

## НУГ. Отчет. Блок про тренажеры.

### *“Автоматическое создание лексических тренажеров для изучающих иностранный язык”.*

#### I Задачи

REALEC — учебный корпус письменных работ студентов бакалавриата НИУ ВШЭ, разработанный в Школе лингвистики в помощь студентам, изучающим английский язык как иностранный ( в частности для подготовки к написанию сочинений на экзамене, близком по формату к Academic IELTS), а также для преподавателей, подбирающих материалы и упражнения для своих студентов. В работе с корпусом проведенный нами анализ его данных, а именно, типов ошибок, которые делают студенты, сделал возможным на материалах корпуса выработать рекомендации для студентов, собрать материалы и автоматически создавать упражнения для изучающих иностранный язык. Для удобства работы преподавателей и студентов с тренажерами необходимо иметь особый интерфейс, отличающийся для групп студентов и преподавателей, в котором сгенерированные упражнения будут находиться и пополняться, редактироваться преподавателями. Выбранному интерфейсу посвящен раздел III. II.

#### II Актуальность

Создаваемые автоматические тренажеры по их формату и типу аутентичны традиционным тренажерам, созданным вручную преподавателями или тестологами. Они призваны значительно повысить эффективность аудиторной и самостоятельной подготовки студентов НИУ ВШЭ к письменной части экзамена по английскому языку, проводимого в конце второго года обучения, а преподавателям они дают возможность меньше времени тратить на подбор и ручную обработку упражнений. Пользовательский интерфейс для тренажеров позволяет студентам выбрать необходимый им курс, потренироваться в написании текстов на сложные темы в удобном формате и сразу же получить обратную связь в виде результата их работы с тренажером; преподаватели же тратят меньше времени на обработку упражнений, проверку студенческих работ и отслеживание прогресса в преодолении собственных сложностей в работе каждого студента.

#### III Было сделано

##### III. I. Упражнения

###### **1. Упражнения на коллокации.**

В процессе анализа корпуса было выявлено, что студенты делают огромное количество лексических ошибок, связанных с незнанием лексических паттернов. Устойчивые словосочетания являются языковой основой, позволяют точнее выразить свои мысли и быть понятыми, они легче в запоминании, чем отдельные лексические единицы в отрыве от контекста, а также позволяют студентам строить английские предложения не как кальки

родного языка, а как копии аутентичных коллокаций, естественно используемых носителями английского.

Первый блок упражнений был целиком посвящен коллокациям. Набор упражнений был создан на основе академического списка коллокаций Пирсона (Pearson's Academic collocation list).

### 1.1. Всего было сгенерировано 5 типов тренажеров:

- Найти соответствия (так называемые *Match collocations*) — студентам предлагается две колонки с первой и второй частями коллокаций в случайном порядке. Необходимо сопоставить одну часть с другой (см. рис. 1).

- Множественный выбор (*Multiple choice*) — студенту необходимо заполнить пропуск в предложении, на выбор ему дается 4 варианта, из которых только один правильный (см. рис. 2, рис.3).

- Заполнить пропуск (*Open cloze*) — дается предложение с пропуском, студенту необходимо заполнить его, вставив целиком правильную коллокацию (см. рис. 4).

- Соотнести коллокацию (*Word bank*) — дается текст/предложения с пропусками и к ним несколько коллокаций. Необходимо вставить правильные варианты в нужное место (см. рис. 5, рис. 6).

- Словообразование (*Word formation*) — Студенту дается предложение и некоторая исходная форма для главной части коллокации, которую необходимо преобразовать в требуемую форму и вставить в пропуск (см. рис. 7, рис.8).

Создание именно таких типов упражнений предполагает необходимость давать эти упражнения студентам в виде цельного курса, начинающегося с более простых упражнений (например, упражнения типа “найти соответствия”), которые сменяются более сложными и заканчиваются упражнением, требующим построить коллокацию, уже выученную студентом в процессе предыдущих упражнений, для заполнения пропуска в данном упражнении.

**Рис.1 Пример упражнения на соответствие коллокации**

<i>universally</i>	<i>income</i>
<i>seminal</i>	<i>accepted</i>
<i>preceding</i>	<i>work</i>
<i>total</i>	<i>new</i>
<i>entirely</i>	<i>minds</i>

**Рис.2 Пример упражнения на множественный выбор**

*Their married life in its essentials will be largely independent of their respective families: even when accommodation is shared, there will seldom be parental \_\_\_\_\_ such as exists in many societies. The relative status and esteem accorded to husband and wife will be roughly equal. Usually each will have some paid employment outside the home and each may have his or her career.*  
Choices:

1. *status*
2. *autonomy*
3. *competence*
4. *domination*

**Рис. 3** Пример упражнения на множественный выбор в интерфейсе пользователя.

**Question 11**

Not yet answered

Marked out of 1.00

Flag question

Edit question

Social support is the support given to individuals from family , friends , peers and professionals , and can be a fundamental factor in reducing \_\_\_\_\_ .

