С.А. СТРОЙКОВ
(Самара)

СТРУКТУРНО-ЛИНГВИСТИЧЕСКИЙ АНАЛИЗ АНГЛОЯЗЫЧНОГО ЭЛЕКТРОННОГО ЛЕКСИКОГРАФИЧЕСКОГО ГИПЕРТЕКСТА ВИКИПЕДИЯ

Даны структурные, лексические, грамматические и синтаксические характеристики англоязычного электронного лексикографического гипертекста Википедия, выявляются средний объем информационных единиц, количество внутренних и внешних гипертекстовых ссылок. Теоретические основы исследования подкрепляются примерами из текстовой части исследуемых информационных единиц гипертекста Википедия.

Ключевые слова: *лексикографический гипертекст, метод лингвистического зондирования, гипертекстовая ссылка, структура, информационная единица, Википедия.*

Википедия представляет собой в настоящее время свободную, общедоступную мультиязычную универсальную интернет-энциклопедию, реализованную на принципах технологии Вики, которая позволяет пользователям изменять структуру и содержимое этой энциклопедии самостоятельно с помощью инструментов, представленных на ее сайте. Возможность создания и редактирования статей любым пользователем, по мнению некоторых экспертов, позволяет умышленно с целью вредительства добавлять, удалять или изменять содержание статей, чтобы скомпрометировать достоверность и авторитетность Википедии, что, в свою очередь, предоставляет право некоторым средствам массовой информации и представителям научного сообщества относиться к ней отрицательно и критично.

Однако С.А. Демченков и Н.Д. Федяева утверждают, что средства массовой информации и научное сообщество скорее преувеличивают негативные эффекты Википедии как лексикографического источника информации. Они приводят в своей работе результаты сравнительного анализа статей Википедии и Британской энциклопедии и приходят к выводу, что процент ошибок в вышеназванных лексикографических источниках примерно одинаковый. Эти ученые также ссылаются на данные социологических опросов ВЦИОМ, которые показывают, что для респондентов сведения, содержащиеся в Википедии, являются даже более достоверными, чем те, которые размещаются на официальных сайтах различных организаций [4, с. 2760]. Кроме того, четкие принципы написания статей Википедии (принцип формата энциклопедии, принцип нейтральности изложения, принцип соблюдения авторского права, принцип уважения других участников [2]) позволяют её администраторам в считанные минуты нейтрализовать большинство последствий проявления вандализма в Википедии.

Проведенные нами исследования также позволили заключить, что Википедия превосходит печатные энциклопедии в том, что содержащаяся в ней информация является более актуальной, нейтральной и независимой, поскольку каждая статья представляет собой результат коллективного труда, электронный формат предъявления энциклопедической информации позволяет отнести, по нашему мнению, Википедию к электронному лексикографическому гипертексту (ЭЛГТ), который, в свою очередь, позволяет решить ряд вопросов современной лексикографии [9].

Интерес научного сообщества к данному энциклопедическому ресурсу в последнее время увеличивается [см., напр.: 1; 3; 4; 6; 7; 9; 10], однако до сих пор электронная энциклопедия Википедия как объект лингвистического исследования не получила всестороннего описания.

В настоящей статье содержатся результаты структурно-лингвистического анализа ЭЛГТ Википедия, который был проведен при помощи метода структурного анализа, метода лингвистического зондирования и метода статистического анализа данных и который позволил выявить средний объем анализируемых статей Википедии, количество внутренних и внешних гипертекстовых ссылок, выступающих в качестве средства связи в ЭЛГТ Википедия, и описать лингвистические особенности англоязычных статей Википедии.

С помощью метода лингвистического зондирования (термин предложен нами в качестве метода отбора статей электронного лексикографического гипертекста [9]) мы отобрали по 1 произвольной статье на каждую букву английского алфавита, в результате исследование проводилось на материале 26 статей (общий объем – 225750 слов), которые явля-

© Стройков С.А., 2016

ются в нашем исследовании информационными единицами (ИЕ).

