**Корпусные технологии в обучении и оценке письменной научной речи русскоязычных студентов в неязыковом вузе.**

*С.А. Стринюк, к.ф.н., доцент департамента иностранных языков НИУ ВШЭ* *sstrinyuk@hse.ru*

**Введение**

Изучению научной речи на английском языке (или, как ее принято называть по-английски, Academic English) в российских вузах в настоящее время уделяется все большее внимание в связи с объективными процессами, происходящими в высшем образовании. Требование к качеству научно-исследовательской работы преподавателей вузов делает необходимым уверенное владение письменной научной речью на английском языке, с тем, чтобы результаты исследований могли быть опубликованы в международных научных изданиях.

В последние годы требования к уровню и качеству владения иностранными языками все больше распространяются и на студентов вузов. Необходимость выполнять квалификационные работы на английском языке трансформируют подходы к определению содержания курсов английского языка: вводятся курсы Academic English (научный английский), как следствие возникает необходимость отбора содержания этих курсов и критериев оценки письменной работы, выполненной в соответствии с каноном научной речи на английском языке.

Вместе с тем, очевидны объективные причины, которые препятствуют успешному выполнению этих задач:

* сокращение учебных часов, выделяемых на дисциплину
* отсутствие практики чтения (excessive reading) научных текстов на английском языке
* несформированность элементарных практических навыков письменной научной речи на английском языке

За годы проведения итогового экзамена по английскому в НИУ ВШЭ в форме письменной работы научного характера, отражающей основное содержание выпускной квалификационной работы, и ее публичной защиты на английском языке удалось собрать достаточно значительный корпус студенческих работ, которые являются ценным материалом для наблюдений над особенностями реализации конвенций научного стиля в письменной речи студентов. Этот объемный эмпирический материал, накопленный за годы проведения экзаменационных испытаний в новом формате, представляет определенный научный интерес, а практические результаты этого исследования могут быть применены для содержательного отбора материалов учебного курса Академическое письмо на английском языке.

Таким образом, **целью** нашей работы является выявление и описание ключевых маркеров научного стиля письменной научной речи на английском языке и атрибутов, его репрезентирующих с тем, чтобы в дальнейшем результаты этого исследования можно было применить в преподавании этой дисциплины. Эта цель может быть достигнута с помощью машинной обработки двух корпусов, эталонного и экспериментального, сравнительного анализа полученных практических результатов, их описания и интерпретации и выработки числовых критериев соответствия письменной речи студентов эталону.

**Функциональный стиль и навыки и умения студентов неязыкового вуза.**

Уровни языковых навыков студентов, изучающих дисциплину Academic English, можно определить, исходя из свободы владения всем разнообразием языковых средств английского языка как нормативное употребление языка, т.е. использование языковых единиц в соответствии с грамматическими нормами на базовом уровне и творческое владение языком, владение языком на уровне, который делает коммуникацию эффективной, а речь выразительной.

Трудности в овладении речепроизводством проявляются на «макроуровнях»: синтаксическом и текстовом. Поскольку стиль письменной научной речи репрезентирован в письменных текстах научно-исследовательского характера, а некоторый «знак качества» этих текстов – их публикация в научных изданиях высокого уровня, проблема недостаточного владения студентами тонкостями употребления речевых средств, присущих научной речи, может быть решена с помощью аналитического чтения научных текстов, которое включает в себя чтение научных текстов разных жанров, анализ их особенностей, структуры, закономерностей функционирования в них языковых средств. Все это будет способствовать осознанию более высокого, уверенного, уровня владения культурой речи.

Однако аналитическое чтение научных текстов является длительным процессом, который в рамках небольшого курса Академическое письмо на английском языке трудновоспроизводим в полном объеме. Применение специализированных справочных и учебных материалов может предложить способ решения задачи, не вполне удовлетворяющий следующим требованиям:

* справочные материалы подобного рода строятся на основании корпусных методов, представляют собой универсальную модель письменного научного языка, не существующую в реальности, так как не являются предметно ориентированными.
* они представляют языковые явления в изоляции, не реализованными в системе целостного научного текста, что иногда приводит к диспропорциональному воспроизведению этих речевых средств в собственных работах, как зачастую происходит с лексическими средствами логической связи («словами-связками»).

**Функциональная стилистика**

Теоретическое обоснование отбора маркеров и атрибутов научного стиля в нашей работе основывается на трудах отечественных ученых М.Н.Кожиной, М.П.Котюровой, Е.А.Баженовой и других ученых, разработавших теорию функциональной стилистики.