Select one:

a. awareness

b. degradation

c. healing

d. stress

**Рис. 4.** Пример упражнения “заполнить пропуск”

*Environmental and \_\_\_\_\_ have combined with agricultural techniques to produce in Japan exceptionally high yields.*

*Правильный вариант ответа - climatic conditions<sup>1</sup>.*

**Рис. 5.** Пример упражнения “соотнести коллокацию”

*Choices: renewed interest, dynamic system, stress level, unrelated topics*

1. *They could see that nature was not static and unchanging, but that it was a \_\_\_\_\_ that was ever changing.*
2. *For example, excessive noise can raise \_\_\_\_\_ and also gives the impression of a lack of privacy.*
3. *Science is a very broad field comprising many varied and seemingly \_\_\_\_\_ from, zoology to astronomy and geology to medicine.*
4. *The black feminist movement again threw the crisis of African-American masculinity and gender relations into relief, and so inspired \_\_\_\_\_ in men's studies by the late 1980s.*

**Рис. 6.** Пример упражнения “соотнести контекст с коллокацией”

---

<sup>1</sup> Предполагается, что студент выучил в курсе данную коллокацию, она относится к определенной теме, поэтому он вставит её, а не возможные другие прилагательные со словом condition или существительные с climatic.

**Question 14**  
Not yet answered  
Marked out of 1.00  
Flag question  
Edit question

The political freedom that Florentine men enjoyed serves to explain Florence's [ ] in the development of the Italian Renaissance. Responding to others - the practitioner-patient relationshipThe challenge facing the health or social care practitioner in light of this range of perspectives on health is how to respond and [ ] with others to form partnerships, be they with patients, colleagues, or members of the wider multidisciplinary team. However, as McDonough and Shaw (1993) state, it is a skill which is difficult to dissociate from listening in many ways, because these two skills often enjoy a dependency in that speaking is only very rarely carried out in isolation; it is generally an interactive skill unless an uninterrupted [ ] is being given. Probably the most famous and [ ] definition of charisma was written by Weber, meaning, 'a certain quality of an individual personality by virtue of which he is considered extraordinary and treated as endowed with supernatural, superhuman, or at least specifically exceptional powers or qualities. The extent of relevance, however, requires a [ ] of the theory and to what extent it relates to more specific elements of the nature of employment.

leading role    communicate effectively    oral presentation    widely used

fuller discussion

**Рис. 7. Пример упражнения на словообразование**

*The provision of private medicine both within and without the NHS has remained a \_\_\_\_\_ issue. (controversy)*

*Правильный ответ — controversial*

**Рис. 8. Пример упражнения на словообразование в пользовательском интерфейсе**

**Question 7**  
Not yet answered  
Marked out of 1.00  
Flag question  
Edit question

A decade of [ ] (intense) studies has identified some signals that pattern the different axes .

*Правильный ответ: intensive or intensifying*

## 1.2. Материалы и публикации

Для создания упражнений на коллокации нам нужен был корпус, содержащий примеры использования языка в академических целях носителями английского языка. В качестве таких корпусов мы выбрали два известных английских ресурса: 1) Британский национальный корпус (British National Corpus) и 2) Британский Академический письменный корпус (British Academic Writing Corpus). При создании упражнений были использованы последние корпусные технологии, а также методы глубокого обучения (deep learning, в частности технология word2vec для создания упражнения на множественный выбор). Упражнения были испробованы студентами и сравнены со стандартными, вручную созданными упражнениями из учебников. Мы доказали, что упражнения, созданные на базе корпуса BAWE, успешно прорешиваются студентами наравне с ручными упражнениями из учебников. Данные исследования об упражнениях на коллокации были описаны в статье и представлены на

семинаре «Компьютерная лингвистика и науки о языке», проходившего 25 апреля на Факультете Компьютерных Наук и докладе на конференции «Диалог» 2016.

