

РЕЦЕНЗІЇ

А. Поликарпов, д. филол. н.*

МГУ им. М. В. Ломоносова (Россия, Москва)

ОБ ОДНОЙ РЕЦЕНЗИИ

С чувством недоумения и горечи прочитал в журнале «Мовознавство» (№5 за 2005 г., ст. 80-83) рецензию В. Широкова на монографию О. Демской-Кульчицкой «Основы национального корпуса украинської мови» (Київ, 2005. – 219 ст.). Рецензия обращает на себя внимание явно неакадемическим стилем и, что главное, наличием в ней многих терминов и сентенций, странных, а то и некорректных с позиций языковеда. Позволим себе привести несколько цитат.

«Насправді лінгвістичний корпус є далекосяжним комп'ютерним узагальненням та розвитком такого об'єкта як звичайна лексична картотека» (ст. 80).

В этом «определении корпуса» чуть ли не каждое слово вызывает вопросы. Что такое «**лингвистический корпус**»? Лингвистика – это наука о языке, лингвистический – относящийся к лингвистике: лингвистическая теория, лингвистическая энциклопедия, лингвистическое образование, лингвистический конгресс и под. В таком случае, лингвистический корпус – это некий корпус, касающийся науки о языке, но не языка. Следует сказать, что этой ошибки в употреблении термина «лингвистический» квалифицированный лингвист не допустит.

Что такое «комп'ютерне узагальнення лексичної картотеки»? «Словник української мови» в 11-и томах дает такое определение существительному «узагальнення»: «загальний висновок, загальні положення, що ґрунтуються на порівнянні окремих предметів, фактів, явищ і виявленні їх спільних рис» (СУМ, т. X, ст. 403). Исходя из этого определения, можно с полным правом утверждать, что приведенное высказывание уважаемого рецензента никакого научного, лингвистического смысла не имеет. Так написать может только человек, весьма далекий от корпусной лингвистики и вообще от языкознания, возможно, не видевший ни одного текстового корпуса. Но ведь рецензент возглавляет «Український мовно-інформаційний фонд», в котором, как он утверждает, создан «лингвистический корпус», на базе которого осуществлен ряд исследований самого разного характера (см. библиографию в конце рецензии – ст. 83). Получается, что уважаемый рецензент сам себе противоречит. Если созданный в этом учреждении корпус соответствует приведенному выше определению, то при чем здесь исследования по грамматике? Напрашивается вывод, что эти исследования не имеют отношения к корпусу, хотя и выполнены в данном учреждении.

Действительно, среди сфер возможных применений текстового корпуса есть и лексикография, создание автоматической лексической картотеки, но это – лишь одно из многих применений корпуса, как хорошо и убедительно показано в рецензируемой В. Широковым монографии (ст. 39 и дальше).

Еще одна цитата:

«Наукова проблема у цій галузі лежить у площині побудови формалізованих моделей лінгвістичної компетенції (принаймні, певних її ділянок), апробації та верифікації цих моделей та створення на їхній основі ефективно діючого програмного забезпечення» (ст. 80). Витиевато и туманно. Конечно, программное обеспечение текстового корпуса – важная задача. Но до ее выполнения должно быть решено много лингвистических научных проблем, касающихся отбора базовых единиц хранения будущего корпуса, обеспечения его репрезентативности, структуризации текстовой базы, разработки лингвистических алгоритмов, обеспечивающих использование корпуса в различных исследованиях, создание базы характеристических данных для осуществления этих алгоритмов при решении различных лингвистических задач и т.п. Все эти проблемы детально рассматриваются в обсуждаемой монографии, но не являются, судя по всему, важными для рецензента В. Широкова. Возможно, именно их он объединил под термином «лингвистическая компетенция»? Если расшифровывать, выявить действительное содержание этого зашифрованного В. Широковым обозначения, то окажется, что это самое важное знание в корпусной лингвистике – знание о принципах и способах создания и анализа корпусов текстов. Ведущим в этом знании являются