Функциональный стиль определяют многие факторы: значение имеет не только соблюдение речевой нормы, но и экстралингвистические факторы, которые влияют на принципы отбора и сочетания языковых средств. Экстралингвистические факторы включают в себя сферу общения, коммуникативную задачу, научный тип содержания, особую форму мышления и т.д.

**Научный стиль речи и стилеобразующие факторы**

Функциональные стили обычно интуитивно опознаваемы, поскольку легко определяется не только их сфера коммуникации и задачи общения, но и «набор» речевых средств, им присущих. В рамках научного стиля отечественные лингвисты обычно выделяют научные подстили: собственно научный, научноучебный, научнотехнический, научнопопулярный, предметно ориентированный (химия, физика, языкознание и т.д.). Научный стиль также может быть классифицирован по жанровому признаку (монографии, статьи, авторефераты, научная информация, по способу изложения (рассуждение, описание).

Стилистическим факторам присуща разная значимость в формировании стиля, они неравнозначны по стилеобразующей значимости. Существенное влияние на формирование функционального стиля оказывают экстралингвистические факторы; в числе основных стилеобразующих факторов, не являющихся собственнолингвистическими, можно назвать логичность, связность и убедительность научной речи.

Отечественными учеными предложена полевая структура функционального стиля: в центре расположены ядерные (стилеобразующие языковые явления), пограничные явления, переходные формы, периферийные/гибридные признаки стиля расположены на периферии.

Рассмотрим основные стилеобразующие факторы функционального научного стиля. Научный текст отличается целостностью, как наиболее общей характеристикой научной речи, что обуславливается его содержанием. Содержание научного текста представляет собой научное знание, он состоит из вторичных текстов, имеет четкую структуру, делим на вторичные тексты, отражающие содержание основного. Научный текст логичен, что определяется самой природой научного (понятийного) мышления. Он характеризуется высокой степенью обобщения и абстрактности, которая также определяется особенностями содержания научного текста.

Таким образом, основные первичными и вторичными признаками научной речи являются: целостность, членимость, логичность, связность, абстрактность. Периферийными признаками являются выражение авторского начала и уклонение от прямолинейных утверждений.

Целостность научного текста соотносится с его логико-семантическим единством, которое достигается благодаря системе терминов: исходные понятия (не определяемые в тексте) основные понятия (связанные с исходными и развиваемые в данном тексте) уточняющие понятия (непосредственно относящиеся к развиваемым) Целостность текста, выраженная в системе понятий логикосемантические отношения: пространства, времени, причинно-следственными целевые отношениями тождества, сходства, аналогии, противоположности, ассоциативных отношений: «часть-целое», «род-вид». (Котюрова М.П.)

Большое значение для целостности научного текста имеют вторичные (иначе – периферийные тексты). Содержание текста отражено в заглавии, оглавлении, аннотации, введении и выводах, их количество и состав варьируются, но периферийные тексты объединяют научный текст.

**Корпусные исследования в стилистике**

Корпусом обычно обозначают совокупность текстов, подобранную по определенным критериям, унифицированную и однородную, как правило, в электронном виде. Тексты в корпусе размечены с целью компьютерной обработки специализированными программными средствами. Корпусы текстов широко используются в самых различных областях филологического и, шире, гуманитарного, социального и др. знания.

Преимущества корпусных исследований заключаются в большей степени объективности исследования, анализе разнообразного языкового материала, поддающегося машинной обработке, кроме того, корпусные исследования имеют неоспоримую прикладную ценность - их можно использовать в учебном процессе как в обучении, так и в организации контроля.

Корпусные исследования с целью дальнейшего применения результатов исследования в учебном процессе могут быть основаны на сопоставлении ***эталонного корпуса*** (корпуса статей изучаемой предметной области - статей из научных журналов высокого уровня, отобранных по предметным областям, с глубиной архива не более пяти лет) и ***экспериментального корпуса***, составленного из студенческих работ, собранных по предметным областям. В исследовании также могут быть использованы два контрольных корпуса – обычно используемые в корпусных исследованиях английского языка BNC (British National Corpus) и BAWE (British Corpus of Academic English).

**Методы и исследовательский материал**

Применение корпусных технологий невозможно без специализированного программного обеспечения, поскольку процесс ручной обработки даже небольшого корпуса текстов является очень трудоемким и длительным. В нашей работе был применен исследовательский прототип, условно названный Paper Cat, инструмент, разрабатываемый группой сотрудников НИУ ВШЭ Пермь, позволяющий проводить машинную обработку естественного языка и генерировать отчеты.