Анализ отобранных ИЕ ЭЛГТ Википедия позволил нам констатировать, что микротекст ИЕ построен по следующему принципу: по форме концептуализации знания ИЕ в энциклопедии представляет собой концепт-фрейм, имеющий определенную когнитивную структуру; по своему жанру, стилю и композиции подобная справочная ИЕ является мини-монографией, которая включает в себя достаточно полную и достоверную информацию, характеризующую предмет, явление или событие, напр.: энциклопедическая ИЕ о каком-либо государстве имеет структуру фрейма, в которую входят следующие данные: географическое положение, столица, границы, территория, население, природно-климатические условия, полезные ископаемые, исторический очерк, общественно-политический строй, административное деление, экономика, культура и др. Кроме того, микроструктура энциклопедических ИЕ ЭЛГТ Википедия характеризуется наличием большого количества элементов креолизованного характера.

В качестве примера рассмотрим структурное построение ИЕ ЭЛГТ Википедия Laboratory. Структура этой ИЕ выглядит следующим образом: 1) название (заголовок), 2) альтернативные значения данного понятия (*For other uses, see Lab (disambiguation) and Laboratory (disambiguation)*), 3) толкование данного слова (*A* ***laboratory*** *(/ləˈbɒrətəri/ or /ˈlæbərətɔri/; informally,* ***lab****) is a facility that provides controlled conditions in which scientific or technological research, experiments, and measurement may be performed),* 4) дополнительная информация *(Laboratories used for scientific research take many forms because of the differing requirements of specialists in the various fields of science and engineering. A physics laboratory might contain a particle accelerator or vacuum chamber, while a metallurgy laboratory could have apparatus for casting or refining metals or for testing their strength* <…>), 5) содержание (contents), 6) текстовая часть ИЕ, которая отражает структуру представленного содержания, 6) раздел «смотри также» (see also), 7) ссылки (references) и 8) внешние ссылки (external links) (https://en.wikipedia.org/wiki/Laboratory).

Анализ отобранных ИЕ ЭЛГТ Википедия позволил нам констатировать, что ИЕ содержат большое количество гипертекстовых ссылок и некоторое количество элементов креолизации. Так, например, ИЕ Laboratory (https://en.wikipedia.org/wiki/Laboratory) содержит 383 внутренние гипертекстовые ссылки, 89 внешних гипертекстовых ссылок и 9 элементов креолизации, а именно фотографий, при нажатии на которые происходит их увеличение на весь экран.

Для подсчета ссылок и деления их на внутренние и внешние мы воспользовались инструментом анализа внешних и внутренних ссылок Link Extractor на сайте http://pr-cy.ru [5]. Всего в данной ИЕ содержится 2176 слов, это означает, что внутренние гипертекстовые ссылки, представленные в данной ИЕ, составляют 17,6% от общего количества слов, внешние гипертекстовые ссылки составляют 4,1% от общего количества слов. Общее количество гипертекстовых ссылок анализируемой нами ИЕ составило 21,7% от общего количества слов. Результаты подсчета по всем изучаемым нами словам представлены в таблице на с. 92.

Анализ приведенной таблицы позволяет сделать вывод, что средний объем исследуемых нами ИЕ составляет 8683 слова, среднее количество внутренних гипертекстовых ссылок – 1039 (12% от общего количества слов), среднее количество внешних гипертекстовых ссылок – 219 (2,5% от общего количества слов). Все гипертекстовые ссылки в ИЕ Википедии выделяются синим цветом и подчеркиваются, после активации гипертекстовая ссылка изменяет свой цвет на фиолетовый для того, чтобы было понятно, какие ссылки читатель уже активировал.

Лингвистический анализ текстовой части англоязычных ИЕ ЭЛГТ Википедия, отобранных методом лингвистического зондирования, позволил установить, что они написаны в научном стиле, характеризуются логическим, точным и однозначным выражением мысли. Отбор лексических единиц для написания ИЕ подчиняется основной задаче адекватного донесения до реципиента описываемого явления в многообразии признаков, которые его характеризуют. В ИЕ ЭЛГТ Википедия слова используются в основных предметно-логических значениях, не были выявлены обсценная лексика, просторечные выражения и сленг. Лексические единицы ИЕ ЭЛГТ Википедия представлены тремя основными пластами: 1) общеупотребительная (нейтральная), 2) общенаучная и 3) специальная (терминологическая) лексика. При этом в некоторых случаях общеупотребительные слова становятся терминами в связи с особым характером их употребления в научном стиле речи. Исследуемые нами ИЕ ЭЛГТ Википедия характеризу-

