

В ожидании другой лингвистики: идея национального корпуса и «революция» в языкознании

Ю.С. Перевезенцева

*кандидат педагогических наук, доцент кафедры «Иностранные языки»
Нижегородского государственного технического университета им. Р.Е. Алексева,
khohlova@pochta.ru*

С.А. Атрощенко

*кандидат педагогических наук, доцент кафедры «Иностранные языки»
Нижегородского государственного университета им. Н.И. Лобачевского
(Арзамасский филиал), atrochshenko_s@mail.ru*

В статье исследуется проблема достоверности, а точнее, обоснованности тезиса известного филолога В.А. Плунояна о современной лингвистике как лингвистике корпусов. В качестве объекта изучения автором статьи был выбран Британский национальный корпус. Рассмотрение основных характеристик корпуса и одной из наиболее употребительных лексем английского языка – *ring* – позволили выявить не только уникальность, но и несостоятельность корпуса для изучения языка.

Ключевые слова: корпусная лингвистика, Британский национальный корпус, контекст.

Для цитаты: Дергачева И.В. Мудрость древнерусской книжности (литературные памятники в Сборнике первой трети XVII в.) [Электронный ресурс] // Язык и текст langpsy.ru. 2014. №1. URL: [http:// langpsy.ru/journal/2014/1/Nurmuhhmetov.phtml](http://langpsy.ru/journal/2014/1/Nurmuhhmetov.phtml) (дата обращения: дд.мм.гггг)

Dergacheva I.V., The wisdom of the old Russian bookishness (the literature artifacts the first triens of the XVII-th century Corpus reported) Jelektronnyj resurs «Jazik I tekst langpsy.ru», E-journal «Language and Text langpsy.ru», 2014. no. 1. Available at: http://_langpsy.ru/journal/2014/1/Nurmuhhmetov.phtml (Accessed dd.mm.yyyy)

В последнее время у многих ученых-филологов интерес вызывает вопрос о перспективах развития лингвистического познания. Не исключено, что одной из причин его возникновения стало содержание лекции известного российского лингвиста В.А. Плунояна, в которой он достаточно категорично высказывает мысль о том, что «современная

лингвистика должна быть лингвистикой корпусов. Корпус позволяет понять, каков язык на самом деле, а не каким мы хотим, чтоб он был. Корпус произвел колоссальную революцию. Корпус вернул лингвистике ее настоящий, полноправный объект» [3]. Позиция В.А. Плунгяна нашла большой отклик, особенно среди методистов и преподавателей иностранного языка. Появилось достаточное количество научных статей, сообщающих о важности и нужности использования корпусов в обучении [1; 4; 5]. Основные их преимущества были изложены в учебном пособии В.П. Захарова «Корпусная лингвистика»:

- 1) достаточно большой (репрезентативный) объем корпуса гарантирует типичность данных и обеспечивает полноту представления всего спектра языковых явлений;
- 2) данные разного типа находятся в корпусе в своей естественной контекстной форме, что создает возможность их всестороннего и объективного изучения;
- 3) однажды созданный и подготовленный массив данных может использоваться многократно многими исследователями и в различных целях;
- 4) спрос на корпусные данные совпал с появлением соответствующих технических возможностей [2, с. 3–4].

Однако нельзя забывать, что вышеизложенные преимущества в ряде случаев могут являться таковыми лишь в теории. Процесс формирования корпуса чрезвычайно трудоемкий (можно вспомнить Чешский национальный корпус, с созданием которого возникло множество трудностей; Национальный корпус русского языка, работа над которым ведется уже достаточно продолжительный период). В силу обстоятельств, любой проект корпуса может оказаться неприемлемым или не до конца реализованным. Кроме того, о корпусах нельзя говорить в общем и целом. Каждый из них по-своему уникален и поэтому заслуживает отдельного внимания и обсуждения.