* © А. Поликарпов, 2006

фундаментальные лингвистические представления о системе языка и формах ее реализации в текстах, а не само по себе программное обеспечение для организации и анализа корпуса. Базовая корпусная информация, заключающаяся в содержании и структуре текстов корпуса, их разметке жанрово-стилистической, хронологической, источниковой, авторской и иной отнесенностью, закладывается еще при создании корпуса и в определенном смысле является далее неизменной (пока не меняется сам состав берущихся для корпуса текстов). Также закладывается в самый начальный период и далее последовательно реализуется лингвистами, работающими с данным корпусом, идеология корпуса – концепция уровня членения каждого текста корпуса, сегментации его единиц, глубина категоризации этих единиц и т.п. Идеология корпуса и базовая корпусная информация существует независимо и до разработки любого соответствующего программного обеспечения. Программные комплексы разрабатываются для обслуживания задач использования и анализа этой базовой лингвистической информации, а не наоборот. В противоположность В. Широкову мы считаем, что создание «эффективного программного обеспечения» не может быть целью корпусной лингвистики. Ее целью является решение различных лингвистических задач на контролируемом текстовом материале, представительно отображающем часть языковой практики данного общества. Считать иначе – значит заниматься непонятным программистским нарциссизмом, не имеющим ничего общего с реальной корпусной работой.

И далее:

«Ми б не стали робити цих зауважень, якби авторка з такою наполегливістю не педалувала цифри, якими вона фіксує як абсолютні, так і відносні пропорції окремих елементів свого гіпотетичного корпусу – без обґрунтування і без смислу. Бачимо, що і з філологією у пані Орісі не дуже добре» (ст. 82). Не удивительно ли, что об уровне филологической подготовки Демской-Кульчицкой рецензент судит по упорству, с которым она «педалирует цифры»? Между тем, эти цифры свидетельствуют о том, насколько серьезно автор монографии рассматривает структуру и представительность будущего корпуса. И смысл в этих ее выкладках безусловно есть, просто рецензент, видимо, не имея необходимых базовых знаний по общей и корпусной лингвистике, его не улавливает. Иное дело, что в монографии О. Демской-Кульчицкой следовало бы детальнее обосновать соотношение представления текстов разных функциональных стилей, возможно, некоторые виды текстов, исчисляющиеся всего тысячей словоупотреблений, в корпус не включать или, наоборот, серьезнее обосновать их включение; есть и опечатки в таблице 14 (ст. 104), на которые необходимо обратить внимание. Т.е., я считаю, что хотя и в монографии О. Демской-Кульчицкой «Основы национального корпуса украинської мови» можно найти недостатки, это явно не означает, что она из одних недостатков состоит. Более того, я считаю монографию О. Демской-Кульчицкой очень важным достижением украинской корпусной лингвистики, т.к. в ней представлены принципы работы по созданию национального корпуса, без обсуждения которых корпусная работа подобного масштаба просто невозможна.

В связи с этим еще одна цитата из рецензии В. Широкова:

«Цілі та завдання, які ставить собі пані Демська-Кульчицька, є надуманими і не мають нічого спільного зі справжніми проблемами сучасної корпусної лінгвістики, насамперед, з проблемою, про яку ми говорили вище: побудовою ефективних засобів автоматичного та автоматизованого маркування лінгвістичних структур у текстах, наявних у корпусі» (стор.81).

Рецензент явно не понимает, что такое создание концепции некоторого проекта. До разработки «эффективных засобів автоматичного та автоматизованого маркування лінгвістичних структур у текстах» при планировании работ по созданию серьезного корпуса надо, как минимум, разработать (а) перечень задач, которые будут решаться на основе использования корпуса, (б) концепцию отбора текстов в корпус (определить оптимальное соотношение в нем объема текстов различных функциональных сфер и типов коммуникации), (в) концепцию основных используемых единиц анализа текстов, (г) концепцию характеристики единиц корпуса различными классами помет и т.п. Подчеркнем, что сама концепция создания, структуризации и разметки корпуса базируется на представлении его разработчиков о том, какие задачи в будущем будут решаться на его основе. В итоге, «способы автоматического и автоматизированного маркирования единиц и структур текстов» могут быть очень разными, вытекающими

именно из той концепции, которая принята авторами-разработчиками корпуса. Рассуждать о таком маркировании без наличия концепции, значит обнаруживать ее отсутствие. Если же что-то делается без общего плана (концепции), то, естественно, это чревато для такой работы теми или иными тупиками. Святая вера некоторых технарей при лингвистике, что достаточно сгрести в кучу какой-то большой объем текстов, нагнать побольше программистов, и дело сделано – находится в вопиющем противоречии с действительностью. Моментами подобного грубого и малоперспективного подхода к проблеме составления национального корпуса украинского языка характеризуется и рецензия пана Широкова.

