

КОРПУСНАЯ РЕВОЛЮЦИЯ: ОТ «ИСКУССТВА» К «НАУКЕ»

А.И. Изотов

УДК 81'32; 811.162.3

Филологические науки

Рассматривается феномен современного гуманитарно-научного знания в его отношении к знанию естественнонаучному. Филологическое знание может быть и естественнонаучным, и гуманитарно-научным, в зависимости от того, какой именно аспект языка/литературы изучается. Развитие корпусных технологий способствует смещению расположенного на оси «искусство»-«наука» массива филологических исследований в сторону «науки»

Ключевые слова: корпусная революция, гуманитарные науки, естественные науки, филология, лингвистика

Изотов Андрей Иванович, д. филол. н.

Московский государственный университет имени М. В. Ломоносова

a.i.izotov@mail.ru

КОРПУСНАЯ РЕВОЛЮЦИЯ: ОТ «ИСКУССТВА» К «НАУКЕ»

По свидетельству ученика А. Н. Колмогорова А. Н. Ширяева (круглый стол на филологическом факультете МГУ имени М. В. Ломоносова весной 2009 года, посвященный филологическим интересам А. Н. Колмогорова), академик как-то мимоходом заметил, что математик имеет дело с **доказательствами**, а у историка доказательств нет и быть не может, а могут быть только **подтверждения** его гипотезы.

Мы полагаем, что в этих словах крупнейшего отечественного ученого двадцатого столетия отмечено принципиальное различие между естественнонаучным и гуманитарно-научным знанием. Речь идет не о традиционном делении научных дисциплин на гуманитарные (имеющие дело с человеком), естественные (имеющие дело со всем остальным окружающим миром) и точные (имеющие дело исключительно с самими собой –

математические науки), а об уникальности/неуникальности объекта исследования.

Естествоиспытатель имеет дело в основном с неуникальными объектами исследования. Знание, установленное в результате естественнонаучного эксперимента, в силу этого **верифицируемо**. Работа с результатами верифицируемого эксперимента - суть естественнонаучного исследования.

Гуманитарий же занимается, как правило, тем, что уже произошло и никогда больше не повторится («нельзя дважды войти в одну реку»). Мы не можем провести эксперимент и доказать истинность или ложность утверждения, что татаро-монголы, разбив в Силезии объединенное войско Великой и Малой Польши, Тевтонского ордена, Тамплиеров, Госпитальеров, а также добровольцев со всей Центральной Европы, опустошив Моравию, Словакию, Венгрию, Хорватию и дойдя в 1242 году до берегов Адриатики, прервали свое движение на Запад, потому что возглавлявшему поход Батю надо было торопиться в Карокорум на выборы нового кагана, а не потому, что монгольская конница была якобы бессильна против неприступных замков Центральной Европы (как принято объяснять чешским школьникам) или же потому, что монгольское войско было обескровлено героическим сопротивлением русских княжеств (как учили в школе автора настоящих строк). Анализ результатов верифицируемого эксперимента гуманитария заменяет критика источников, ср. [24, с. 42].

Поэтому естествоиспытатель оперирует доказанными фактами, а гуманитарий – данными, истинность/ложность которых установлена с большей/меньшей степенью вероятности.

Филологическое знание, как мы склонны считать, может быть и естественнонаучным, и гуманитарно-научным, в зависимости от того, какой аспект языка / литературы изучается. Если нас интересуют **неуникальные** явления (представленные в одном или нескольких языках суффиксы, приставки, синтаксические конструкции; литературные жанры и т.д.), речь идёт прежде всего о естественнонаучном знании, если же объект исследования

уникален (история конкретного языка / конкретной литературы) – прежде всего о гуманитарно-научном.

Возможна, безусловно, и комбинация естественнонаучного и гуманитарно-научного знаний, ср., например, [3; 5; 6], где речь идет, с одной стороны, об общих грамматических свойствах форм с теми или иными суффиксами, многократно употребленными в том или ином синтаксическом контексте, с другой же стороны - об историко-культурном фоне возникновения конкретной партиципальной системы в конкретном языке.