В данной статье мы обратимся к наиболее известному корпусу английского языка, по образцу которого создавались многие другие, – Британскому национальному корпусу – British National Corpus (BNC) [6]. Вопрос о том, насколько данный корпус может оправдать исследовательские ожидания лингвистов и помочь в обучении языку, для нас станет основополагающим. Для его решения мы предлагаем охарактеризовать корпус в целом и на примере леммы *ring* изучить его возможности. По результатам мы, возможно, поймем, для чего предназначен BNC, а соответственно и ряд других корпусов, созданных на его основе: для исследовательской работы или для освоения языковых навыков.

Объем корпуса составляет 100 млн. слов. Разработан он был в Оксфордском университете при участии Ланкастерского университета и Британской библиотеки. Работа над созданием корпуса продолжалась с 1991 по 1994 год. Тексты, представленные в Британском национальном корпусе, отбирались по трем основным критериям: время, область, которую данный текст описывает, и тип издания. По времени все тексты принадлежат примерно одному периоду, начиная с 1975 года, исключения делались только для развлекательной литературы, поскольку некоторые из произведений очень популярны и по сей день. В BNC присутствуют литературные произведения с 1964 года, что подтверждает синхроничность корпуса. К области развлекательной литературы принадлежит 25% текстов. 75% письменных текстов были взяты из информативных изданий (наука, искусство, коммерция и финансы, досуг, социология, мировое обозрение). Учитывались также размер (количество слов), тема, обсуждаемая в тексте, имя автора, возраст, пол, место рождения, место жительства, возрастная группа людей, которым предназначен данный текст, а также уровень сложности данного текста.

Подкорпус, представляющий письменный английский язык, включает в себя газеты, периодические научные издания и журналы, издаваемые для различных возрастов, популярную научную фантастику, опубликованные и неопубликованные письма, школьные и университетские сочинения и другое.

Подкорпус устной речи включает речь добровольно вызвавшихся участвовать в проекте людей различных возрастов, проживающих в разных частях Великобритании и принадлежащих к различным социальным слоям. Разговорная речь представлена в окружении множества контекстов: от речи формальных деловых или правительственных встреч до радио-шоу и телефонных разговоров.

Все тексты Британского национального корпуса сегментированы на предложения. Словам внутри предложения присвоены соответствующие маркеры, обозначающие грамматический класс слова или его часть речи.

Весь подкорпус устной речи разделен на две примерно равные части:

- 1) демографическую часть, содержащую транскрипции спонтанных, естественных диалогов,
- 2) контекстно-управляемую часть (часть, в которой важную роль играл контекст), содержащую записи, сделанные на каких-либо публичных мероприятиях.

Полный доступ к корпусу в режиме on-line отсутствует. В этом режиме доступны далеко не все возможности поискового интерфейса, поставляемого вместе с полной версией корпуса, выдача ограничивается 50 случайными примерами.

Существует доступ к версии корпуса, разработанной Марком Дэвисом [7]. Далее проиллюстрируем возможности корпуса BNC на примере лексемы *ring*, поскольку таковая является одной из наиболее употребляемых в английском языке. Это подтверждают данные о частотности *ring* в корпусе, составляющие около 6694 случаев.

Сравним частотность *ring* в разных регистрах. Таблица 1 включает непосредственные подсчеты и нормированные подсчеты. Непосредственные подсчеты показывают количество случаев встречаемости слова в каждом регистре. Однако эти регистры включают различное количество слов. По этой причине сравнение непосредственных подсчетов нельзя использовать как критерий для вывода о большей или меньшей частотности слова в одном регистре по сравнению с другим. Достоверные основания для сравнения по регистрам обеспечивают нормированные подсчеты, которые преобразуют количество случаев встречаемости слова по стандартной шкале, обычно в пересчете на 1 млн. словоупотреблений.

Таблица 1.

Частотность RING в определенных регистрах, нормированная на 1 млн. слов

Регистр	Примерное количество слов в регистре	Непосредственные подсчеты	Нормированные подсчеты

Разговорная речь	9.963.663	1558	156,37
Художественная литература	15.909.312	1920	120,68
Журналы	7.261.990	515	70,92
Газеты	10.466.422	694	66,31
Научно-популярная литература	16.495.185	634	38,44
Научная литература	15.331.668	301	19,63
Разное	20.835.159	1072	51,45

Как видно из таблицы 1, наибольшей является частотность слова *ring* в разговорной речи (156,37).