Paper Cat – это исследовательский и обучающий инструмент, который позволит сделать обучение английскому языку ориентированным на предметную область и предоставит в качестве учебного материала научный текст, признанный качественным экспертами в данной области.

Функциями этого инструмента является:

* обработка эталонных корпусов по заданным критериям
* сбор данных по заданным критериям
* построение эталона – выделение наиболее значимых критериев, подкрепленных частотностью появления в текстах эталонных корпусов
* обработка корпуса студенческих работ по заданным критериям
* сбор данных по заданным критериям
* критическое осмысление полученных данных и формирование экспертной оценки – содержание и контроль качества обучения

Для проведения пилотного исследования нами были собраны два корпуса; эталонный корпус представляет собой корпус статей по тематике «компьютерные науки», в который входит 21 статья, опубликованная в англоязычных рецензируемых журналах высокого уровня по компьютерным наукам, (80654 слов); и корпус студенческих работ, включающий 28 работ русскоязычных студентов, обучающихся по направлению подготовки бизнес информатика и программная инженерия (44 942) слова.

Нами был проведен анализ речевого материала, представленного в наиболее популярных изданиях (учебных и справочного характера), предназначенных для студентов, преподавателей и научных работников, заинтересованных в приобретении устойчивых навыков владения иностранным языком. К существенным критериям, поддающимся поиску при машинной обработке, мы отнесли: существительные (процентное соотношение по отношению к другим частям речи), личные местоимения, that of/those of, десемантизированные глаголы (be, become, seem, remain, grow, consider) и глаголы широкой абстрактной семантики (например, be, exist, have, appear, occur, alter, continue, contribute, discuss, involve, investigate, conduct, consider и др.), видо-временные формы глаголов, залог, усилительные наречия (clearly, dramatically, completely, considerably, essentially, significantly, markedly, perfectly), атрибуты в препозитивной и постпозитивной позиции, сложные связки (not merely, but also, both and, as … as, neither … nor, the … the, not so … as), архаизмы (thereby, therewith, hereby), сложные предлоги (throughout, within, in accordance with, instead of, according to, because of, due to, regardless of), лексические коннекторы (since, therefore, (it) follow\s (that), so, thus, lead\s to, result\s in.

Мы полагаем, что эти атрибуты стиля научной письменной речи на английском языке, которые можно идентифицировать в корпусе с помощью специализированного программного обеспечения и которые были положены в основу нашей модели оценки качества научного текста. Описанные речевые единицы вошли в список аннотаций, и были применены к двум аннотированным корпусам текстов: корпусу предметно ориентированных статей и студенческих работ по теме компьютерные науки.

***Связность*** – одна ключевых характеристик научной речи, реализующаяся на разных языковых уровнях, лексическом и синтаксическом.

***Целостность текста*** определяется как логико-семантическое единство, выраженное в системе терминов. К атрибутам научного стиля, характеризующим ***целостность*** текста мы отнесли:

* terminology
* nominalization
* nouns as nominal premodifiers
* prepositional phrases as nominal postmodifiers
* appositive noun phrases

Поскольку поиск терминологии требует формирования объемного списка терминов, что заняло бы много времени, от обработки этой категории было решено отказаться, по крайней мере, на первоначальном этапе исследования, но был оставлен маркер *nominalization*, показывающий морфологический состав научной речи.

***Маркеры стиля логичность и связность*** включают в себя:

* lexical cohesive devices
* complex conjunctions (not merely, but also, both and, as … as, neither … nor, the … the, not so … as)
* archaisms (thereby, therewith, hereby)
* complex prepositions (throughout, within, in accordance with, instead of, according to, because of, due to, regardless of)
* logic connectors (since, therefore, (it) follow\s (that), so, thus, lead\s to, result\s in
* contextual referencing (pronouns)
* repetition (repeating key words)

Маркер стиля ***абстрактность*** реализован на лексико-семантическом уровне:

* Noun with abstract suffix (-ment, -ion, -ation, -ition, -tion, -sion, -f, -ness, -ce, -cy, -ity, -dom, -th, -ery, -ry, -ise, -ice, -hood, -ics, -ship)
* Noun with -or suffix
* desemanticised verbs (be, become, seem, remain, grow, consider)
* verbs of broad abstract semantics (be, exist, have, appear, occur, alter, continue, contribute, discuss, involve, investigate, conduct, consider, illustrate, assume, find, calculate, demonstrate, identify, analyse, support, challenge, examine, affect, provide, include, classify, establish)
* passive voice

В соответствии с полевой структурой стиля мы полагаем, что в ядре функционального научного стиля расположены:

* nominalization
* cohesive devices
* desemanticised and abstract vocabulary

***Периферия:***

* personal stance
* hedging

В модели на настоящем этапе исследования атрибуты стиля членимость со структурными элементами, включающими периферийные тексты, более крупные текстовые единицы (главы, параграфы и т.д.), абзацы и предложения нами не исследуются.

