

УДК 811.111'33'374

Нина Борисовна Мечковская, д. филол. н.

Белорусский государственный университет, Минск, Беларусь

эл. почта: nina.mechkovskaya@gmail.com

Nina Borisovna Mechkovskaya, Dr of Sc. (Philology)

Belarusian State University, Minsk, Belarus

e-mail: nina.mechkovskaya@gmail.com

НАЦИОНАЛЬНЫЙ КОРПУС, ЧАСТОТНЫЕ И ТОЛКОВЫЕ СЛОВАРИ НА БУМАГЕ И ОНЛАЙН: УРОКИ ВЗАИМОДЕЙСТВИЯ

Показаны инновационные черты толковых (одноязычных) онлайн-словарей издательств *Macmillan Education* и *HarperCollins*, основанные на достижениях корпусной лингвистики и частотной лексикографии. Рассмотрены факторы углубляющейся диверсификации толковых онлайн-словарей (избыточность и необозримость корпусных данных; трудности в демонстрации словообразовательных связей лексем; экранные и другие ограничения в онлайн-представлении и восприятии лексикографической информации).

Ключевые слова: Macmillan English Dictionary, Collins Online English Dictionary, облигаторные и факультативные разделы частотных словарей, высокочастотные слова как лексика словарных дефиниций, диверсификация онлайн-словарей.

NATIONAL CORPUS, FREQUENCY AND EXPLANATORY PRINTED AND ONLINE DICTIONARIES: INTERACTION EXPERIENCES

The innovative features of the explanatory (mono-lingual) online dictionaries of the Macmillan Education and HarperCollins publishing houses, based on the achievements of corpus linguistics and frequency lexicography, are shown. The factors of the deepening diversification of explanatory online dictionaries are considered (redundancy of corpus data; difficulties in demonstrating word-building connections of lexemes; screen and other restrictions in the online presentation and perception of lexicographic information).

Key words: Macmillan English Dictionary, Collins Online English Dictionary, obligatory and facultative sections of frequency dictionary, high-frequency words as vocabulary of dictionary definitions, diversification of online dictionaries.

1. Частотные словари, в особенности те из них, которые основаны на национальных корпусах и разработаны с использованием IT-технологий, выходят далеко за пределы обязательных для жанра частотного словаря двух разделов – ранжированного по частоте списка лемм и алфавитного списка лемм с указанием их частот. Леммы со своими *ipm*, оказываются включенными также в подразделения, классы и подклассы, объединенные по разным признакам – по типам коммуницирования (устное vs письменное), по частям речи, стилям, прагматическому потенциалу слов, по “полям интересов” исследователей (interest boxes), как, напр., «Частота названий дней», «Частота цветообозначений» в словаре Дж. Лича и сотрудников (2001). Один из впечатляющих результатов отдельного ранжирования лексики устной и письменной речи (что обозначено уже в названии словаря Лича: *Word Frequencies in Written and Spoken English*): из 50 топ-слов письменной речи только 33 леммы совпадают с 50 топ-словами устной речи.

2. Достижения частотной лексикографии и интернетных технологий преобразили всемирно известные английские толковые словари *Collins Concise Dictionary* и *Macmillan English Dictionary* [далее сокращенно: Collins; MED]. В 100-тысячном алфавитном словнике MED красным цветом выделены первые по частоте 7.500 слов; в их составе словарь различает три группы слов с разной частотой: 1) первые по частоте 2.500 слов; 2) слова с частотными рангами в диапазоне от 2.501 до 5.000; 3) лексем с рангом в диапазоне от 5.001 до 7.500. Так были найдены количественные корреляты лексикологическим метафорам типа “лексическое ядро языка”, “периферия словаря”, “основной словарный запас” и т.п. Самые частотные лексем, выявленные в MED и Collins (соответственно 2.500 и 2.000), были использованы как закрытый набор семантических компонентов и с их помощью были построены дефиниции всех лексем (в своем словаре). Так лексикографы MED и Collins, реализуя давнее правило здравого смысла и логики – не определять неизвестное через неизвестное, – фактически провели семантический компонентный анализ лексики в невиданном прежде масштабе, попутно показав практическую полезность метода.

3. В словарной статье MED информация о слове подается нумерованными блоками (или ярусами) с эскалацией содержания при переходе от одного блока к следующему. На некоторых продвинутых этапах на экранах, которые пользователь видит перейдя по очередной ссылке, появляется стандартная ссылка *THESAURUS*, ведущая на сайт словаря к информации корпуса, которая не входит собственно в словарь. Как сказано на сайте MED, «эти два ресурса [*Тезаурус* и Словарь], идут рука об руку, чтобы предоставить именно ту помощь, которая вам нужна». На сайте Collins эскалация лексикографических сведений на продвинутых этапах сопровождается ссылками на приложение *Trends* с данными своего корпуса. Таким образом, пользователи MED или Collins, обращаясь к информации словаря о конкретной лексеме, по мере углубления своего запроса, получают о слове информацию бóльшую, чем вмещает словарь, – информацию от Корпуса.

4. Современная толковая лексикография предстает как сосуществование ряда взаимно связанных лексикографических предприятий, у каждого из которых свои функции в общем деле: 1) национальный корпус для исследователей и преподавателей языков – это сырьевая база, содержащая множество словоупотреблений каждого слова; 2) частотный словарь определяет коммуникативную ценность разных слов; 3) онлайн 100-тысячный словарь – это ранжированная сырьевая база для бумажного толкового академического словаря объемом в 100 тысяч словарных статей и центр лингвистической культурно-образовательной платформы. Физически такой словарь возможен в двух формах существования: 1) как большая настольная книга; 2) как электронная копия настольного словаря, доступная пользователю не по отдельным словарным статьям, но постранично.