Программа конкордансов может создавать список частотности слов, который обычно представляется в алфавитном порядке, по порядку встречаемости слова. В корпусе можно произвести подсчеты частотности для каждой грамматической формы слова (рис.1). Так, на 1000 примеров словоупотребления слово *ring* встречается как нарицательное существительное 592 раза, как имя собственное – 4 раза, как глагол – 404 раза.

Part of speech coloring: noun pronoun proper noun adjective verb adverb preposition

592	HX8	W_misc	A	B	C	are available from the Office --	RING	HILDA NOW -- We must raise the
593	G2A	W_misc	A	B	C	City Centre approx. 1 mile ,	Ring	Road approx. 2 miles . # DETAILED
594	KRT	S_brdcast_news	A	B	C	. The A415 , that is the Abingdon	Ring	Road , that is very busy there
595	J1G	W_email	A	B	C	intersection of A34 and Oxford	Ring	Road . If anyone is going past
596	K97	W_newsp_other_report	A	B	C	complex) turn right onto	Ring	Road , right hand lane to go over
597	FM1	S_speech_unscripted	A	B	C	And er all I want you to do is just	ring	the answers . Only takes about

Рис.1

Лексема *ring* образует большое число сочетаний. Проанализируем коллокации (collocates) слова. Для каждой коллокации (collocation) существует сильная тенденция ассоциироваться с одним смыслом или значением. Поэтому, выделяя наиболее частые коллокации слова, можно проанализировать его значения.

В таблице 2 приведены коллокации из 100-словного списка для слова *ring* в регистрах: научная литература и художественная литература.

Таблица 2.

Частые коллокации *ring* в двух регистрах корпуса

Научная литература	Художественная литература
Левые коллокации	

inner	13	'll	99
wedding	11	wedding	52
outer	8	engagement	19
gold	7	gold	19
Правые коллокации			
road	3	me	55
laser	3	him	46
electrodes	3	back	26
round	3	round	16

Из данных, представленных в таблице, следует, что слово *inner* является наиболее частым левым коллокатом для *ring* в научной литературе, а *me* – частым правым коллокатом в художественной литературе.

Корпусы позволяют исследовать значения слов путем использования конкордансов. Анализ смыслов слова можно проверить, просмотрев списки конкордансов, которые показывают данные словосочетания.

В корпусе BNC информация о лексеме и ее коллокатах представлена следующим образом. По каждому регистру указывается *source information* – источник информации, включающий дату, а также *expanded context* – расширенный контекст.

Например, для коллокатов *ring* в художественной литературе можно отметить интересные сходство и различия со словосочетаниями в научной литературе. Так, совместная встречаемость *wedding ring* (явно в значении – обручальное кольцо) и *gold ring* (в значении – золотое кольцо) очень похожа в двух регистрах. Сказанное иллюстрируют следующие примеры конкордансов:

Регистр – художественная литература	Регистр – научная литература
<p>Source information: (1985-1994), It might have been Jerusalem. Healy, Thomas. Edinburgh: Polygon Books, 1991, pp. 1-81. 3119 s-units.</p> <p>Expanded context:</p> <p>' Always fucking trouble,' he said.' If it's not one thing it is another.'" It's not my fault,' Phillis said,' we were robbed.' Rab pulled on his trousers. He looked for the time. But a naked wrist.' The bastard's stole my watch,' he said.'</p>	<p>Source information: (1985-1994), Investigating society. Robert G. Burgess (ed.) Harlow: Longman Group UK Ltd, 1989, pp. 32-171. 1622 s-units</p> <p>Expanded context:</p> <p>post J.B. Oh God, yes. R.B. You did what every teacher does and that is you dressed differently. J.B. I was wearing this jumper. (laughs) R.B. I don't remember that. But you dressed differently, I remember that and you also wore a</p>