Статистика гипертекстовых связей статей ЭЛГТ Википедия

№ п/п	ИЕ	Общее число слов	Количество внутренних гипертекстовых ссылок	Отношение к общему количеству слов, %	Количество внешних гипертекстовых ссылок	Отношение к общему количеству слов, %
1	apple	8201	1065	13	235	2,7
2	boy	4420	360	8,1	93	2,1
3	clay	2134	320	13,8	119	5,6
4	DOTA	3463	426	13,3	112	3,2
5	ester	3639	576	15,8	100	2,7
6	ferry	4765	747	15,7	103	2,1
7	game	5728	530	9,2	154	2,7
8	hypertext	3630	288	7,9	108	3
9	Internet	14107	1302	9,2	390	2,8
10	jacket	1901	423	22,3	48	2,5
11	KSL-TV	3644	52	1,4	16	0,4
12	laboratory	2176	383	17,6	89	4,1
13	mirror	8804	1010	11,4	184	2,1
14	nursing	11352	1134	10	138	1,2
15	oxygen	13305	2011	15,1	351	2,6
16	pet	4509	623	13,8	145	3,2
17	quest	2057	247	12	22	1,1
18	robot	12638	1179	9,3	300	2,4
19	sheep	13046	1318	10,1	284	2,2
20	Taiwan	23707	2629	11,1	519	2,2
21	U2	13749	1660	12,1	272	2
22	velvet	1869	468	25	63	3,4
23	Wikipedia	24138	2166	9	812	3,4
24	xylophone	3591	414	11,5	112	3,1
25	YouTube	16422	2016	12,3	543	3,3
26	Zinc	18755	3668	19,6	387	2,1

ются также наличием неологизмов, поскольку результаты научных исследований и изобретения, безусловно, приводят к порождению новых лексических единиц для их обозначения.

С точки зрения синтаксической организации предложения ИЕ ЭЛГТ Википедия характеризуются достаточно определенной и строгой системой союзной связи, которая вытекает из логически последовательной системы изложения, например: *Hypertext is text displayed on a computer display or other electronic devices with references (hyperlinks) to other text which the reader can immediately access, or where text can be revealed progressively at multiple levels of detail (also called StretchText). The hypertext pages are interconnected by hyperlinks, typically activated by a mouse click, keypress sequence or by touching the screen. Apart from text, hypertext is sometimes used to describe tables, images and other presentational content forms with hyperlinks. Hypertext is the underlying concept defining the structure of the World Wide Web, with pages often written in the Hypertext Markup Language (HTML). It enables an easy-to-use and flexible connection and sharing of information over the Internet* <…> (https://en.wikipedia.org/wiki/Hypertext).

Усложнение синтаксической структуры предложения в анализируемых нами статьях ЭЛГТ Википедия происходит также за счет использования герундиальных, причастных и инфинитивных конструкций, например: 1) <…> *The term xylophone may be used generally, to include all such instruments, such as the marimba, balafon and even the semantron* <…> (https://en.wikipedia.org/wiki/Xylophone); 2) <…> *The earliest evidence of a true xylophone is from the 9th century in southeast Asia, while a similar hanging wood instrument – a type of harmonicon – is said to have existed in 2000 BC in what is now part of China, according to the Vienna Symphonic Library* <…> (https://en.wikipedia.org/wiki/Xylophone); 3) <…> *A team at the Palo Alto*

Research Center attributed this slowing of growth to the project's increasing exclusivity and resistance to change <…> (https://en.wikipedia.org/wiki/Wikipedia).

Тексты ИЕ ЭЛГТ Википедия характеризуются наличием форм страдательного залога, например: 1) <…> *It is cultivated worldwide as a fruit tree, and is the most widely grown species in the genus Malus* <…> (https://en.wikipedia.org/wiki/Apple); 2) <…> *When he becomes an adult, he is described as a man* <…> (https://en.wikipedia.org/wiki/Boy); 3) <…> *Games are often classified by the components required to play them* <…> (https://en.wikipedia.org/wiki/Game); 4) <…> *One of the first known experiments on the relationship between combustion and air was conducted by the 2nd century BCE Greek writer on mechanics, Philo of Byzantium* <…> (https://en.wikipedia.org/wiki/Oxygen); 5) <…> *Velvet is a type of woven tufted fabric in which the cut threads are evenly distributed, with a short dense pile, giving it a distinctive feel* <…> (https://en.wikipedia.org/wiki/Velvet).

