

## **ЯЗЫКОВЫЕ И ИНФОРМАЦИОННЫЕ РЕСУРСЫ: ПРОБЛЕМЫ ИДЕНТИФИКАЦИИ И КЛАССИФИКАЦИИ**

### **Вводные замечания**

В современном мире «создание и эффективное использование информационных ресурсов» рассматривают «как важнейший фактор социального и экономического развития человечества» [2, с. 8]. Очевидной становится насущная необходимость формирования ИКТ-компетенций у профессионалов всех видов деятельности, работающих в различных предметных областях. Филология, разумеется, не составляет исключения, и ИКТ становятся неотъемлемым компонентом в университетской подготовке будущих филологов. Специалистами также подчеркивается важность «информационного подхода» при обучении языкам в XXI веке: «Информационный подход при обучении иностранному языку, в частности РКИ, даёт надежду на ликвидацию главного дефицита современного мировосприятия — дефицита цельного взгляда на мир» [5, с. 145]. Соответственно назрела необходимость уточнения, какие именно информационные ресурсы необходимы филологам, какое место среди них занимают языковые ресурсы, какие виды языковых ресурсов существуют и как они могут применяться в лингводидактических целях. Для этой цели необходимо, прежде всего, сопоставить три понятия: *информационные ресурсы*, *языковые ресурсы*, *лингводидактические ресурсы*.

### **Лингвистические ресурсы. Отличие собственно языковых ресурсов от информационных ресурсов**

Термин *информационные ресурсы* часто рассматривают как эквивалент термина *электронные ресурсы*. Такое видение вполне понятно, но не совсем правильно, ведь информационные ресурсы существуют столько, сколько существуют специально используемые человеком для передачи знаний и / или информации семиотические системы, среди которых особое место по праву принадлежит естественным языкам. Соответственно, особо актуальный сегодня термин *компьютерные ресурсы* (computational resources) и приобретающий популярность термин *электронные ресурсы* (digital resources) номинируют видовые понятия, родовым по отношению к которым является понятие, выражаемое термином *информационные ресур-*

сы. Отметим, что термин *электронные ресурсы*, в свою очередь, обозначает родовое понятие, по отношению к которому понятие компьютерных ресурсов выступает в качестве вида. Таким образом, квазиэквивалентность терминов *информационные ресурсы* и *электронные ресурсы* обусловлена концептуализацией по признаку «новых технологий» и являет собой пример метонимического переноса.

Развитие ИКТ, однако, предоставляя новые возможности в создании различных видов информационных ресурсов, часто идентифицируемых как *новые, современные* и пр., не умаляет роли «традиционных» информационных ресурсов: появление интернет-изданий и автономных электронных изданий, например, лишь расширило «издательский рынок», но не вытеснило традиционные «бумажные» публикации.

Неопределенным, в связи с диверсификацией, является и понятие *ресурсы*. Представляется целесообразным выделение необходимых и достаточных признаков, позволяющих идентифицировать понятие ресурсов в целом: ресурсы — это источник или средство поддержки чего-либо, они готовы к использованию и могут помочь достичь определенной цели. Как и все внешнее окружение человека состоит из естественно существующих фактов и созданных человечеством артефактов, так и ресурсы следует разделять на (1) естественные и (2) искусственные, т.е. специально создаваемые человеком. Очевидно, что все документальные информационные ресурсы относятся ко второй группе. К ней же относятся и все электронные информационные ресурсы, представляющие собой новую форму существования информации, характеризующуюся (особенно если речь идет об интернет-ресурсах) значительно большей доступностью по сравнению с другими видами информационных ресурсов.