Тем не менее отнесение филологии к чему-то однозначно гуманитарному характерно как для массового сознания (ср. обидные статьи на <http://lurkmore.to/>), так и для специалистов - от позднелатинского автора Марциана Капеллы, давшего Филологии в услужение семь Свободных Искусств, до С. С. Аверинцева, определявшего филологию как «содружество гуманитарных дисциплин - лингвистич., литературоведч., историч. и др., изучающих историю и выясняющих сущность духовной культуры человечества через языковой и стилистич. анализ письменных текстов» [1, с. 410].

Эти представления не смогли разрушить ни структурализм, ни «хомскианская революция». Возможно, их сможет хотя бы поколебать «корпусная революция», о которой заговорили и у нас, см. [2].

Однако оправданно ли употребление слова «революция», когда речь идет о лингвистической дисциплине («корпусная лингвистика»), которая появилась около полувека назад, однако была до самого недавнего времени настолько никому не интересна, что у неё до сих пор нет даже собственного УДК-классификатора (приводимый обычно УДК-81'32 - это «Математическая лингвистика»)? Конечно, оправданно, и вот почему:

Во-первых, сама дисциплина развивается так стремительно, что учебные материалы по ней устаревают, едва будучи опубликованными. Например, в замечательном (без всякой иронии) учебнике [21], вышедшем в 2011 году, в Приложении 1 (Список Национальных корпусов) объем Национального корпуса русского языка (<http://ruscorpora.ru/>) определен в **70 млн.** слов, тогда

как уже в январе 2012 года этот объем равнялся **364 млн.** Что же касается **полуторамиллиардного** на сегодняшний день Национального корпуса чешского языка (<http://ucnk.ff.cuni.cz/>), то он описывается в этом учебнике как ещё 100-миллионный.

Во-вторых, развитие корпусной лингвистики за последнее десятилетие позволило перевести изучение языкового материала на принципиально иной уровень. Если еще 10-15 лет назад при описании не отличающихся высокой частотностью явлений исследователь был вынужден опираться во многом на собственную интуицию, то появление крупных (свыше миллиарда словоупотреблений) репрезентативных корпусов, снабженных соответствующим программным обеспечением, расширило эмпирическую базу исследования до немыслимых прежде объемов, ср. [11; 12]. Ведь нетрудно подсчитать, что если бы кто-то захотел просто прочесть вслух составляющие входящий в состав Национального корпуса чешского языка 1,3-миллиардный репрезентативный подкорпус SYN тексты (150 слов в минуту, 8 часов в день без выходных и праздничных дней), ему понадобилось бы для этого **пятьдесят лет!** Компьютер же просмотрит весь этот массив в поисках заданной словоформы, лексемы, грамматической конструкции (а также их комбинации) в течение нескольких минут.

Эти возросшие возможности хорошо иллюстрируются сравнением трех статей на одну и ту же тему (побуждение, выраженное перформативной конструкцией) [7; 8; 19]. Первая из них писалась на материале, собранном в «докорпусную эпоху» - около 10 000 высказываний с актуальной коммуникативной функцией побуждения, извлеченных сплошной выборкой из соотносительных по жанру, объему и времени написания произведений художественной литературы; в составе эксплицитных перформативных конструкций было документировано употребление **6** чешских перформативных глаголов. Привлечение материала появившегося 100-миллионного подкорпуса SYN2000 позволило довести число рассматриваемых чешских перформативных глаголов во второй статье до **30**. Наконец, современный 1,3-миллиардный

объем подкорпуса SYN документирует употребление в составе побудительных высказываний **37** чешских директивных и директивно-комиссивных перформативных глаголов.

