части предложения не связаны по смыслу из-за употребления неправильного союза (в данном случае «that» необходимо заменить на «so that»).

Прагматические ошибки связаны с интерференцией родного языка учащихся в изучаемый (английский) язык. Это означает, что студенты применяют правила родного языка к правилам английского языка, тем самым производя «деформированный» английский, что наблюдается в их письменных работах [8, с. 134]. В пример можно привести предложение из работы испаноязычного студента, употребившего слово «miles» вместо «thousands» (на английский язык испанское «miles» переводится как «тысячи»): *Miles of people are now suffering from unemployment around the world.*

Для проведения исследования, каждой ошибке присваивается свой символ, согласно системе кодирования ошибок Дайан Николлс. Так, например, D означает ошибку в словообразовании (wrongly Derived) или CN обозначает ошибку в определении концепта исчисляемости существительного (countability of Noun error) [15]. Каждая ошибка была занесена в Excel, где затем проводился общий статистический анализ.

Таким образом, из общей выборки текстов (568 текста) было найдено 703 ошибки. Наибольшее количество составили лингвистические ошибки – 587. Также было найдено 116 ошибок дискурсивного характера. Прагматические ошибки не анализировались ввиду большого количества родных языков учащихся, чьи работы легли в основу содержимого корпуса. Результаты исследования представлены в таблице 1.

Количественный анализ ошибок в письменных работах учащихся

Тип ошибок	Лингвистические	Дискурсивные	Прагматические	ИТОГО
Кол-во ошибок				
Численный показатель	587	116	-	703
В % соотношении	83,5%	16,5%	-	100%

Основываясь на типологии ошибок, наиболее часто совершаемых студентами при выполнении письменных работ, и их количественном соотношении, представляется возможным разработать ряд рекомендаций для улучшения компетенций учащихся в рамках формирования навыков иноязычной письменной речи. Эмпирической базой, на которой учащиеся смогут проверить и исправить свои ошибки в данном случае выступает корпус. В качестве иллюстрации, разберем одну из найденных нами ошибок. Так, например, известно, что русскоговорящие учащиеся часто испытывают проблемы в использовании исчисляемых и неисчисляемых существительных. Благодаря корпусным данным нами были найдены две ошибки:

1. *A recent news reveals the global financial crisis is coming soon.*
2. *It is acceptable to have salt in your food ration in a little dosis.*

В первом случае неисчисляемое существительное «news» используется с артиклем, что противоречит правилу употребления артикля только с исчисляемыми существительными (a car, a banana). Поэтому в данном случае можно употребить партитивную конструкцию «a piece of»: *A recent piece of news reveals the global financial crisis is coming soon.* Во втором случае используется неверная конструкция для выражения количества. Согласно правилу, с неисчисляемыми существительными должны использоваться такие слова и выражения «little», «much», «a lot of», «a bit of», «a great deal of», «a small/big amount of» и т.д. Таким образом, целесообразно внести следующие изменения в предложение с ошибкой: *It is acceptable to take small amounts of salt in your food ration.*

Отмечая роль педагога в исправлении ошибок, стоит отметить, что наличие разметки в учебном корпусе с комментариями по исправлению ошибок оказывает неоценимую помощь студентам при самостоятельной учебно-исследовательской деятельности. На сегодняшний день создаются подобные системы, в частности испано-англоязычный проект INTELeNG (Innovación tecnológica para la enseñanza/aprendizaje de las lenguas), представляющий собой базу примеров из учебного корпуса с последовательным пояснением к коррекции ошибок [10, с. 194]. Нужно отметить, что в отличие от нашего исследования, где невозможно проследить природу прагматических ошибок учащихся ввиду разнообразия их

родных языков, в рамках двуязычных систем, таких как INTELeNG, это является реально достижимым. В особенности, такие системы проявляют себя в рамках обучения переводу, где прагматический аспект является одним из главенствующих [4, с. 56]. На основе корпуса можно создавать упражнения по переводу как часть контрастивного подхода, изучающего моменты, в которых грамматики родного и изучаемого языка подвергаются взаимной интерференции, в результате чего происходят различные ошибки. Образец упражнения был выбран из базы INTELeNG на примере английского и испанского языков:

Переведите данное предложение на английский язык, раскрыв скобки. Используйте паритивные конструкции, где это возможно.

Example: **Me dió un buen consejo (advice)** - > He gave me some good advice OR He gave me a good piece of advice

1. Tuvimos un tiempo buenísimo en vacaciones el año pasado (weather).
2. Los trabajadores causaron daños en el material y tuvieron que pagarlos (damage, equipment) etc.

Помимо того, что корпус может выступать эмпирической базой примеров ошибок, на основе его данных можно создавать коррекционные упражнения как дедуктивного, так и индуктивного типа. Под дедуктивным типом упражнений подразумеваются те, которые направлены на формирование у учащихся умения устанавливать причинно-следственные связи, переходя от общего понимания лингвистического явления к его частным элементам (например, правилам). При использовании упражнений индуктивного типа учащимся наоборот заранее предлагаются правила и образцы, после изучения и анализа которых они приходят к выводу о том, как функционирует та или иная языковая структура. Примером последнего может выступить упражнение по использованию наречий степени, где учащимся предлагается набор прилагательных с различными усилителями:

Данные прилагательные, глаголы и причастия часто используются с наречиями «highly», «strongly» и т.д. Основываясь на примерах, попробуйте определить значение, которое приобретает прилагательное/глагол/причастие.