Электронные языковые ресурсы еще не фиксируются в качестве особого вида ресурсов. Как правило, их относят к информационным ресурсам, т.е. идентифицируют как вид информационных ресурсов. Важно, что и те, и другие не существуют естественно, а создаются людьми для достижения определенных целей. Тем не менее, между ними существует принципиальное отличие: информационные ресурсы далеко не всегда вербальны, т.к. информация может передаваться в любой материальной форме, понятной целевому адресату, а не только в языковой. Вербальные информационные ресурсы, в свою очередь, могут классифицироваться с точки зрения передаваемой ими информации. Она может быть предметной, или собственно информацией, нашедшей выражение **в языке** (соответствующий вид ресурсов — *информационные ресурсы* (content resources), в дальнейшем мы их будем идентифицировать как *вербализованные ресурсы*); предметной информации противопоставлена информация о **языке** (соответствующий вид ресурсов — *языковые ресурсы*).

Вербализованные (content) и языковые ресурсы (language resources) различаются по своим основным характеристикам: для всех информационных ресурсов, включая вербализованные, — это достоверность (абсолютная характеристика, т.к. информация может быть либо достоверной, либо

недостовойной), а для языковых ресурсов — это аутентичность (относительная характеристика, обуславливается аутентичностью того языкового материала, выбор которого, в свою очередь, обуславливается целью создания языкового ресурса).

Подобно информационным ресурсам, языковые ресурсы существуют в традиционной (словари и грамматики) и в электронной форме (о различных видах языковых ресурсов см., например, [7]). Важно, что грамотно построенные языковые ресурсы всегда отражают естественно существующие ресурсы того языка(ов), для которого(ых) они создаются.

В лингвистике в последнее время получил хождение еще один термин — *лингвистические ресурсы*, который часто используют как синоним термина *языковые ресурсы*. Представляется, однако, что установление эквивалентности между понятиями, обозначаемыми данными терминами, не правомочно, поскольку между ними существуют парциальные отношения. Языковые ресурсы составляют лишь часть лингвистических ресурсов, поскольку к лингвистическим ресурсам относятся и информационные (content) ресурсы, имеющие отношение к предметной области лингвистики, представляющей сегодня весьма разветвленную область знания.

### **Отождествление в Интернете языковых и лингводидактических ресурсов**

Как уже отмечалось выше, языковые ресурсы в Интернете не идентифицируются как особый вид ресурсов. В последнее время, однако, поисковые машины Интернета идентифицируют их как ресурсы, применяемые для обучения языкам, т.е. как лингводидактические ресурсы. Таким образом, здесь мы также наблюдаем пример замещения родового понятия видовым; к тому же, не все лингводидактические ресурсы могут идентифицироваться лишь как языковые ресурсы, а, следовательно, отнесены к лингвистическим ресурсам. Тем не менее, электронные ресурсы лингводидактической направленности весьма важны для обучения языкам. В классификации таких электронных ресурсов на первое место выходит методическая целесообразность. Отсюда в качестве классифицирующих могут выступать, например, следующие признаки: аспект изучения языка (для изучения фонетики / лексики / грамматики / стилистики и др. // комплексные обучающие системы); сфера употребления (подъязык / LSP, коммуникативная сфера, стиль, регистр, вид дискурса и пр.); вид учебной деятельности (собственно обучение / тренировка / контроль // комплексные системы) и др.

Лингводидактические и языковые электронные ресурсы объединяет тот факт, что и те и другие основаны на ресурсах языка. Тем не менее, между ними существуют значительные отличия. Языковые ресурсы, как правило, многоцелевые, т.е. их создатели не ставят перед собой как изначальные исключительно дидактические цели. Как и традиционные словари и грамматики не учебного назначения, электронные языковые ресурсы выполняют, прежде всего, справочные функции. Тем не менее, особенности

языкового материала, на котором они создаются, позволяют классифицировать их и по признаку применимости в лингводидактических целях.