И, наконец, **в-третьих**: обращение филолога к корпусу как основному источнику фактического материала в условиях современного развития Интернета и компьютерных технологий не в разы, но в порядки повышает **верифицируемость** используемых им данных. Например, для того, чтобы верифицировать даже какую-нибудь мелочь из кандидатской диссертации автора настоящих строк [4], для которой он когда-то с аспирантским энтузиазмом в течение трех лет собирал фактический материал, еще совсем недавно пришлось бы повторить этот трудоёмкий аналог естественнонаучного эксперимента (выборку примеров по заданному алгоритму из заданного массива текстов). Пользуясь же материалом Национального корпуса чешского языка, швейцарский исследователь М. Гигер не только проверил сделанные нами 20 лет назад выводы о статусе и функциональной нагруженности редкого причастного типа на $-(v)š\acute{i}$, но и существеннейшим образом откорректировал их, рассмотрев те аспекты их функционирования, анализ которых был в докорпусную эпоху немислим [25].

Корпус бесценен для лексиколога и лексикографа, см. [14]. Он позволяет, например, при составлении словника исходить не из интуиции и неких умозрительных соображений, а из реальной употребительности тех или иных лексем в современных текстах. Это же реальное функционирование лексических единиц в речи определяет и выбор вариантов, и порядок следования синонимов, и сочетаемость, см. подготовленные на корпусном материале словари [17; 18].

Одним из мировых лидеров в корпусных исследованиях остается Чехия, поэтому неудивительно, что именно в Праге возникла традиция использовать корпусный материал для анализа не только лексического, но также и грамматического строя языка. На базе Института чешского языка Академии наук Чешской республики в ноябре 2005 года прошла конференция Grammar &

Согрога, в которой приняло участие более 100 исследователей из 14 стран, см. [9]. В сентябре 2007 года здесь же была проведена вторая международная конференция с тем же названием и с не менее солидным числом участников, см. [13], а 28-30 ноября 2012 года – четвертая (третья конференция Grammar & Согрога проходила 22-24 сентября 2009 в Маннгейме). Продуктивность обращения к корпусному материалу практически во всех областях лингвистического исследования убедительно показала конференция, посвященная 100-летию Института Чешского языка, см. [16].

Являясь своего рода огромной энциклопедией, хотя и организованной не парадигматически, а синтагматически, корпус может быть весьма полезен и лингвокультурологу, см. [15].

Важно и то, что использование материала корпуса не мешает оставаться в рамках уже существующих научных школ и направлений, ср. [10; 20]. Речь отнюдь не идет о некоей новой научной парадигме. Так, в представляющих совершенно разные лингвистические школы [22] и [23] обращение к корпусному материалу отнюдь не заставило бы перерабатывать статьи целиком, однако позволило бы дополнить весьма интересный анализ убедительной статистикой.

Таким образом, использование корпусного материала в филологическом исследовании ощутимо сдвигает это исследование по оси «гуманитарное знание» - «естественнонаучное знание» в сторону последнего.

Список литературы

- 1. Аверинцев С.С.** Филология // БСЭ. Т. 27. М., 1977. С. 410-412.
- 2. Голубкова Е. Е.** Использование корпусных данных при изучении словосочетаний устойчивого характера (на материале английского языка) // Вестник МГЛУ. Вып. 572. Изучение разноструктурных языковых знаков: проблемы и перспективы. Москва: ИПК МГЛУ "Рема", 2009. - С. 30-36.
- 3. Изотов А.И.** К вопросу о соотносительности систем причастий в чешском и русском языках // Вестник Моск. ун-та, Серия 9, Филология, 1988. № 4. – С. 103-108.