Предложенная нами модель атрибутов стиля научной речи может быть положена в основу принципов отбора элементов курса Academic English. На ее основании может быть организован контроль качества освоения научного стиля в его письменном варианте (проверка письменной работы), а также проведен целый ряд корпусных исследований собственно научного стиля письменной научной речи на английском языке и научноучебной речи студентов.

**Практические результаты и интерпретация**

***Таблица 1. Показатели, полученные с помощью прототипа Paper Cat по отобранным маркерам научного стиля***

|  |  |  |
| --- | --- | --- |
| **Маркер** | **Competent** | **Learners** |
| **BI\_PE** | **Law** | **Politology** | **History** | **Economics** | **BI\_PE** | **Law** | **Politology** | **History** | **Economics** |
| **Abstract semantic v** | 0,0413600 | 0,0598570 | 0,0616600 | 0,0571200 | 0,0348091 | 0,0756800 | 0,0678800 | 0,0878500 | 0,0849000 | 0,1897730 |
| **Archaisms** | 0,0000482 | 0,0002980 | 0,0000781 | 0,0001300 | 0,0001270 | 0,0002600 | 0,0000590 | 0,0001000 | 0,0001400 | 0,0003640 |
| **Complex conjunctions** | 0,0045300 | 0,0093330 | 0,0151900 | 0,0130000 | 0,0043450 | 0,0056800 | 0,0065900 | 0,0076000 | 0,0090500 | 0,0162730 |
| **Desem.verbs** | 0,0281000 | 0,0468330 | 0,0449400 | 0,0446700 | 0,0222600 | 0,0504200 | 0,0427600 | 0,0631000 | 0,0615700 | 0,1237270 |
| **Future** | 0,0013600 | 0,0035360 | 0,0028100 | 0,0015700 | 0,0012450 | 0,0193700 | 0,0080000 | 0,0150500 | 0,0072900 | 0,0352270 |
| **IPronoun** | 0,0003500 | 0,0069290 | 0,0029400 | 0,0057300 | 0,0005360 | 0,0013700 | 0,0016500 | 0,0031500 | 0,0063800 | 0,0136360 |
| **Intensifying adverbs** | 0,0017000 | 0,0020000 | 0,0024100 | 0,0016300 | 0,0010450 | 0,0013700 | 0,0010000 | 0,0017500 | 0,0021000 | 0,0045450 |
| **Logic connec** | 0,0044900 | 0,0073690 | 0,0067000 | 0,0045500 | 0,0030000 | 0,0053700 | 0,0064100 | 0,0090000 | 0,0083300 | 0,0172730 |
| **Noun** | 0,4833490 | 0,5782140 | 0,7663100 | 0,6827200 | 0,4155450 | 0,7650000 | 0,7614700 | 0,9343000 | 0,9566200 | 0,8726580 |
| **OrSuffix** | 0,0042200 | 0,0041310 | 0,0055000 | 0,0043300 | 0,0047910 | 0,0063700 | 0,1024000 | 0,0114000 | 0,0101400 | 0,0164090 |
| **PassiveVoice** | 0,0243400 | 0,0195710 | 0,0215600 | 0,0246700 | 0,0118090 | 0,0501100 | 0,0283500 | 0,0396500 | 0,0361400 | 0,0868640 |
| **Past** | 0,0200200 | 0,0272020 | 0,0445200 | 0,0981000 | 0,0116820 | 0,0181100 | 0,0164700 | 0,0275500 | 0,0563300 | 0,0541820 |
| **PersPron** | 0,0185800 | 0,0430120 | 0,0331300 | 0,0625200 | 0,0131636 | 0,0225300 | 0,0339400 | 0,0393500 | 0,0486700 | 0,0598640 |
| **Prepositive attributes** | 0,2237100 | 0,2431790 | 0,3351100 | 0,2831500 | 0,1990640 | 0,3374200 | 0,3091200 | 0,3740000 | 0,3597100 | 0,7196360 |
| **Postpositive attributes** | 0,0955400 | 0,1311670 | 0,1591600 | 0,1343000 | 0,0788910 | 0,1427400 | 0,1820000 | 0,2084500 | 0,2045700 | 0,3475910 |
| **Present** | 0,0799900 | 0,1035600 | 0,0956100 | 0,0550700 | 0,0555000 | 0,1242600 | 0,1135900 | 0,1413500 | 0,1261000 | 0,2777270 |
| **ThatThose** | 0,0002200 | 0,0005240 | 0,0003400 | 0,0006300 | 0,0002090 | 0,0000500 | 0,0000000 | 0,0001000 | 0,0001400 | 0,0000000 |
| **WePronoun** | 0,0108600 | 0,0095600 | 0,0106300 | 0,0055500 | 0,0078180 | 0,0057900 | 0,1735000 | 0,0155500 | 0,0105700 | 0,0105450 |
| **YouHeShe** | 0,0001100 | 0,0002380 | 0,0002800 | 0,0003200 | 0,0001090 | 0,0007400 | 0,0010000 | 0,0012000 | 0,0012900 | 0,0019550 |

