

В. Б. Борщев, Е. В. Рахилина, Т. И. Резникова, Е. Н. Хасина

Структура и особенности системы “Российская лингвистика” (RUSLING)*

Сообщается о конкретной системе, разрабатываемой в Отделе лингвистических исследований ВИНТИ. Это информационный Интернет-портал “Российская лингвистика” (RUSLING). Прототипом системы является знаменитая система LINGUIST List – основной мировой лингвистический ресурс, систему RUSLING можно рассматривать локализацией системы LINGUIST List для российской лингвистики. Обсуждается полезность такой локализации.

Ключевые слова: Информационная система, научное информационное пространство, каталогизация научных ресурсов, Интернет, российская лингвистика.

ВВЕДЕНИЕ

Статья является сообщением о конкретной системе, разрабатываемой в Отделе лингвистических исследований ВИНТИ. Это информационный Интернет-портал “Российская лингвистика” (RUSLING). Пилотный вариант первой очереди этой системы размещен на научном портале ВИНТИ <http://science.viniti.ru/> в разделе Лингвистика и на сайте <http://www.rusling.narod.ru/>.

Статья устроена следующим образом. В первом разделе будет кратко охарактеризован информационный ресурс, послуживший прототипом для нашей системы – международный лингвистический портал LINGUIST List. Основной (второй) раздел статьи посвящен описанию структуры и принципов функционирования создаваемой системы – российского ресурса RUSLING. Наконец, в третьем разделе обсуждаются дальнейшие перспективы и пути совершенствования системы RUSLING.

1. ПРОТОТИП

Прототипом нашей системы является знаменитая система LINGUIST List (<http://www.linguistlist.org/>) – основой мировой информационный ресурс по лингвистике, которым пользуются лингвисты всего мира. Система эта подробно описана в работе [2] и, конечно, на ее сайте. Поэтому здесь мы только назовем ее основные черты.

Метасистема. LINGUIST List пытается описать как можно полнее информационное пространство лингвистики как научной области, нарисовать что-то вроде “информационной карты” лингвистики и снабдить эту карту средствами навигации по ней. Поэтому он является (в основном) метасистемой. Основные информационные ресурсы разрабатываются и существуют, как правило, вне системы. Система, так сказать, каталогизирует и аннотирует их в естественной форме и обеспечивает доступ к ним, являясь как бы структурированным каталогом объектов этого информационного пространства (например, персоналий, т. е. лю-

дей, работающих в лингвистике, а также организаций, конференций, публикаций, языковых ресурсов и т. д.).

Принцип самообслуживания. Чтобы добиться “полноты охвата” этого информационного пространства, авторы системы использовали принцип самообслуживания. Они тщательнейшим образом разработали структуру системы и ее логику, естественную для ее пользователей – лингвистов, реализовали ее, в основном, в виде совокупностей баз данных и предложили создателям информационных ресурсов самим заполнять эти базы данных сведениями об этих ресурсах и, при необходимости, корректировать записанную там информацию. Иными словами, лингвисты могут внести информацию о себе и своих научных интересах, а также о своих публикациях, диссертациях и т. п., организаторы конференций – об организуемых ими конференциях, разработчики – о создаваемых ими лингвистических ресурсах, проектах, программных средствах и т. д. и т. п. Внесенная информация проходит проверку у администраторов сайта (чтобы не допустить распространение спама и других не имеющих отношения к лингвистике данных), а затем помещается в соответствующие базы данных и тем самым становится доступной для пользователей.

Комфортные сервисные средства. Пользователям системы, как тем, кто заполняет ее, так и тем, кто “потребляет” записанную там информацию, был предоставлен набор удобных сервисных средств, делающих работу с системой по-настоящему комфортной.

2. ИНТЕРНЕТ-ПОРТАЛ

“РОССИЙСКАЯ ЛИНГВИСТИКА” (RUSLING)

2.1. Зачем создавать российский аналог LINGUIST List'a?

Как уже говорилось, портал RUSLING разрабатывался по образцу портала LINGUIST List. Возникает естественный вопрос – а зачем создавать российский аналог LINGUIST List'a? Ведь

*Статья является изложением доклада (точнее, основной его части), сделанного на 7-й Международной конференции “НТИ-2007” [1].

LINGUIST List, этот действительно замечательный портал, открыт для всех. Пополнение его происходит, как мы уже говорили, методом “самообслуживания” (posting с авторизацией): каждый пользователь может добавить свою запись в соответствующую рубрику. Т. е. в принципе российские лингвисты, внося в LINGUIST List соответствующие дополнения, могут отразить там и инфраструктуру российской лингвистики.

