

Интерес к электронным учебным корпусам (computer learner corpora) в преподавании ИЯ объясняется возможностью анализировать образцы письменных и устных высказываний, созданными пользователями языка - не носителями (learners). Специально разработанные программы (инструменты) позволяют автоматизировать процессы анализа языковых данных, статистической обработки и поиска информации, выявлять нарушения норм языкового выражения, что представляет определенную ценность не только для лингвистов, но и для пользователей.

В рамках описываемого проекта учебный корпус REALEC использовался для анализа эссе в формате IELTS студентов ГУ ВШЭ согласно частотности и уровню употребления лексических единиц, академической лексики и некоторых других параметров.

IELTS (International English Language Testing System) - система тестирования по английскому языку, признанная в 120 странах, позволяет определить уровень владения языком. Успешная сдача экзамена дает возможность студентам российских вузов продолжить образование или пройти стажировку в за рубежом, в англоязычной стране.

Вот уже несколько лет студенты бакалавриата ВШЭ сдают итоговый экзамен в формате IELTS Academic. Экзамен IELTS состоит из 4 частей, направленных на проверку определенного навыка - чтения, письма, восприятия на слух, речи и имеет два формата - академический и общий. Письменная часть экзамена длится 60 минут и состоит из двух частей - описание графической информации и написание эссе. За отведенное время студентам необходимо выполнить письменные задания (в объеме 150 и 250 слов соответственно), которые будут оценены согласно следующим критериям: task response, coherence and cohesion (логичность и связанность), lexical resource (лексическая насыщенность), grammatical range and accuracy (грамматическое разнообразие и точность).

Известно, что знание лексики при изучении иностранного языка является одним из главных показателей языкового уровня.

Эксперимент В рамках исследовательского проекта Высшей школы экономики был проведен эксперимент по автоматическому оцениванию лексического наполнения письменных работ студентов бакалавриата в формате IELTS (lexical resource). Работы, оцененные экспертами подверглись автоматическому оцениванию и сравнению со всем корпусом работ. При этом, такие критерии как task response, coherence and cohesion, grammatical range and accuracy не анализировались. Целью данного эксперимента являлось установление взаимосвязи между оценкой, поставленной экзаменатором, и автоматическим

оцениванием лексического наполнения работ на основе выделенных критериев. Была выдвинута гипотеза, что при помощи разработанного приложения можно предварительно оценить лексическую вариативность письменных работ методом сравнения данной работы со всем корпусом работ по ряду критериев, а также путем определения уровня слов по общеевропейской шкале CEFR и частотности слова (СОСА).

Итог: Данный эксперимент показал, что оценки экзаменаторов коррелируют с параметрами, которые были выделены для автоматической оценки эссе, т.к. у работ, оцененных на 75 баллов и выше (из 100 возможных) по всем параметрам, кроме длины слова, результат был выше среднего по корпусу. Длина слова как параметр для оценивания работы не является значимой характеристикой. Разработанное приложение можно использовать для предварительного оценивания лексического наполнения письменной работы. Однако это не исключает необходимости проверки эссе вручную, поскольку такой параметр, как правильность лексического выбора может быть оценен только экзаменатором, а не машиной.

Заключение В связи с изменившимися способами проверки и оценки знаний студентов по английскому языку в соответствии с международными стандартами экзаменов по иностранному языку возникла необходимость в использовании электронных ресурсов в вузе для автоматического оценивания письменных работ, что позволит сократить время преподавателей на проверку, а также даст возможность студентам тренировать данный вид заданий самостоятельно. И первым шагом в этом направлении можно считать приложение, позволяющее оценить эссе с точки зрения лексической насыщенности. Как показал эксперимент, данное приложение может представлять мощный инструмент для решения определенных задач.

Новизна Возможность получения эссе с аннотацией и информацией об уровнях слов и частотности является инновационным и перспективным в дальнейших корпусных исследованиях и разработках на базе учебного корпуса, т.к. позволяет определить уровень лексического наполнения письменных работ до проверки экзаменатором. Данная разработка может быть также полезна для преподавателей, что особенно актуально в сложившихся условиях.

Перспективы В дальнейшем представляется перспективным доработка отдельных инструментов, таких как установление таймера, возможность переключать страницы и пр. Интересным также представляется автоматический анализ синтаксических структур.