4. **Изотов А.И.** Система причастных форм в современном чешском литературном языке в сопоставлении с русским (формообразование, семантика, функционирование): Автореф. дис. ... кандидата филол. наук. М., 1991. - 24 с.
5. **Изотов А.И.** Об асимметрическом дуализме формы и содержания полных причастий современного чешского языка // Вестник Моск. ун-та, Серия 9, Филология, 1992. № 6. – С. 34-39.
6. **Изотов А.И.** Система полных причастных форм чешского литературного языка в сопоставлении с русскими // Славяноведение, 1994. № 1. – С. 50-59.
7. **Изотов А. И.** Об иллокутивно специализированном побуждении в современных чешском и русском языках // Язык, сознание, коммуникация: Сб. статей. – М.: МАКС Пресс, 2000. Вып. 15. – С. 62-70.
8. **Изотов А. И.** Чешские перформативные глаголы в побудительном высказывании // Язык, сознание, коммуникация: Сб. статей. – М.: МАКС Пресс, 2001. Вып. 16. – С. 23-31.
9. **Изотов А. И.** Грамматика и корпус: Международная конференция в Праге // Вестник Моск. ун-та, Серия 9, Филология, 2006. № 3. – С. 215-219.
10. **Изотов А.И.** Функционально-семантическая категория императивности в современном чешском языке как прагмалингвистический феномен: Автореф. ... доктора филол. наук. М., 2007. - 40 с.
11. **Изотов А. И.** Опыт корпусного анализа функционально-семантической категории императивности в современном чешском языке // Вестник Моск. ун-та, Серия 9, Филология, 2007. № 3. – С. 47-63.
12. **Изотов А. И.** Чешский национальный корпус и аналитический императив: опыт корпусного анализа малоупотребительных и маргинальных языковых единиц // Вестник Оренбургского гос. ун-та, 2007, № 2. – С. 4-11.
13. **Изотов А. И.** Грамматика и корпус: Очередная международная конференция по корпусной лингвистике в Чехии // Вестник Моск. ун-та. Серия 9. Филология, 2008. № 3. – С. 211-215.

- 14. Изотов А. И.** Новый чешско-русский словарь // Вестник Моск. ун-та. Серия 19. Лингвистика и межкультурная коммуникация. 2008. № 4. – С. 178-184.
- 15. Изотов А. И.** Ключевые явления американской культуры в зеркале национального корпуса чешского языка (англоязычная литература) // Язык, сознание, коммуникация: Сб. статей. – М.: МАКС Пресс, 2010. Вып. 41. – С. 4-20.
- 16. Изотов А. И.** Чешский язык в синхронии и диахронии: столетние корни Института чешского языка // Вестник Моск. ун-та. Серия 9. Филология, 2011. № 5. – С. 212-216.
- 17. Изотов А. И.** Учебный чешско-русский и русско-чешский словарь: около 40 000 слов и выражений. - М.: Филоматис, 2012. - 832 с.
- 18. Изотов А. И.** Новый чешско-русский словарь: около 100 000 слов и выражений. - М.: Дрофа, 2012. - 1024 с.
- 19. Изотов А. И.** Опыт корпусного анализа чешских директивных и директивно-комиссивных перформативных глаголов // Вестник РУДН. Серия Лингвистика, 2012. № 3. - С.99-107.
- 20. Изотов А. И.** Особенности концептуализации авторитарного побуждения в русской и чешской языковых картинах мира // Вестник ЦМО МГУ. Филология. Культурология. Педагогика. Методика. 2012. № 4. - С. 17-22.
- 21. Захаров В.П., Богданова С.Ю.** Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – 161 с.
- 22. Паршина Е.О.** Имена собственные в функции нарицательных как средство оценочной категоризации // Филологические науки. Вопросы теории и практики. 2012. № 7 (18): в 2-х ч. Ч. I. С. 159-162.
- 23. Пацюкова О.А.** Морфемный статус финали -истик(а) в именах существительных // Филологические науки. Вопросы теории и практики. 2013. № 1 (19). С. 142-145.
- 24. Хайдеггер, М.** Время и бытие: Статьи и выступления. - М.: Республика, 1993. - 447 с.

25. Giger M. The Czech active past participle in -(v)ší in contemporary journalist texts // *Korpus - Gramatika - Axiologie. Journal for corpus research and evaluation of language.* 2010/ #2. pp. 3-23. ISSN: 1804-137X

Изотов А.И. Корпусная революция: от «искусства» к «науке» // *Филологические науки: Вопросы теории и практики.* - Тамбов: Грамота, 2013. - №4-1. - 68-71.