

*О.Н. Камшилова, Н.С. Капотова, В.В. Разумова*

## **ИССЛЕДОВАНИЕ ДЛИНЫ И СТРУКТУРЫ ПРЕДЛОЖЕНИЯ В СПЕЦИАЛЬНОМ КОРПУСЕ ТЕКСТОВ**

### **1. Длина и структура предложения как признак аутентичности английского специального текста**

#### ***1.1. Объект исследования***

Образование многочисленных разновидностей глобального английского языка (европейский, русский, китайский, корейский английский и др.) дает обширный материал для исследования ассимиляции синтаксического строя английского языка к потребностям продуктивного, информативного и адекватного процесса коммуникации в условиях международного дискурса. Исследование фактов частных разновидностей глобального английского способствует установлению общих нормообразующих тенденций *в синтаксическом строе глобального английского языка*, которые фиксируются в тематически соотнесенных текстах, функционирующих в пределах одной коммуникативной сферы, «*между представителями определенных социальных групп или внутри обособленной области общественных отношений*»<sup>1</sup>.

#### ***1.2. Материал исследования***

Исследование проводилось на материале двух исследовательских сопоставительных корпусов псевдопараллельных специальных текстов, выровненных по предложениям и синтаксически размеченных, общим объемом около 9 000 предложений. Первый из корпусов объединил тексты научных докладов международ-

---

<sup>1</sup> *Протасова Е.Н.* Функциональная прагматика: вариант психолингвистики или общая теория языкознания? // Вопросы языкознания. 1999. № 1. С. 142–155.

ных конференций, второй включил тексты деклараций, относящихся к области внутренней и внешней политики европейских государств. Каждый из корпусов имеет два подкорпуса – тексты носителей английского языка и тексты, представляющие собой разновидности глобального английского: корейский и европейский английский.

## **2. Методика исследования длины и структуры предложения**

### **2.1. Выбор единицы исследования**

Предварительные данные о длине предложений в сопоставляемых текстах были получены на основе стандартной статистической программы, в которой в качестве единицы длины используется *графическое слово*. В результате была выявлена устойчивая тенденция: длина предложений в глобальных текстах короче в среднем на 2 графических слова в европейском английском и на 6 в корейском английском.

Для определения существенных расхождений в синтаксисе аутентичных и глобальных текстов в качестве единицы анализа была выбрана *клауза*: структурный и количественный анализ клауз в сопоставляемых корпусах дает возможность установить степень сложности предложений и глагольных форм, что реально влияет на длину предложения.

### **2.2. Выбор инструмента исследования**

Наиболее отвечающим задачам и материалу исследования оказался синтаксический анализатор ENGCG, созданный исследователями из Хельсинского университета<sup>1</sup>.

---

<sup>1</sup> [www.connexor.eu/technology/machinese/demo/syntax](http://www.connexor.eu/technology/machinese/demo/syntax) – визуальное представление; текстовое представление: [www2.lingsoft.fi/cgi-bin/engcg](http://www2.lingsoft.fi/cgi-bin/engcg) или <http://visl.sdu.dk/visl/en/parsing/automatic/parse.php>

Текстовое представление данной программы оказалось удобным при работе с большим по объему корпусом для анализа предложений выше средней длины. Дифференцированный анализ предложения, простота передачи данных в другие приложения сделали возможным использование результатов для статистического анализа и систематизации полученных данных.

Функциональные идентификаторы, расположенные после знака @: @+FAUXV – finite auxiliary verb, @+FMAINV – finite main verb, @-FMAINV – nonfinite main verb, @SUBJ – subject и пр. стали опорой для создания и анализа формальных моделей финитных и нефинитных клауз в исследуемых текстах.

### **3. Характеристика длины и структуры предложения в сопоставляемых корпусах**

#### ***3.1. Параметры длины и структуры предложений***

Сравнительный анализ структуры предложения в аутентичных и корейских английских текстах состоял из нескольких этапов.