Например, это видно и по такой фразе этой рецензии:

«Це, взагалі, унікальна ситуація у мовознавчій науці – описувати властивості об'єкта, якого не існує в природі» (ст. 82). Разве рецензенту не известно, что концепция любого проекта создается до реализации самого проекта, а не тогда, когда проект уже осуществлен? Разве рецензент никогда не видел ни одного проспекта будущего большого издания, будь то грамматика, словарь или атлас? Концепция и состоит в том, что в ней как можно подробнее описываются принципы и возможные свойства еще не существующего объекта, с тем, чтобы первоначально проверить эти принципы на восприимчивости коллег и далее продуктивно использовать таким образом проверенный план при реализации проекта. Иное, т.е. работа без плана, используя лексику рецензента – просто шарлатанство.

Дело в том, что у неопитов от лингвистики часто путаются такие понятия как «корпус текстов» и «коллекция текстов». Текстовых коллекций, т.е. простых собраний (библиотек) текстовых файлов сейчас пруд пруди в Интернете. Пользоваться ими и их анализировать в некоторых отношениях может любой, имеющий доступ в Интернет. Корпус же текстов – это более организованная и структурированная на основе некоторых принципов, а также многоаспектно размеченная совокупность текстов. Он дает дополнительные возможности лингвистам для разнообразного анализа текстов и их единиц.

Видел ли кто-нибудь продукт Украинского языково-информационного фонда под названием «Украинский национальный лингвистический корпус»? Каковы принципы его создания и основные структурные характеристики? Каковы результаты анализа? По крайней мере, опубликованная в 2005-ом году (на самом деле, в 2006-ом) коллективная монография под общей редакцией В. Широкова «Корпусная лингвистика» (К.: Довіра, 2005) не содержит ясно сформулированных принципов и конструкторских решений для создания такого корпуса. Так, например, в разделе 4 «Інформаційні та лінгвістичні засади побудови українського національного лінгвістичного корпусу» (ст. 103-193) речь идет о лингвистических проблемах и лингвистических технологиях информационного общества и общества знаний; о принципах формирования и основных направлениях использования и развития УНКЛН (Украинского национального лингвистического корпуса); о грамматической подсистеме и принципах ее моделирования; о парадигматизации и лемматизации; о средствах семантической разметки. Все эти разделы с точки зрения корпусной лингвистики не являются проблемами концептуализации корпуса. Не менее далеки от концепции, а также и от корпусной лингвистики формулировки типа: «зрозуміло, що кожному суспільно-виробничому укладу відповідає характерний для цього укладу комплекс технологічних інструментів. Зокрема, для інформаційного суспільства це були інформаційно-комунікаційні технології. Для суспільства знань це, очевидно, мусять стати технології оперування знаннями причому на всьому її життєвому циклі...» (ст. 106); «Інформація є об'єктивною характеристикою об'єктивних процесів. На нашу думку, існує глибока аналогія між визначенням понять інформації, з одного боку та енергії, яка також є певною об'єктивною характеристикою, – з іншого.» (ст. 107); «Однією з домінант новітньої історії, як відомо, є змагання за джерела енергії. Особливого значення та гостроти ця проблема набула з розвитком технотронної ери, коли виявилось, що джерела енергії не є невичерпними й у 70-х роках у світі спалахнула енергетична криза» (ст. 113) и далее «Склалося враження, що відбувається збільшення виробництва при зменшенні вкладених в нього ресурсів» (ст. 113); «УНЛК належить до типу мовно-інформаційних систем, складна структура та широкий набір функцій якої спонукають до якомога глибокого осмислення, насамперед, його лінгвістичних принципів» (ст. 120). Но самыми парадоксальными в разделе, посвященном концепции создания текстового корпуса национального языка, являются даже не размышления о

глобальном мировом кризисе и его преодолении, а таблицы, присутствующие в нем: табл. 4.1 «Зіставлення властивостей енергії та інформації» (ст. 107), табл. 4.4 «Динаміка енергоспоживання в розвинених країнах у 1970-1982 рр. (в млн. тонн нафтового еквівалента)» (ст. 114).

Все это не только не является концепцией текстового корпуса, но вводит в заблуждение читателя, поскольку происходит профанация корпусной лингвистики и подмена вымыслами В. Широкова об энергии реальных проблем и задач корпусного направления в лингвистике.