<p>They were in here tae.' Rab rubbed his eyes.' They must hiv been. I was wearing my watch when I went tae bed.'" I've still got my <i>wedding ring</i>.' Phillis (her fingers had thickened, sudden weight) now wore her ring, on a length of string, around her neck.' Then ye're lucky.'" I would have felt them,' Phillis said.' I never felt a thing,' Rab said.' They could hiv cut my throat I never felt a thing.' Phillis bare-foot (she had great long dirty toes) in a tom dressing gown.' They might hiv killed the snapper,'</p>	<p>wedding ring. J.B. Yes. I did yes. Deliberately. R.B. Can you tell me... J.B. (laughs) R.B. Can you tell me why? J.B. Because Father Curtis was there and I thought you know he won't want some married woman who goes round not wearing a <i>wedding ring</i> in charge of all these kids will he? Even though I thought the job was really mine I thought I had to create the impression of somebody who was, you know, above board, for, you know, the Catholics. R.B. Had anybody said anything to you about doing that? J.B. No. No, well you see I wear a ring intermittently like I'm wearing that one today. R.B. Yes. Well as you waved your hand through the air I saw that ring there you</p>
<p>Source information: (1985-1994), On the edge. Cross, Gillian. Oxford: OUP, 1989, pp. 66-170. 3470 s-units.</p> <p>Expanded context:</p> <p>and thought he was the handsomest boy they had ever seen...'. AC4 Day Five -- Thursday 11th August # 6.30am # 'Where are you off to?' Joe shot out of the milking house as Jinny came past. Even in wellingtons, he could move as lightly as a squirrel. Jinny looked down guiltily.' Just thought I'd have a walk.' Joe watched her for a second, his long finger delicately turning the gold ring in his ear. Then he shrugged.' You can get in and do the milking. Then I should be able to get the roots hoed before breakfast. That'll put me half a morning ahead, and I could do with that.' It wasn't fair. She had forced herself out of bed especially early, without waking Oz, just so that she could have a spare half-hour or so to take a walk up Back Clough Dale. And here was Joe, brushing her plans</p>	<p>Source information: (1985-1994), Minoans: life in Bronze Age Crete. Castleden, Rodney. London: Routledge & Kegan Paul plc, 1993, pp. 68-157. 1498 s-units.</p> <p>Expanded context:</p> <p>of worship and submission. Professor Stylianos Alexiou reminds us that there were other divine boys who survived from the religion of the pre-Hellenic period -- Linos, Plutos, and Dionysos -- so not all the young male deities we see depicted in Minoan works of art are necessarily Velchanos. Jacquetta Hawkes (1968) suggests that perhaps there were once seashore shrines dedicated to a Sea Goddess. If so, they have been washed away by the sea. Nilsson (1949), thinking along the same lines, proposes that there was a Goddess of Navigation. A gold ring from the port of Mochlos shows a goddess sailing on a boat with a shrine apparently built on to the afterdeck (deleted:figure) conveying a portable shrine from one coastal site to another. The lost 'Ring of Minos' similarly showed a priestess steering a shrine-laden boat across a bay (Figure 39). On the rocky shore in the background, no less than three separate shrines or temples were shown, one with a presiding goddess indicating the sacral horns on her altar, and two with sacred trees growing out</p>

Однако коллокация *ring round* является примером того, как пара сочетающихся слов может ассоциироваться с разными смыслами в разных регистрах. Рассмотрим примеры конкордансов:

Регистр – художественная литература	Регистр – научная литература
<p>Source information: (1985-1994), Love over gold. James, Susannah. London: Corgi Books, 1993, pp. 115-290. 4105 s-units.</p> <p>Expanded context:</p> <p>though everyone knew that what she was really chasing was the ravaged-looking thriller writer she had met at the launch of his last book and slept with the very same night. Felicity had come out in a severe facial rash and spent the time either screaming or staring fixedly at the paperknife on her desk.' It's probably that face pack you put on,' Alexandra said.' Write it up.'" I can't do that unless I know for sure, can I?' shrieked Felicity.' They could sue.'" <i>Ring round</i> some of the other magazines and see if the same thing's happened to them,' suggested Alex.' The manufacturers are bound to have sent samples to other people, not just us.'" That's an idea." And for heaven's sake give me that paperknife." But supposing it hasn't happened to anyone else,' Felicity said.' Supposing I'm the only one?" Then it could be something you ate, couldn't it?</p>	<p>Source information: (1985-1994), Explaining and understanding international relations. Smith, S and Hollis, M. New York: OUP, 1991, pp. 1-113. 1519 s-units.</p> <p>Expanded context:</p> <p>any retreat, as in quantum theory in physics or for the philosophical reasons in the last chapter, is reluctant. But the most obvious fact about the social world is that what happens in it has meaning for the inhabitants. Here are four senses in which this is true and which do not apply to the atoms studied in physics or chemistry, even if biologists might want to use some concept of meaning when thinking about animals. First, people find meaning in their experience. Here we should distinguish between signs and symbols. When we say that a <i>ring round</i> the moon means rain, we refer to a connection in nature. A ring round the moon is a natural sign of or evidence of rain to come, just as a paw print in the dust is a sign that an animal has passed by. When we say that a flag at half-mast means that someone has died, we refer to a social convention that death shall be marked in this way. Natural signs and social symbols might seem to form a continuum, marked by the presence of</p>

Как видно из приведенных примеров в художественной литературе пара *ring round*, по-видимому, используется в значении глагола «*окружать*», а в научной литературе – в значении существительного с предлогом «*кольцо вокруг*».

Важнейшим типом парадигматических отношений в лексике являются семантические, а соответствующие им объединения слов по значению на основе их тесной взаимосвязи и взаимозависимости в содержательном плане, исходя из той или иной связи обозначаемых

ими понятий, занимают первостепенное место в анализе лексики. В таблице 3 представлена информация о взаимных оценках для коллокатов (*The Mutual Information score for the collocates*). Как правило, оценка 3,0 или выше показывает семантическую связь в паре.

Таблица 3.

Взаимная оценка для коллокатов слова *ring*

Коллокаты	Процент от всех вхождений коллоката со словом <i>ring</i>	Взаимная оценка
Левые коллокаты		
П	0.28	5.34
wedding	4.41	9.31
engagement	4.77	9.42
inner	1.10	7.30
gold	0.54	6.27
outer	1.35	7.60
Правые коллокаты		
me	0.11	3.98
road	0.48	6.11
him	0.07	3.41
back	0.07	3.26
round	0.12	4.13
electrodes	1.66	7.90
laser	0.30	5.43

В целом, Британский национальный корпус создает ощущение погруженности в стихию языка из-за привлечения его составителями разных типов источников. Благодаря корпусу, исследователь-филолог, человек, изучающий язык, может увидеть многообразие употребления английских слов (нами была рассмотрена одна из наиболее часто используемых лексем – *ring*), причем в самых различных сферах человеческой жизни. Однако в упомянутом преимуществе кроется и серьезный недостаток: работающий с корпусом вынужден понимать речь абсолютно разных человеческих сообществ, что порой

© 2007-2013 ГБОУ ВПО «Московский городской психолого-педагогический университет»

весьма трудно. Более того, следует допускать, что некоторые из значений лексемы успели устареть, поскольку корпус создавался 19 лет тому назад и, исходя из его общей характеристики, не обновлялся. Иными словами, корпус статичен и потому не способен точно указать на текущее состояние развития языка, что может стать серьезным упущением для ученого-лингвиста. Для человека, который не имеет достаточных языковых навыков, корпус неудобен тем, что требует от него невероятных интеллектуальных усилий, чтобы растолковать из обрывков тестов слова или целые выражения. При изучении языка для общения на бытовом уровне представляется более удобным обратиться к словарю, чем выполнить трудоёмкие манипуляции по интерпретации контекстов. В этой связи корпусные технологии, имея, несомненно, преимущества, все же пока не могут претендовать на то, чтобы стать новой парадигмой обучения английскому языку или его исследования.

