

КОРПУСНАЯ РЕВОЛЮЦИЯ И ПЕРЕОЦЕНКА ЯЗЫКОВОЙ  
СИТУАЦИИ В ЧЕХИИ

А. И. Изотов

студентов. Учитывая, что синтаксический аспект начинается уже на втором курсе, осуществимой оказывается лишь задача «научить строить / породить предложение», а серьезному теоретическому материалу и знакомству с основными синтаксическими концепциями в таких условиях места не находится. Вынесение концептуальных синтаксических вопросов возможно в последующий курс «Теоретическая грамматика», где, по понятным причинам, они могут быть представлены лишь крайне обобщенно.

#### Цитируемая литература

- Бояджиев Т., Куцаров И., Пенчев Й. Съвременен български език (фонетика, лексикология, морфология, синтаксис). – София, 1998.  
Грамматика на съвременния български книжовен език. Т. 3: Синтаксис. – София, 1983.  
Пенчев Й. Строеж на българското изречение. – София, 1984.  
Радева П. Записки по синтаксис на съвременния български книжовен език (словосъчетание и просто изречение). – Велико Търново, 1997.  
Съвременни лингвистични теории: Помагало по синтаксис. – Пловдив, 2001.

*А.И. Изотов (Москва)*

## Корпусная революция и переоценка языковой ситуации в Чехии

Развитие компьютерных и интернет-технологий в последние десятилетия обеспечили доступ к электронным корпусам для максимально широкого круга пользователей. По выражению одного из создателей Русского национального корпуса В.А. Плунгяна, «люди, настроенные романтически, даже могли бы сказать, что в лингвистике произошла корпусная революция» [Плунгян 2009].

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»)? Мы уверены, что вполне 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 [Захаров, Богданова 2011], 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 немислимых прежде объемов. Ведь нетрудно подсчитать, что если бы кто-то захотел просто прочитать 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.

И, наконец, **в-третьих**: обращение филолога к корпусу как основному источнику фактического материала в условиях современного развития интернета 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 [Изотов 1991], 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  $-(v)ší$ , но и существеннейшим образом откорректировал их, рассмотрев те аспекты их функционирования, анализ которых был в докорпусную эпоху немислим [Giger 2010].

Корпус бесценен для лексиколога и лексикографа. 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 [Изотов 2012a], [Изотов 2012б].

Являясь своего рода огромной энциклопедией, хотя и организованной не парадигматически, а синтагматически, корпус может быть весьма полезен и лингвокультурологу, см. [Izotov 2010].

Наконец, корпус может быть весьма полезен и в лингводидактике, как при составлении упражнений, так и в самостоятельной работе с корпусом обучаемого.

И если на территории 1/6 части суши «корпусные исследования» для многих остаются лишь модным сочетанием слов (неизбежный вывод после ознакомления с текстами статей в РИНЦ, содержащих слова «корпус», «корпусный» и т.д.), в Чехии основная масса лингвистических исследований проводится с привлечением корпусного материала. Именно на основе корпусных данных строятся новые грамматические описания чешского языка, выходящие в Карловом университете в Праге и в Институте чешского языка национальной Академии наук, см. [Svrgček et al. 2010], [Kapitoly... 2011]. Продуктивность обращения к корпусу практически во всех областях лингвистического исследования убедительно показала конференция, посвященная 100-летию Института чешского языка национальной академии наук, см. [Изотов 2011].

При этом исследователи закономерно обращаются к тем корпусам, которые имеются в наличии, то есть к корпусам письменных и устных чешских текстов (а не к несуществующим корпусам «литературных» и «обиходно-разговорных» текстов), поэтому и на выходе получают описания «письменного чешского» (чеш. *psaná čeština*, англ. *Written Czech*) или «устного чешского» (чеш. *mluvená čeština*, англ. *Spoken Czech*). В результате термин *obecná čeština* (англ. *Common Czech*), в свое время введенный в научный обиход для обозначения речи/языка «обычных», то есть не получивших формального образования чехов и противопоставленный термину *spisovná čeština*, обозначающему речь/язык чехов, такое формальное образование получивших) постепенно вытесняется как из академической, так и из учебной литературы, ср. [Slovník... 2006], [Mluvená... 2011], [Svejnová 2012], что, впрочем, отнюдь не означает его немедленного исчезновения.

#### References

- Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – 161 с.
- 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.
- 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.
- Изотов А.И. Новый чешско-русский словарь: около 100 000 слов и выражений. – М.: Дрофа, 2012. – 1024 с.

- Изотов А.И.* Учебный чешско-русский и русско-чешский словарь: около 40 000 слов и выражений. – М.: Филоматис, 2012. – 832 с.
- Плунгян В.А.* Почему современная лингвистика должна быть лингвистикой корпусов // <http://polit.ru/article/2009/10/23/corpus/> [обращение 29.04.2013]
- Svejnová J.* Česky, prosím II: Učebnice češtiny pro cizince. – Praha: Nakladatelství Karolinum, 2012. – 318 s.
- Cvrček V. et al.* Mluvnice současné češtiny. 1. – Praha: Nakladatelství Karolinum, 2010. – 354 s.
- Izotov A.* American Cultural Literacy Phenomena in the Mirror of Czech National Corpus: Literature, Mythology, Folklore. – Moscow: Azbukovnik, 2010. – 200 p.
- 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. P. 3-23.
- Kapitoly z české gramatiky / Ed. František Štícha.* – Praha: Academia, 2011. – 1190 s.
- Mluvená čeština: hledání funkčního rozpětí / Eds. Čmejrková S., Hoffmannová J.* – Praha: Academia, 2011. – 492 s.
- Slovník nespisovné češtiny. 2. rozš. vyd. / Ed. Hugo J.* – Praha: Maxdorf, 2006. – 460 s.

Изотов А.И. Корпусная революция и переоценка языковой ситуации в Чехии // *Славянские языки и литературы в синхронии и диахронии: Материалы международной научной конференции / Ред. кол. М.Л.Ремнёва и др.* – М.: МАКС Пресс, 2013. – 129-132.