

McCarthy, Michael. Accessing and interpreting corpus information in the teacher education context. - Language Teaching. 2008, 41. Cambridge University Press. 563-574.

Майкл Маккарти (Michael McCarthy)

**Получение и интерпретация корпусной информации:
обучение преподавателей работе с корпусом**

Корпусная информация становится все более важным элементом преподавания, однако развитию у преподавателей навыков оценки и использования корпуса по-прежнему уделяется сравнительно мало внимания. Как правило, преподаватели слышали о корпусе, но толком не знают, что это такое и как им пользоваться, опасаясь, что потребуются слишком специальные навыки, связанные с обработкой данных или информационными технологиями. Кроме того, многих интересует, не надо ли быть носителем языка, чтобы правильно понимать и использовать корпусную информацию. Чтобы ответить на эти типичные и, отметим, вполне оправданные вопросы, необходимо описать общую перспективу, включающую корпусную информацию в преподавание.

С корпусной революцией произошла смена целой научной парадигмы (в понимании Kuhn 1962). Изменился не только наш взгляд на язык и наше понимание роли устной и письменной речи, разницы между ними, изменилась технология создания учебных материалов, ресурсов и программ для преподавания языка, а также в какой-то мере и ход самих занятий. Сегодня ни один уважающий себя издатель учебной литературы не обходится без корпусных данных. Однако по-прежнему считается, что преподаватели должны пассивно потреблять эти данные. Необходимо пересмотреть отношения между преподавателем, ученым и издателем учебно-справочной литературы – преподаватель должен превратиться из пассивного пользователя основанных на корпусах дидактических материалов в активного участника «корпусной революции».

Как можно этого добиться? Чтобы стать квалифицированным пользователем корпусных ресурсов? преподаватель должен пройти определенную подготовку. Первый шаг – помочь ему понять природу корпуса, что такое корпус, как он выглядит. Я всегда объясняю преподавателям: мы каждый день пользуемся корпусом, ведь Интернет – тот же корпус, только очень большой. Мы все занимаемся поиском в этом крупнейшем в мире корпусе: мы забиваем слова или фразы и получаем тысячи ответов. Это и есть одна из основных операций, которые корпусные лингвисты совершают над корпусными базами языка – они посылают в крупные языковые базы данных запросы, устные или письменные, и получают ответы, зачастую тысячи ответов, в форме статистики (т.е. подсчет частотности, статистика сочетаемости и т.д.). Им нужно как-то осмыслить эти ответы, точно так же, как нам нужно осмыслить результаты, полученные при поиске в Интернете.

Далее, необходимо обучить преподавателей оценивать основанные на корпусе методические ресурсы – ведь только в этом случае они смогут квалифицированно формулировать свои требования к ученым-разработчикам корпусов и издателям основанных на корпусах дидактических материалов. Важный вопрос, на который должен уметь отвечать преподаватель: является ли предъявляемый ему ресурс строго корпусным или он просто учитывает корпусные данные, опирается на них (McCarthy 1998: 22f.)? Разница здесь такая: **строго корпусные материалы** – это материалы, которые абсолютно беспристрастно отражают то, что сообщает компьютер об использовании языка, нравится вам это или нет. Покойный Джон Синклер, мой учитель и наставник, был абсолютно убежден в том, что ни в коем случае нельзя корректировать данные, полученные при

анализе корпуса. Скажем, если анализ корпуса разговорной речи демонстрирует, что все люди, даже самые образованные, употребляют после *there is* форму множественного числа, значит, нравится нам это или нет, мы должны принять это как необратимое изменение в повседневной языковой практике.

Подготовка учебных материалов с учетом корпуса – это более тонкий подход. В этом случае мы отбираем то, что считаем полезным, подходящим, правильным для обучения студентов, и тогда задача преподавателя или составителя – выступить в роли посредника, фильтруя корпусную информацию для дидактических целей. В роли «фильтров» могут выступать наши представления об учебном процессе, о рамках учебной программы, особенности культуры, традиций и т.д. Поэтому, обучая учителей работе с корпусом, надо не просто рассказать им, какие кнопки они должны нажимать и предоставить их самим себе. Важно, чтобы они решили для себя: корпус направляет нас или мы сами направляем корпус в нужное нам русло?