Литература

1. *Евстигнеев М.Н.* Структура ИКТ-компетентности учителя иностранного языка // Язык и культура. 2011. № 1. С. 119-125.
2. *Захаров В.П.* Корпусная лингвистика: Учебно-метод. пособие. СПб, 2005. С. 3-4.
3. *Плунгян В.А.* Почему современная лингвистика должна быть лингвистикой корпусов [Электронный ресурс] //Публичные лекции на Полит.ру. [М., 2009]. URL: <http://www.polit.ru/lectures/2009/10/23/corpus/html> (Дата обращения: 13.08.2013).
4. *Сысоев П.В.* Лингвистический корпус в методике обучения иностранным языкам // Язык и культура. 2010. № 1. С. 99-111.
5. *Чернякова Т.А.* Использование лингвистического корпуса в обучении иностранному языку // Язык и культура. 2011. № 4. С. 127-132.
6. British National Corpus [Электронный ресурс] // Сайт Британского национального корпуса. URL: <http://www.natcorp.ox.ac.uk/> (Дата обращения: 15.08.2013).
7. BYU-BNC British National Corpus [Электронный ресурс] // Сайт Британского национального корпуса. URL: <http://corpus.byu.edu/bnc/> (Дата обращения: 15.08.2013).

Pending «another» linguistics: idea of the national corpus and «revolution» in study of language

Y.S. Perevezentseva,

*Ph.D, associate professor, of the Department of Foreign Languages
of Nizhny Novgorod Alekseev State technic University, khohlova@pochta.ru*

S.A. Atroshchenko,

*Ph.D, associate professor, of the Department of Foreign Languages of Nizhny Novgorod
Lobachevskij State University (Arzamas filiation)atrochshenko_s@mail.ru*

The article inquiries into the question of trustworthiness. To be specific, this article is devoted to the question of justification of Plungyan's thesis. V.A. Plungyan considered contemporary linguistics to be corpus-based linguistics. The author of the article has chosen British National Corpus as a target of research. Study corpus' general characteristics and one of the most useful lexical item – ring – gives an opportunity to show up both uniqueness and unconvincing of corpus to study/research the language.

Key words: corpus-based linguistics, British National Corpus, context.

References

1. *Yevstignejev V.N.* Struktura IKT-kompetentnosti uchitela inostrannogo jazika [Structure of the foreign language teacher IKT- competence] // Jazik I kultura [Language and culture]. 2011 № 1. P. 119-125.
2. *Zakharov V.P.* Korpysmaya lingvistika [The linguistics of the Corporuses]. St-Petersburg: 2005. P. 3-4.
3. *Plungan V.A.* Pochemu sovremennaja lingvistika dolzhna bit lingvistikoj korpusov [Why the modern linguistics must be the linguistics of the Corporuses] [Elektronnij resurs] // // Publichniye lektsii na Polit.ru.. [Moscow, 2009]. Available at: <http://www.polit.ru/lectures/2009/10/23/corpus/html> (Accessed 13.08.2013).
4. *Sisojev P.V.* Lingvisticheskiy korpus v metodike obuchenija inostrannim jazikam [The linguistics Corpus in the methods of the foreign language training] // Jazik I kultura [Language and culture]. 2010. № 1. P. 99-111.

5. *Chernakova T.A.* Ispolzovanije lingvisticheskogo korpusa v obucheniji inostrannomu jaziku [Usage of the linguistics Corpus in the foreign language training] // Jazik I kultura [Language and culture]. 2011. № 4. P. 127-132.
6. British National Corpus [Elektronnij resurs] // Sajt British National Corpus . Available at: <http://www.natcorp.ox.ac.uk/> (Accessed 15.08.2013).
7. BYU-BNC British National Corpus [Elektronnij resurs] // Sajt British National Corpus . Available at: <http://corpus.byu.edu/bnc/> (Accessed 15.08.2013).