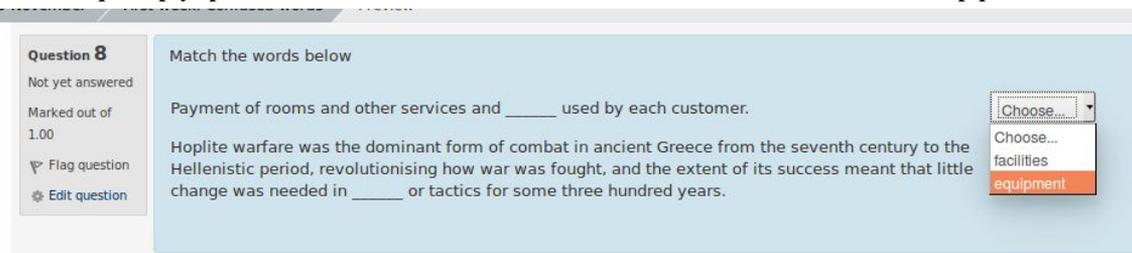
## 2. Упражнения на “смешиваемые слова” (EASILY CONFUSABLE WORDS)

Как показал лексический анализ корпуса REALEC, у студентов зачастую возникают трудности с так называемыми смешиваемыми словами (commonly confusing/confused words). Это слова, которые очень похожи между собой либо по написанию, либо произношению, либо по смыслу. Употребление подобных слов может вызывать затруднение, поэтому упражнения на них обязательны для отработки студентами.

На материале корпуса REALEC и BAWE мы сгенерировали упражнения на пары/тройки данных слов (например, raise vs rise; residence vs residents ; believes vs beliefs; say; speak; tell и многие другие). Из корпуса находятся данные пары слов и случайным образом отбираются предложения с их употреблением. Студентам предлагается упражнение вида: два предложения с пропуском, два подходящих варианта ответа — необходимо соотнести правильное слово с предложением. Пример упражнения на смешиваемые слова представлен на рисунке 9.

Из сгенерированных упражнений составлен 2-х месячный курс в онлайн-интерфейсе, в котором студенты могут планомерно отрабатывать данные слова.

**Рис. 9. Пример упражнения на “смешиваемые слова” в пользовательском интерфейсе**



## 3. Лексические упражнения на базе REALEC

Корпус REALEC может давать не только разного рода статистическую информацию об ошибках для анализа, но и быть предметом генерации самих упражнений. На каждый тип лексических ошибок можно генерировать упражнения вида — найдите ошибку в предложении и напишите правильный ответ. В корпусе уже есть вся необходимая информация для генерации: предложение с ошибкой, тип ошибки, правильный вариант. На основе корпуса, скриптами на языке Python вычленяется необходимая информация и генерируются упражнения, также осуществляется их перевод в формат moodle для пользовательского интерфейса на сайте.

Упражнения объединены в общий лексический курс. Каждое упражнение состоит из тестов на один тип ошибок. Всего лексических ошибок в REALEC 13: Word choice, Choice of lexical item, Word often confused, Choice among synonyms, Choice of a part of lexical item, Absence of certain components of a collocation, Redundant words, Derivation, Conversion, Formational affixes, Formational suffix, Formational prefix, Confusion of category. Упражнения представлены для каждой категории отдельно, плюс смешанное упражнения на все виды.

### Пример лексического упражнения на ошибку категории Word choice:

**Question 27**  
Not yet answered  
Marked out of 1.00  
Flag question  
Edit question

However, without **noticing** functions of the object we cannot understand what it is and for what it is used.

Answer:

### Пример лексического упражнения на ошибку категории Choice among synonyms:

**Question 10**  
Not yet answered  
Marked out of 1.00  
Flag question  
Edit question

As we take this idea into account, the **pluses** of the tendency of early language teaching will be inarguable.

Answer:

### Пример лексического упражнения на ошибку категории Word often confused:

**Question 2**  
Not yet answered  
Marked out of 1.00  
Flag question  
Edit question

People have been arguing about what is the main criterion of a good **work** since ancient times.

Answer:

### Пример лексического упражнения на разные виды ошибок:

**Question 1**  
Not yet answered  
Marked out of 1.00  
Flag question  
Edit question

There is a table showing us how   the comfort and service from using the planes. The percentage of people, who were satisfied with their  had increased in 8 years. So, we can easily make conclusions, that  are trying to do all the things which increase the comfort of consumers. People, who focus their attention on courtesy of the airport workers and staff have not chosen a lot in 8 years. To sum up, many passengers  about the prices and do not concentrate their attention at .

people	appreciate	fluing experience
air companies	think a lot	the problems with their comfort