Правая часть таблицы представляет собой результаты пилотного анализа двух типов аннотированных корпусов с помощью прототипа Paper Cat по отобранным маркерам научного стиля. Интерпретировать диаграмму можно исходя из нескольких соображений. Во-первых, она дает представление о построении многоаспектной модели, созданной с учетом основных атрибутов стиля. В этом смысле доминирование некоторых маркеров объяснимо с точки зрения теории функциональной стилистики и полевой модели стилей: некоторые атрибуты являются ключевыми, т.е. стилеобразующими. В данном случае, это морфологический состав научной речи – очевидно преобладание существительных в трех категориях, собственно количество существительных, а также большое количество групп существительных, выполняющих атрибутивную функцию.

***Диаграмма 1. Сравнение по эталонным корпусам***

Во-вторых, полученные эмпирические данные показывают, что между стилистической нормой письменного научного английского языка, принятым в различных областях знания, существует различие. Наибольший интерес представляют данные, полученные от сравнения разных эталонных корпусов. Поскольку номинализация является одной из существенных стилеобразующих категорий, дающих самый высокий количественный результат, представляет интерес сравнение данных по этому атрибуту в разных предметных областях.

Третий, не менее важный вывод, который следует из представленного отчета, в целом научноучебная речь студентов на английском языке довольно сильно отличается от речи исследователей, репрезентированной в статьях, опубликованных в англоязычных научных журналах высокого уровня. Очевидно, что количественные показатели практически по всем выделенным нами атрибутам в речи студентов более ярко выражены. Что это может означать? Судя по всему, речь студентов скорее «наукообразна», нежели научна в некоторых своих проявлениях.

***Диаграмма 2. Сравнение по двум корпусам: корпусу статей (эталонному корпусу) и крпусу студенческих работ***

**Заключение**

**Литература:**

|  |
| --- |
| Комарова А.И. Функциональная стилистика: научная речь. Язык для специальных целей (LSP). Издательство УРСС, Москва 2004Котюрова М.П. Об экстралингвистических основаниях смысловой структуры научного текста (функционально-стилистический аспект). Красноярск, 1987.Котюрова М.П., Баженова Е.А. Культура научной речи: Текст и его редактирование. 2\_е изд. Пермь, 2007 |
|  |

Кожина М.Н. К основаниям функциональной стилистики. Пермь, 1968. С. 142—206, 240—249.

Кожина М.Н. О диалогичности письменной научной речи. Пермь, 1986.

Кожина М.Н. О специфике художественной и научной речи в аспекте функциональной стилистики. Пермь, 1966.

Кожина М.Н. О функциональных семантико-стилистических категориях //

Филологические науки. 1987. № 2.

Кожина М.Н. Стилистика русского языка. М., 1993.

Скворцов Л.И. Культура русской речи Словарь справочник. Москва, 2006

Gáliková, S. Academic Writing : Selected Topics in Writing an Academic Paper. Series: Uni Slovakia, volume 3 Frankfurt am Main : Peter Lang GmbH, Internationaler Verlag der Wissenschaften. 2016

Sawaki, T. Analysing Structure in Academic Writing. Series: Postdisciplinary Studies in Discourse London : Palgrave Macmillan. 2016

Hyland, K.; Jiang, F. Is academic writing becoming more informal? In English for Specific Purposes. January 2017

Strongman, L.Academic Writing. Newcastle upon Tyne : Cambridge Scholars Publishing. 2013