На первом этапе было установлено преимущественное употребление простого предложения во всех текстах: его доля составляет 65% и 45% в корейском и аутентичном английском текстах соответственно. Доля сложноподчиненных предложений в корейских текстах составляет 28% от общего количества предложений в тексте, тогда как в аутентичных текстах доли простого и сложноподчиненного предложений практически равны (46% и 45% соответственно).

На основании предварительного анализа утверждалось, что разница в длине предложений по сопоставляемым корпусам составила 6 словоупотреблений. Одна финитная клауза может состоять из min 2 графических слов (модель SUBJ +FMAINV) или из 3–4 графических слов (модели SUBJ +FMAINV PCOMPL-S/O; SUBJ +FAUXV-FMAINV; SUBJ+FAUXV-FAUXV-FMAINV)

без учета второстепенных членов и распространений различного рода. Таким образом, можно предположить, что от 2 до 4 слов, составляющих указанную выше разницу, приходится на зависимые клаузы сложноподчиненных предложений.

Анализ моделей, полученных в результате автоматического парсинга, выявил качественные различия одних и тех же типов предложений. Для аутентичного предложения в целом характерна большая степень распространенности второстепенными членами, а для сложного предложения – большее количество клауз, включенных в его состав. В корейских английских текстах обнаруживается тенденция разделять сложные предложения (в основном сложносочиненные) на два или несколько простых, которая существенно влияет на показатель средней длины предложения и предпочтительный выбор простых структур:

*The state is assumed to coerce the collective interest of standards over private interests. And its decision making process has a public nature.* (KE 5)

На втором этапе исследования были рассмотрены нефинитные клаузы, которые по результатам анализа встречаются в корпусе довольно редко. Нефинитные клаузы не регулярны в корейских английских текстах, тогда как в аутентичных процент их использования стабильно высокий. Исключение составляют инфинитивные предикативные обороты. При этом в корейских текстах инфинитивные конструкции носят характер штампов и регулярно воспроизводятся в текстах.

На третьем этапе были исследованы характерные для корпуса модели сказуемых: +FMAINV, +FMAINV PCOML-S/O, +FAUXV-FMAINV, +FAUXV-FAUXV-FMAINV, +FAUXV-FAUXV PCOMPL-S. Особый интерес представляет модель +FMAINV PCOML-S/O (составное именное сказуемое), которая считается отличительной чертой глобальных текстов. Данное утверждение не подтвердилось результатами исследования, которые показали, что процент использования такой модели в двух

подкорпусах примерно одинаков и составил 17% и 20% (стабильно высокий) для аутентичных и глобальных текстов соответственно.

### ***3.2. Показатель степени сложности предложений и распространенности предложений***

Количество клауз в предложении определяет степень сложности предложения, которая может быть измерена количеством финитных и нефинитных клауз на предложение<sup>1</sup>. Чем выше коэффициент сложности, тем выше показатели длины предложения. Однако средние значения коэффициента сложности  $K_{прс}$  для аутентичных и глобальных текстов (2,3 и 2,2 соответственно) отличаются незначительно. Для выявления значимых различий был введен параметр распространенности предложения  $N_{сл}$ , определяющий значение среднего количества словоупотреблений, приходящихся на одну предикацию.

Значения параметра распространенности для аутентичных и европейских английских текстов составили 11,7 и 10,5 сл/пр соответственно. Согласно полученным значениям  $K_{прс}$ , предложение в текстах деклараций включает в среднем две предикации. Следовательно, предложение на эталонном английском будет, в среднем, на два слова длиннее. Таким образом, использование коэффициента распространенности подтверждает результаты предварительного ручного подсчета.

---

<sup>1</sup> Kučera H. Computational analysis of predicational structures in English // COLING. Volume 1: The 8th International Conference on Computational Linguistics. Tokyo, 1980. P. 32–38.