Известно, что самыми «продвинутыми» электронными языковыми ресурсами являются лингвистические (снабженные, помимо структурной разметки, специальными видами метатекстовой и собственно лингвистической разметки) корпуса. В последнее время в корпусной лингвистике, которая занимается проблемами создания корпусов, их типологией и применимостью для различного рода задач, проявляется ярко выраженная тенденция перехода от разработки мегаресурсов в виде национальных (генеральных) корпусов к «малым» корпусам (small corpora), создаваемым с конкретными целями или под конкретные задачи. Среди таких «целевых» корпусов можно выделить корпуса, классификация которых возможна по дидактическим целям. Рассмотрим некоторые такого рода корпуса, нашедшие применение в обучении русскому языку, в том числе русскому языку как иностранному.

### **Специализированные лингвистические корпуса в обучении русскому языку**

Среди корпусов, создаваемых для решения лингводидактических задач, прежде всего, необходимо выделить собственно учебные корпуса, представляющие собой такой вид электронных языковых ресурсов, которые создаются именно в дидактических целях. Как правило, такие корпуса невелики по объему; глубоко аннотированы (иногда нетрадиционно по сравнению с другими лингвистическими корпусами); представляют исключительно норму языка (обучающий корпус в рамках Национального корпуса русского языка — НКРЯ [4]), но могут быть построены и на языковом материале СМИ. Так, например, на аутентичном языковом материале русскоязычной публицистики построен корпус ХАНКО [8] — один из наиболее известных зарубежных учебных корпусов русского языка, созданный в Хельсинкском университете (Финляндия).

Корпусами, получившими наибольшее распространение в преподавании языков как иностранных, являются «корпусы обучающихся» (Learner Corpora), ранее названные нами LC-ресурсами [6]. Такие корпуса иногда также называют «учебными» [3]. Среди LC-ресурсов особая роль принадлежит «пропедевтическим корпусам» (подробнее об LC-ресурсах и пропедевтических корпусах см. [6]), создаваемым с целью предотвращения характерных для определенных категорий учащихся ошибок. Примером такого рода корпусов дидактической направленности может служить RULEC (Russian Language Learner Corpus of Academic Writing) — лингвистический корпус на основе письменных работ студентов РКИ продвинутого уровня, создаваемый на кафедре русского языка и литературы Портландского университета (США, штат Орегон) [1].

### **Информационные и языковые ресурсы с точки зрения системности**

Под информационными ресурсами в общем случае принято понимать «отдельные документы и отдельные массивы документов, документы

и массивы документов в информационных системах, банках данных, других информационных системах» [2, с. 8]. Это определение хорошо тем, что отражает одновременную системность и асистемность информационных ресурсов, которые строго **системны датологически** (для обеспечения возможности поиска) и **асистемны инфологически**. Как результат — они достаточно легко классифицируются по датологическим и поисковым основаниям, но с большим трудом могут быть систематизированы по контенту, за исключением баз и банков данных, формируемых по изначально разработанным и заданным инфологическим моделям, сопряженным в случае банков данных.

В отличие от информационных ресурсов, **языковые ресурсы всегда системны**. Как особый вид информационных ресурсов языковые ресурсы отличает принципиальный **фактологический характер**, т.к. в основе электронных языковых ресурсов всегда лежит языковой материал. Важным фактором для электронных языковых ресурсов является фактор первичности / непервичности, который релевантен как для самого языкового материала, на котором строятся ресурсы, так и для собственно языковых ресурсов. Например, основной корпус НКРЯ — первичный языковой ресурс, а построенный на его основе обучающий корпус — вторичный. При создании электронных языковых ресурсов, которые в самом общем случае могут быть исчислены как разноструктурные лингвистические базы данных, всегда осуществляется предварительное / изначально **дефинирование набора характеристик**, которые задают основу инфологической модели конкретной лингвистической базы данных либо кладутся в основу системы разметки / аннотирования, применяемой в конкретном лингвистическом корпусе. Более того, в отличие от информационных ресурсов, создание электронных языковых ресурсов поддерживается стандартами, в т.ч. разрабатываемыми под эгидой ИСО — Международной организации стандартизации.