Что же касается утверждения в рецензии В. Широкова о надуманности целей, которые поставила перед собой Демская-Кульчицкая в своей монографии, то с этим никак нельзя согласиться. В монографии представлена стройная, логически непротиворечивая концепция структуры будущего корпуса, сфер его использования, его предметной области, очерчены первостепенные задачи, которые необходимо решить на первом этапе его создания и последующие, касающиеся формирования реализационной стратегии (как создания корпуса, так и его использования), разработана детальная схема маркирования слов текста (определения их грамматических характеристик). Иными словами, в монографии определены лингвистические предпосылки и задачи создания корпуса, частично создана база лингвистической характеристики для будущего маркирования единиц текста. Большого от концепции вряд ли нужно требовать. Что же касается построения автоматических или автоматизированных способов самого маркирования, процесса его осуществления, то это – задача следующего этапа – этапа выработки различных способов эксплуатации и развития созданного корпуса. При этом способы маркирования могут меняться (в т.ч. и автоматизироваться во все большей степени, т.к. может накапливаться и формализоваться информация для построения алгоритмов анализа текстов на основе статистической обработки ранее проведенной и верифицированной разметки представительных обучающих выборок), однако сама концепция лингвистической базы разметки остается той же самой, если она с самого начала имела твердые лингвистические основания.

Рецензент В. Широков считает недостатком монографии О. Демской-Кульчицкой то, что в ней много места уделено обзору чужих работ. Я же думаю, что это не недостаток, а достоинство работы. А как можно создавать концепцию, не ознакомившись с уже существующим опытом коллег по созданию других текстовых корпусов? Только на основе учета опыта других и можно не только избежать некоторых часто возникающих ошибок, но и создать такой корпус, который оказался бы уникальным, выполняющим свою, особую роль. В данном случае – это роль наиболее представительного корпуса текстов современного украинского языка, который на основе первичной разметки основными морфологическими и лексико-семантическими категориями и последующего многостороннего его анализа-развития содержал бы многоуровневую и точную информацию о различных морфологических, лексических, словообразовательных, синтаксических и иных единицах текстов, употребленных в различных охарактеризованных жанрово-стилевых сферах их бытования. Именно задачу подготовки развернутого плана работ по созданию подобного корпуса и поставила перед собой О. Демская-Кульчицкая в своей монографии. И неплохо справилась с нею. Конечно, есть в ее монографии некоторые спорные места, а также отдельные решения, требующие более детального обоснования (в т.ч., иногда немотивированное употребление англоязычных терминов, вместо существующих украинских), но эти аспекты не настолько значимы, чтобы можно было сделать вывод: «ніякої концепції національного корпусу української мови у авторки немає». Просто, рецензент В. Широков, не будучи лингвистом, не увидел тех лингвистических проблем, которые подняты и решены в монографии (и без решения которых вести работы по созданию большого корпуса любого языка просто невозможно). Так бывает, когда представитель технических или физико-математических наук возомнит себя специалистом в такой непростой предметной науке, как языкознание, и пытается заменить реальные проблемы, существующие в ней, и методы, достаточно отработанные и очевидные в ней, своим ограниченным, а значит, и непрофессиональным видением этих проблем и методов.

Возможно, такое критическое рвение рецензента вызвано тем, что он усмотрел конкуренцию создаваемого с участием О. Демской-Кульчицкой корпуса тому корпусу, который создан под его руководством? Видимо, поэтому он и пытается доказать, что никакого иного корпуса, чем тот, который у него, создавать не нужно. Эта позиция не

только не свидетельствует о широком научном кругозоре рецензента, но и, возможно, говорит о его мафиозно-монополистических замашках. Будущий Национальный корпус украинского языка, концепция которого представлена в монографии О. Демской-Кульчицкой, коренным образом будет отличаться от корпуса, создаваемого в Украинском языково-информационном фонде, в частности, если учесть то некорректное определение корпуса, которое дано В. Широковым в его рецензии.

Наличие в любой национальной лингвистической традиции не одного, а многих текстовых корпусов – не недостаток, с которым надо бороться, а нормальное, даже очень полезное для развития этой области состояние. Пытаться подавлять коллег-конкурентов может лишь тот, кто возомнил себя либо носителем абсолютной истины в данной области, либо тот, кто считает, что главное в науке – только застолбить какой-либо предметный участок, заявить на него свои, неизвестно откуда взявшиеся, эксклюзивные права на него, чтобы туда никому другому доступа не было. Эта позиция, естественно, далека от норм научного общежития. Все-таки, во все времена нормой было искать истину, а не огораживаться, критиковать коллег (если есть предмет для критики), корректно конкурировать с ними, а не пытаться их уничтожить. Силовые методы приносят в науку из других областей и долго в ней не задерживаются.