Closely: related...

Highly: recommended, respected...

Greatly: desired, encouraged...

Strongly: believed, competitive...

В заключение, хотелось бы сказать, что возможности корпусной лингвистики кажутся практически безграничными, и в контексте лингводидактики находят свое отражение как в выявлении и анализе ошибок учащихся, так и в создании корректирующих упражнений на материале корпуса. Стоит признать, что из-за ограниченности используемых корпусных

ресурсов, и, возможно, недостатка опыта в данной сфере деятельности, данное исследование нельзя назвать полным. В будущем планируется проведение более сложного лингвистического анализа ошибок на более обширной эмпирической базе. Однако есть надежда, что анализ распространенных ошибок при письме и перечисленные в данной статье практические рекомендации, окажут некоторую помощь преподавателям, которые задаются вопросом, как бороться с ошибками студентов, а студентам, изучающим иностранный язык, помогут повысить свою общую компетентность в области письменной речи.

СПИСОК ЛИТЕРАТУРЫ

1. Вербицкая М. В., Махмурян К. С. Методические рекомендации для учителей, подготовленные на основе анализа типичных ошибок участников ЕГЭ 2019 по иностранным языкам // Педагогические измерения. – Москва, 2019. – №03. – С. 41–69.
2. Горина О. Г. Использование технологий корпусной лингвистики для развития лексических навыков студентов-регионоведов в профессионально-ориентированном общении на английском языке: дис. ... канд. пед. наук. – М., 2014 – 321 с.
3. Ибишова В. Н. Когезия как внутренняя логико-семантическая структура текста // Фундаментальные и прикладные исследования: проблемы и результаты. – Новосибирск, 2015. – № 20. – С. 115–126.
4. Ивлева А. Ю. Особенности языковой подготовки бакалавров, обучающихся по профилю «Перевод и переводоведение» на младшем этапе обучения // Иностранные языки: лингвистические и методические аспекты. – Тверь, 2015. – № 32. – С. 54–61.
5. Назарова Г. И. Дидактические и психолого-лингвистические факторы построения системы обучения письменной речи // Вестник Казанского государственного педагогического университета. – Казань, 2003. – № 01. – С. 108–112.
6. Сафонкина О. С., Иргизова К. В. Использование корпусной лингвистики в условиях цифрового образовательного пространства // Нижегородское образование. – Нижний Новгород, 2019. – №02. – С. 112–117.
7. Смирнова Н. В., Щемелева И. Ю. Роль письма в современном университете: анализ зарубежной практики обучения академическому письму // Вестник Санкт-Петербургского университета. Серия 9 «Филология, востоковедение, журналистика». – СПб., 2015. – №02. – С. 185–197.
8. Тишулин П. Б. Виды языковых ошибок и возможности их исправления при обучении иностранному языку // Известия высших учебных заведений. Поволжский регион. Гуманитарные науки. – Пенза, 2012. – № 1 (21). – С. 132–137.

9. Aijmer K., Altenberg B. English Corpus Linguistics: Studies in Honor of Jan Swartvik // Crystal D. Stylistic profiling. – New York: Routledge, 2013. – P. 221-239.
10. Campoy-Cubillo M. C., Belles-Fortuno B., Gea-Valor M. L. Corpus-Based Approaches to English Language Teaching // Mendikoetxea A., Bielsa M., Rollinson P. Focus on Errors: Learner Corpora as Pedagogical Tools. – London: Continuum International Publishing Group, 2011. – P. 180-195.
11. Corder S. P. The Significance of Learners' Errors [Электронный ресурс]. – Режим доступа: <https://www.uky.edu/~tmclay/Corder%201967.pdf> (дата обращения 30.09.2020).
12. Dulay H., Burt M., Krashen S. Language Two [Электронный ресурс]. – Режим доступа: <https://qdoc.tips/dulay-burt-amp-krashen-1982-languagetwo-pdf-free.html> (дата обращения 30.09.2020).
13. Granger S. Error-Tagged Learner Corpora and CALL: A Promising Synergy [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/228602144_Error-tagged_learner_corpora_and_CALL_A_promising_synergy (дата обращения 30.09.2020).
14. Korovina I.V., Safonkina O. S. Effective writing: tips and practice: учебное пособие. – Саранск: МГУ им. Н. П. Огарёва, 2019. – 108 с.
15. Nicholls D. The Cambridge Learner Corpus: Error Coding and Analysis for Lexicography and ELT [Электронный ресурс]. – Режим доступа: https://www.academia.edu/22736960/The_Cambridge_Learner_Corpus_error_coding_and_analysis_for_lexicography_and_ELT (дата обращения 30.09.2020).