#### **«Региональные» лингвистические корпуса как новый вид языковых ресурсов**

Закономерное смещение фокуса с описания нормы на исследование узуса речи результативало в повышенном внимании современной лингвистики к различным формам существования языка и речевым разновидностям. Как следствие, формируется новое ответвление корпусной лингвистики, связанное с созданием такого вида языковых ресурсов, как «региональные корпуса». Именно таким ресурсом является иллюстративный корпус СМИ Гродненщины (РКГ), созданный в рамках российско-белорусского проекта с целью выявления семантической и социокультурной специфики русской речи Гродненщины. Аутентичный языковой материал, положенный в основу РКГ, отражает взаимодействие языков, прежде всего, белорусского и русского, сосуществующих в условиях белорусско-русской лингвокультурной общности. Данные РКГ, соответственно, позволяют исчислять наиболее типичные случаи близкородственной языковой интерференции, специфичную лексику, включая фразеологизмы, а также

выявлять прецедентные для Беларуси и Гродненщины феномены, что важно для обучения иностранцев русскому языку в условиях Республики Беларусь. РКГ вошел в состав нового модуля НКРЯ — корпуса региональной и зарубежной прессы. Таким образом, задавая соответствующие региону подкорпусы, можно видеть реальные расхождения в узусе русской речи, т.е., по сути дела, осуществлять не только внутриязыковое сопоставление, но и выявлять отличия в системе ценностей близких, но не тождественных культур.

Подводя итог вышесказанному, можно сделать **вывод** о том, что языковые ресурсы, технологии их создания и направления имплементации формируют новую область междисциплинарных лингвистических исследований, актуальную не только для собственно исследовательских и / или дидактических целей, но и для обеспечения экономической и информационной составляющих национальной безопасности. Очевидно, таким образом, что именно эта область знаний должна стать приоритетной в подготовке филологов по направлению «компьютерное обеспечение», а корпусные технологии должны быть инкорпорированы в практико-ориентированную парадигму подготовки преподавателей-филологов и, особенно, будущих преподавателей иностранных языков, в т.ч. русского языка как иностранного.

#### ЛИТЕРАТУРА

- 1 Алсуфьева, А.А. Опыт разработки лингвистического корпуса на основе письменных работ студентов РКИ продвинутого уровня // Труды международной конференции «Корпусная лингвистика — 2013». — СПб., 2013. — С. 9–13.
- 2 Антопольский, А.Б. Информационные ресурсы России. — М., 2004.
- 3 Камшилова, О.Н. Учебный корпус текстов : потенциал, состав, структура. — СПб., 2012.
- 4 Национальный корпус русского языка [Электронный ресурс]. — Режим доступа : <http://www.ruscorgo.ru>. — Дата доступа: 17.02.2014.
- 5 Ольхова, Л.Н. Что есть основа обучения русскому языку как иностранному в XXI веке // Сборник материалов научной программы российской экспозиции международной выставки «ЭКСПОЛИНГВА–2007». — М., 2008. — С. 145–149.
- 6 Рычкова, Л.В. Корпусные технологии в исследовании русской речи иностранных студентов // Труды Казанской школы по компьютерной и когнитивной лингвистике TEL-2012. Сер. Интеллект. Язык. Компьютер. Вып 15. — Казань, 2012. — С. 40–47.
- 7 Рычкова, Л.В. Языковые ресурсы: традиции и инновации // V Международная научно-практическая конференция «Прикладная лингвистика в науке и образовании» памяти Р.Г. Пиотровского. — СПб., 2010. — С. 306–311.
- 8 Хельсинкский аннотированный корпус русских текстов ХАНКО [Электронный ресурс]. — Режим доступа: [www.slav.helsinki.fi/hanko](http://www.slav.helsinki.fi/hanko). — Дата доступа : 17.02.2014.