Помимо этого, исследуемые нами англоязычные ИЕ ЭЛГТ Википедия характеризуются наличием атрибутивных конструкций типа «существительное + существительное» (Noun + Noun constructions), например: 1) <…> *end-effectors and other environmental conditions may create hazards, and as such risk assessments should be done before using any industrial motion-control application* <…> (https://en.wikipedia.org/wiki/Robot); 2) <…> *the British Isles are most closely associated with sheep production* <…> (https://en.wikipedia.org/wiki/Sheep); 3) <…> *U2 have released 13 studio albums and are one of the world's best-selling music artists of all time, having sold more than 170 million records worldwide. They have won 22 Grammy Awards, more than any other band; and, in 2005, were inducted into the Rock and Roll Hall of Fame in their first year of eligibility* <…> (https://en.wikipedia.org/wiki/U2).

Проведенное исследование позволило прийти к выводу о том, что, несмотря на существование отрицательных характеристик, Википедия является в настоящее время авторитетной, актуальной, достоверной, нейтральной, построенной на четырех принципах на основе технологии Вики энциклопедией, которая представляет собой электронный лексикографический гипертекст, состоит из ИЕ и гипертекстовых ссылок, позволяет решать некоторые вопросы современной лексикографии и все больше привлекает ученых в качестве объекта научных, в том числе и лингвистических исследований. Структурно-лингвистический анализ ИЕ ЭЛГТ Википедия позволил нам заключить, что ИЕ представляет собой минимонографию и включает следующие элементы: название (заголовок), альтернативные значения понятия, толкование, дополнительную информацию, содержание, текстовую часть ИЕ, раздел «Смотри также», ссылки и внешние ссылки. Средний объем исследуемых нами ИЕ ЭЛГТ Википедия составляет 8683 слова, среднее количество внутренних гипертекстовых ссылок – 1039 (12% от общего количества слов), среднее количество внешних гипертекстовых ссылок – 219 (2,5% от общего количества слов). Текстовые части ИЕ написаны в научном стиле и представляют с точки зрения жанровой стратификации научного стиля самостоятельные энциклопедические статьи. Лексика ИЕ ЭЛГТ Википедия представлена 1) общеупотребительными, 2) общенаучными и 3) специальными единицами. Синтаксическая организация предложений статей ЭЛГТ Википедия характеризуется точно определенной системой союзной связи, которая вытекает из строгой логически последовательной системы изложения, при этом усложнение синтаксической структуры предложений в ИЕ ЭЛГТ Википедия происходит также за счет использования герундиальных, причастных и инфинитивных конструкций, текст ИЕ характеризуется наличием форм страдательного залога и атрибутивных конструкций.

Список литературы

1. Астраханцев Н.А. Автоматическое извлечение терминов из коллекции текстов предметной области с помощью Википедии // Труды ИСП РАН. 2014. №4. С. 7–20.
2. Википедия. URL: https://ru.wikipedia.org (дата обращения: 15.04.2016).
3. Волошина И.С. Параметры энциклопедического дискурса в лексикографическом описании термина *on-line* // Вестник ЧелГУ. 2014. № 23 (352). С. 13–16.
4. Демченков С.А., Федяева Н.Д. Википедия как инструмент лексикографических исследований (на материале русскоязычного корпуса статей) // Фундаментальные исследования. 2014. № 11–12. С. 2759–2763.
5. Инструмент анализа внешних и внутренних ссылок Link Extractor URL: http://pr-cy.ru/link_extractor (дата обращения: 15.04.2016).
6. Коршунов А.В. Извлечение ключевых терминов из сообщений микроблогов с помощью Википедии // Труды ИСП РАН. 2011. С. 269–282.
7. Ремизов Д.С. Создание лингвострановедческого словаря на базе технологии «wiki» // Известия РГПУ им. А.И. Герцена. 2008. № 86. С. 227–231.

8. Стройков С.А. Основные понятия лингвистической концепции электронного лексикографического гипертекста // Известия Самарского научного центра Российской академии наук. 2010. Т. 12, № 5 (3). С. 808–811.

9. Стройков С.А. Лингвистическое исследование электронного лексикографического гипертекста: проблемы и перспективы // Гипертекст как объект лингвистического исследования: материалы Всерос. науч.-практ. конф. с междунар. участием. Самара, 15 марта 2010. Самара: ПГСГА, 2010. С. 125–129.