Важный вопрос встает в тот момент, когда мы знакомимся со словарями, учебниками, справочными материалами: какой корпус использовался при их создании? Издатель, который говорит, что новый словарь базируется на корпусных данных, должен указывать, какие именно данные он использовал. Тесно связанные с этим вопросы – каков объем корпуса и как он построен. Все крупные издательства учебной литературы по английскому языку задействуют корпусные ресурсы, но абсолютно разные по объему и структуре. Размер корпуса надо оценивать не слепо в количестве словоупотреблений, а в зависимости от тех целей, для которых он использовался. Для анализа лексики, составления словарей необходимы сотни миллионов слов, поскольку даже в крупном корпусе какое-то слово или выражение может встретиться всего один-два раза. Скажем, действительно ли выражение «It's raining cats and dogs» («дождь льет как из ведра») так уж популярно или его можно услышать только на занятиях английским языком? В Кембриджском международном корпусе (http://www.cambridge.org/elt/corpus/international_corpus.htm) это выражение встречается один раз на 50 миллионов словоупотреблений; чтобы сделать какие-либо содержательные выводы об этом выражении, нам нужен корпус, состоящий как минимум из полумиллиарда слов. А вот для грамматического корпуса вполне хватит и ста миллионов слов, ведь грамматические явления встречаются гораздо чаще.

Еще один вопрос, который должен задавать себе учитель: насколько компетентны те люди, которые составляли на основе корпуса учебные материалы? Насколько они опытные, насколько хорошо сами умеют пользоваться корпусом? Я сам начал работать с корпусом 25 лет назад, но до сих пор чувствую, что пока еще только учусь. Не менее важно понять, какие именно части учебного материала действительно составлены на базе корпуса. Часто декларируется, что ресурс основан на корпусе, но в реальности из корпуса взят только словарь, а все остальное – грамматика, задания, тексты и т.д. – к корпусу никакого отношения не имеет.

Важная тема – включение в корпус данных устной разговорной речи. Первое издание знаменитого словаря COBUILD было основано исключительно на данных письменной речи, но сейчас, с упрощением технологий записи и расшифровки устных текстов, оправдать невключение или слабую представленность в корпусе устных текстов уже гораздо труднее. Только равный объем устного и письменного подкорпусов (как в десятиллионном корпусе, который я составил вместе с Рональдом Картером) позволяет сделать обоснованные выводы о различиях между письменным языком и устной речью.

Какие именно источники включены в корпус и как их ранжировать? Например, если вы хотите создать по-настоящему всесторонний корпус британских газет, надо ли учитывать, что таблоиды занимают 78% рынка, тогда как серьезные издания лишь 22%? Нужно ли при составлении корпуса сохранять эту пропорцию, чтобы было в три раза больше примеров из желтой прессы, чем из *Observer*, *Guardian* или *Independent* – то есть стремиться к тому, чтобы отразить реальный газетный язык, с которым каждый день

сталкивается британское общество? Это вопрос баланса, актуальный для любого корпуса. Стоит ли снисходить до масс? У нас по-прежнему сильно предубеждение против использования данных, полученных от людей с более низким образовательным цензом. Меня самого обвиняли в том, что в описание разговорной грамматики английского языка я включил «полуграмотных носителей» - я записывал речь людей из Кардиффа, Дублина и т.д. Роль демографических факторов, обсуждение общественных и профессиональных предрассудков и ожиданий – важные аспекты корпусной лингвистики, которые нужно донести до преподавателя. Столь же важен вопрос о включении в корпус речи неносителей языка. Сегодня все еще трудно убедить издателей учебных пособий по английскому языку в том, что было бы неплохо использовать устную и письменную речь неносителей. Это помогло бы точнее квалифицировать типичные ошибки в английском и продумывать методику борьбы с ними.

Вернемся теперь к мысли, высказанной нами в самом начале. Ситуация, при которой корпуса и корпусные материалы создаются без участия учителей, сегодня неприемлема. На настоящем этапе необходим диалог между разработчиками корпусов и преподавателями. Учитель должен формулировать свои пожелания к корпусам и ресурсам, перестать быть немым, пассивным адресатом основанных на корпусе образовательных ресурсов. Пока что такой диалог совершенно не развит; но уже сейчас можно отметить, что преподаватели, если они получают право голоса, говорят о необходимости создания специализированных, тематических корпусов. Обычные диверсифицированные корпуса оказываются недостаточными для педагогической деятельности: даже крупный, многомиллионный корпус содержит недостаточно текстов по сельскохозяйственной тематике для того, чтобы студент смог работать со специфической для этой области лексикой.

Выше речь шла об использовании ресурсов, при создании которых используются корпуса; как максимум, о предъявлении определенных требований к таким ресурсам. Но из потребителя ресурсов преподаватель должен превратиться в активного исследователя и практика корпуса. Он должен иметь широкий доступ к корпусным данным, в открытом доступе в интернете или на CD. Сегодня существует немало специализированных, ориентированных на дидактическое использование Интернет-ресурсов, из которых преподаватель может сам получать корпусную информацию. Самый заметный из них – <http://www.lex tutor.ca>; здесь можно получить консультацию, списки слов и другие полезные ссылки.