Не будем далеко ходить за примером – в России уже существует или создается несколько языковых корпусов современного русского языка: Национальный корпус русского языка, Полистилевой корпус современного русского языка, Большой корпус русского языка, Корпус текстов А. С. Пушкина и др. Аналогично в США, Англии, Германии и других странах, в том числе и в других славянских, создано по несколько корпусов различного целевого назначения. Каждому ученому, если это действительно серьезный ученый, понятно, что ни одна научная проблема не может быть раз и навсегда решена каким-то одним ученым или даже коллективом. А уж на корпусном материале столько разных проблем может быть поставлено и решено! Почему же на Украине кто-то должен иметь монопольное право на истину, а также на приватизацию для кого-то лакомых участков территории науки? Наука – самая свободная и открытая сфера деятельности человека. Поэтому "стоять насмерть" на застолбленном участке и тормозить работы других коллективов не пристало ученому. Да и бесперспективно такое огораживание и стояние.

Разумеется, степень эффективности каждого из создаваемых Корпусов покажет время, хотя многое, как показывает уже имеющаяся практика, вытекает из концепции того или иного корпуса (если она, конечно же есть).

В заключение своей рецензии В. Широков упрекает члена-корреспондента НАН Украины И. Р. Выхованца и Е. Г. Городенскую за то, что они не выступили против рекомендации к печати монографии О. М. Демской-Кульчицкой: "Невже і ви знайшли у пані Орісі щось нове в українській морфології?" (ст. 83). Этот выпад рецензента свидетельствует о его полном непонимании задач как общей, так и корпусной лингвистики, а именно в той части, что разметка того или иного корпуса морфологическими пометами – это не открытие нового в морфологии, а реализация наиболее точных схем морфологической квалификации словоформ и лексем, выработанных в лингвистике к данному моменту в данной лингвистической традиции. Только на основе такой первичной разметки и возможен дальнейший углубленный лингвистический анализ морфологических проблем, например, того, какие семантические классы лексем, в каких формах, в каких синтаксических условиях (в сочетании с лексемами каких семантических и морфологических классов, в каких синтаксических конструкциях), в каких типах текстов и с какой частотой реализуют те или иные свои морфологические категории. Т.е. только на этой основе и возможно то искомое ПОЛУЧЕНИЕ НОВОГО ЗНАНИЯ, не известного ранее, которое и является целью науки. Корпусная лингвистика дает возможность для получения такого нового знания, т.к. в структуре корпусного материала синтезируется, с одной стороны, информация о естественных единицах и их конфигурациях в тексте с категориальной информацией, накопленной об этих единицах в лингвистике и закрепленной в корпусе в ходе первичной разметки текста, с другой. Автоматизированное обследование коррелятивных связей между этими и иными динамически добавляемыми типами информации приводит к выявлению скрытых структур, открытию нового знания, а точнее, подтверждению или опровержению той или иной гипотезы о структурах текстов определенного содержательного и стилового диапазона, которая вытекает из той или иной концепции.

Если В. Широков с такой постановкой вопроса не согласен, то это означает либо то, что у него имеется своя, правда, никому не известная НОВАЯ концепция украинской морфологии (что маловероятно, если судить по его работам), либо он не понимает того, о чем он пишет. Второе – более вероятно.

И последнее. В Широков в конце рецензии пишет, что она была обсуждена и одобрена на заседании ученого совета организации «Український мовно-інформаційний фонд». Хочется спросить: неужели среди членов ученого совета этой организации не нашлось специалиста, который указал бы рецензенту на ошибки в его тексте? И в редколлегии «Мовознавства» почему-то никто не заметил несуразности текста, отнюдь не поднимающего авторитет журнала? Но, как оказывается, В. Широков является заместителем редактора этого журнала. Конечно, внутреннее дело академического института, кого кем избирать или назначать в редакцию профильного для Института языкознания НАНУ органа, но закономерно возникает вопрос: неужели на Украине такой дефицит языковедов, что заместителем редактора ведущего журнала пришлось сделать доктора технических наук, явно не разбирающегося в лингвистической проблематике, что хорошо видно по написанной им рецензии, а также по его трудам? А куда смотрел редактор этого журнала? Академик В. М. Русановский достаточно серьезный ученый, он же не мог не увидеть, что рецензия В. Широкова далека от науки, от профессионализма, от нормальных принципов научной дискуссии. Почему же он одобрил ее публикацию? Чем можно объяснить такую нетребовательность?

20. 04. 2006