10. Стройков С.А. Языковые и стилевые особенности англоязычного электронного гипертекста «Википедия» // Материалы III Международной научно-практической конференции «Германистика на рубеже тысячелетий» / под ред. Ю.Б. Ясаковой. Набережные Челны: Изд.-полиграф. отдел НФ ГОУ ВПО «НГЛУ им. Н.А. Добролюбова», 2010. С. 53–56.

* * *

1. Astrahancev N.A. Avtomaticheskoe izvlechenie terminov iz kollekcii tekstov predmetnoj oblasti s pomoshh'ju Vikipedii // Trudy ISP RAN. 2014. №4. S. 7–20.

2. Vikipedija. URL: https://ru.wikipedia.org (data obrashhenija: 15.04.2016).

3. Voloshina I.S. Parametry jenciklopedicheskogo diskursa v leksikograficheskom opisanii termina *online* // Vestnik ChelGU. 2014. № 23 (352). S. 13–16.

4. Demchenkov S.A., Fedjaeva N.D. Vikipedija kak instrument leksikograficheskih issledovanij (na materiale russkojazychnogo korpusa statej) // Fundamental'nye issledovanija. 2014. № 11–12. S. 2759–2763.

5. Instrument analiza vneshnih i vnutrennih ssylok Link Extractor URL: http://pr-cy.ru/link_extractor (data obrashhenija: 15.04.2016).

6. Korshunov A.V. Izvlechenie kljuchevyh terminov iz soobshhenij mikroblogov s pomoshh'ju Vikipedii // Trudy ISP RAN. 2011. S. 269–282.

7. Remizov D.S. Sozdanie lingvostranovedcheskogo slovarja na baze tehnologii «wiki» // Izvestija RGPU im. A.I. Gercena. 2008. № 86. S. 227–231.

8. Strojkov S.A. Osnovnye ponjatija lingvisticheskoj koncepcii jelektronnogo leksikograficheskogo giperteksta // Izvestija Samarskogo nauchnogo centra Rossijskoj akademii nauk. 2010. T. 12, № 5 (3). S. 808–811.

9. Strojkov S.A. Lingvisticheskoe issledovanie jelektronnogo leksikograficheskogo giperteksta: problemy i perspektivy // Gipertekst kak ob#ekt lingvisticheskogo issledovanija: materialy Vseros. nauch.-prakt. konf. s mezhdunar. uchastiem. Samara, 15 marta 2010. Samara: PGSGA, 2010. S. 125–129.

10. Strojkov S.A. Jazykovye i stilevye osobennosti anglojazychnogo jelektronnogo giperteksta «Vikipedija» // Materialy III Mezhdunarodnoj nauchno-prakticheskoj konferencii «Germanistika na rubezhe tysjacheletij» / pod red. Ju.B. Jasakovoj. Naberezhnye Chelny: Izd.-poligraf. otdel NF GOU VPO «NGLU im. N.A. Dobroljubova», 2010. S. 53–56.

Structural and linguistic analysis of the English electronic lexicographic hypertext Wikipedia

The article deals with the structural, lexical, grammar and syntactic characteristics of the English electronic lexicographic hypertext Wikipedia. It reveals the average volume of information units, the quantity of inner and outer hypertext references. The theoretical foundations of the research are proved by the examples from Wikipedia texts.

Key words: *lexicographic hypertext, method of linguistic research, hypertext reference, structure, information unit, Wikipedia.*

(Статья поступила в редакцию 19.04.2016)

Т.Г. НИКОЛАЕВА
(Самара)

ТИПЫ МАКРОСИТУАЦИЙ В АНГЛИЙСКИХ ПРЕДЛОЖЕНИЯХ, ОСЛОЖНЁННЫХ ВТОРИЧНО-ПРЕДИКАТИВНЫМИ СТРУКТУРАМИ

Описаны основные типы макроситуаций в основе семантически осложнённых предложений с вторично-предикативными структурами в современном английском языке. Макроситуация как элемент объективной реальности объединяет в себе несколько событий. Анализ данных ситуаций ведется по направлению от объективной реальности к пропозиции, а от неё − к предложению.

Ключевые слова: *пропозиция, семантически осложнённое предложение, вторично-предикативные структуры, макроситуация, объективная действительность, событие, логико-семантический уровень.*

В своих предыдущих исследованиях мы подробно освещали проблему взаимодействия понятий «ситуация» и «пропозиция», а также взаимной зависимости содержания ситуаций

© Николаева Т.Г., 2016