Более того, пройдя определенную подготовку, преподаватель вполне может научиться создавать свой собственный корпус, и в этом ему необходимо оказывать всяческую поддержку. Прекрасный пример этого – корпус Уолша, созданный на основе расшифровки занятий с учащимися (Walsh 2006). Основное препятствие здесь – огромное количество времени и ресурсов, необходимых для обработки материалов. Но и эта трудность может быть преодолена: в одном из проектов создания учебного корпуса участвуют совместно десятки преподавателей – материалы, предоставляемые каждым, становятся доступны для всех участников проекта.

Итак, задачи обучения преподавателей можно разбить на три этапа. На первом этапе следует объяснить преподавателям, что такое корпус и как именно его используют для подготовки учебных пособий. Второй этап – научиться критически оценивать предъявляемые ему основанные на корпусе материалы и, опираясь на эти навыки, самостоятельно отстаивать свои интересы среди ученых и издателей дидактической литературы. Наконец, третий этап – развивать у преподавателей собственные навыки использования корпуса и помогать им создавать их собственные корпуса, лучше отвечающие их нуждам. Только таким образом учителя могут из пассивных

пользователей словарей и грамматик превратиться в активных участников корпусной революции.

Литература

- Barlow, M. & S. Burdine (2006). *American phrasal verbs* (CorpusLAB Series). Houston, TX: Athelstan.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *Longman grammar of spoken and written English*. London: Longman.
- Campbell, D., C. McDonnell, M. Meinardi & B. Richardson (2007). The need for a speech corpus. *ReCALL* 19.1, 3–20.
- Carter, R. A., R. Hughes & M. J. McCarthy (2000). *Exploring grammar in context*. Cambridge: Cambridge University Press.
- Carter, R. A. & M. J. McCarthy (1995). Grammar and the spoken language. *Applied Linguistics* 16.2, 141–158.
- Carter, R. A. & M. J. McCarthy (2006). *Cambridge grammar of English*. Cambridge: Cambridge University Press.
- Carter, R. A. & D. Nunan (2001). *The Cambridge guide to teaching English to speakers of other languages*. Cambridge: Cambridge University Press.
- Collins Cobuild (1987). *Collins Cobuild English language dictionary*. London: Collins ELT.
- Edge, J. (1989). *Mistakes and correction*. London: Longman.
- Farr, F. (2005). Relational strategies in the discourse of professional performance review in an Irish academic environment: The case of language teacher education. In K. P. Schneider & A. Barron (eds.), *Variational pragmatics: The case of English in Ireland*. Berlin: Mouton de Gruyter, 203–234.
- Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness* 17.1, 25–43.
- Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- McCarthy, M. J. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M. J. & M. Handford (2004). ‘Invisible to us’: A preliminary corpus-based study of spoken business English. In U. Connor & T. Upton (eds.), *Discourse in the professions: Perspectives from corpus linguistics*. Amsterdam: John Benjamins, 167–201.
- McCarthy, M. J., J. McCarten & H. Sandiford (2004–2006). *Touchstone: Student books Levels 1–4*. Cambridge: Cambridge University Press.
- McCarthy, M. J. & F. O’Dell (2005). *English collocations in use*. Cambridge: Cambridge University Press.
- McCarthy, M. J. & F. O’Dell (2008). *Academic vocabulary in use*. Cambridge: Cambridge University Press.
- O’Keeffe, A. & F. Farr (2003). Using language corpora in language teacher education: Pedagogic, linguistic and cultural insights. *TESOL Quarterly* 37.3, 389–418.
- O’Keeffe, A., M. J. McCarthy & R. A. Carter (2007). *From corpus to classroom*. Cambridge: Cambridge University Press.
- Prodromou, L. (2008). *English as a Lingua Franca: A corpus-based analysis*. London: Continuum International.
- Rundell, M. & P. Stock (1992). The corpus revolution. *English Today* 30, 9–17; 31, 21–38; 32, 45–51.
- Scott, M. (1999, 2004). *Wordsmith tools. Software, Versions 3 and 4*. Oxford: Oxford University Press.
- Sinclair, J. McH. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

- Timmis, I. (2002). Native-speaker norms and International English: A classroom view. *ELT Journal* 56.3, 240–249.
- Vásquez, C. & R. Reppen (2007). Transforming practice: Changing patterns of participation in postobservation meetings. *Language Awareness* 16.3, 153–172.
- Walsh, S. (2006). *Investigating classroom discourse*. London: Routledge.

MICHAEL MCCARTHY is Emeritus Professor of Applied Linguistics at the University of Nottingham, UK, Adjunct Professor of Applied Linguistics at the Pennsylvania State University, USA, and Adjunct Professor of Applied Linguistics at the University of Limerick, Ireland. He is co-director (with Ronald Carter) of the five-million word CANCODE spoken English corpus project and the (co-)author and (co-)editor of more than 40 books and 80 academic articles.