Однако, как показывает реальная практика, российская лингвистика пока очень слабо представлена в LINGUIST List'e. Скажем, в списке из 18 с лишком тысяч лингвистов присутствует всего около 60 лингвистов из России. В базах данных системы плохо представлены российские университеты и академические институты, не значатся основные журналы. В соответствующие списки внесена ничтожная доля российских диссертаций (порядка 10), там почти нет российских проектов, издательств и т. п. Конечно, виноват в этом не LINGUIST List: система открыта для всех профессионалов, и это их задача предоставлять туда информацию о различных составляющих российской научной лингвистической жизни.

В чем же причины? Наверное, их несколько, и главная — историческая. В советское время советская лингвистика в каком-то смысле не была частью мировой лингвистики, как и вся советская наука — частью мировой науки. Международные контакты были затруднены, поездки на зарубежные конференции были редкостью, публиковаться можно было только в советских журналах. Все это хорошо известно.

Железного занавеса уже давно нет, и многое, конечно, изменилось. Российские лингвисты намного больше, чем раньше, ездят на международные конференции, публикуются на Западе, международные конференции проводятся и в России. Выпускники российских университетов начали поступать в аспирантуру в американские и европейские университеты.

Но во многом российская лингвистика так и остается “отдельным континентом”. Российские лингвисты до сих пор публикуются, в основном, по-русски и в российских изданиях. Трудно сказать, какая доля российских лингвистических статей дается по-английски, но очевидно, что очень небольшая (хотя это явно не худшие сочинения). Более того, работы, публикуемые по-английски, как правило, издаются и по-русски, часто раньше их английских версий. Наверное, это естественно, особенно для исследований по русскому языку. И, скорее всего, это будет продолжаться довольно долго. Не только публикации, но и вся российская лингвистическая жизнь — проекты, преподавание и (в меньшей степени) конференции — будет происходить в основном на русском языке (хотя, наверное, “англоязычная составляющая” российской лингвистической жизни будет все же постоянно увеличиваться).

Подведем итог. Основная идея системы LINGUIST List — это *глобализация* информационного пространства лингвистики. Это естественная идея, но она, по крайней мере пока, плохо реализуема в отношении российской лингвистики. Российские лингвисты пока не очень активно

взаимодействуют с международными системами, российская лингвистика остается в значительной мере русскоцентричной. Идея нашего портала — это *локализация*¹ системы LINGUIST List, точнее — создание ее русского аналога (конечно, не *вместо* системы LINGUIST List, а *в дополнение* к ней). Так же, как и LINGUIST List, наш портал строится как метасистема, его цель — предоставить российским лингвистам карту существующих в Интернете информационных ресурсов по лингвистике, прежде всего российских и русскоязычных, дополнив её необходимой справочной информацией. При этом в дальнейшем можно думать и об англоязычной версии портала, ориентированной на “внешнего пользователя”, интересующегося российской лингвистикой.

Замечание

Обсуждая идеи локализации, интересно рассмотреть российские Интернет-системы, существующие в других областях. Так в математике существуют по крайней мере две российские информационные Интернет-системы, Math-Net.Ru (<http://www.mathnet.ru/>) и Math Tree (<http://www.mathtrce.ru>).

Math-Net.Ru. Как сказано на сайте системы, “информационная система Math-Net.Ru — это общероссийский математический портал, предоставляющий российским и зарубежным математикам различные возможности в поиске информации о математической жизни в России”. Система создана в Математическом институте им. В. А. Стеклова РАН совместно с Отделением математических наук РАН. Подробное описание системы приведено в работе [3].

Math Tree. Как сказано на сайте системы, “Проект посвящен созданию каталога математических Интернет-ресурсов. Цель проекта систематизация математических знаний, в частности, Интернет-ресурсов.

Инициаторами проекта являются институты Сибирского отделения РАН”.

Подробное описание системы приведено в работе [4].

2.2. Структура системы RUSLING

Набор рубрик нашего портала аналогичен набору основных рубрик LINGUIST List'a и включает следующие категории: **Новости, Персоналии, Организации, Журналы, Научные исследования и публикации, Лингвистические ресурсы, Обучение и преподавание, Ссылки** (см. рис. 1 и, конечно, сайт портала).

Состав рубрик будет со временем расширен. Каждая запись в каждой рубрике представляет собой имя (название) соответствующего “объекта” (лингвистика, организация, журнала и т. п.), содержащее ссылку на его сайт (если таковой имеется) и в некоторых случаях снабженное краткой справкой об этом объекте. Представляется, что такого рода портал будет полезен российским лингвистам, в частности, лингвистам, работающим в провинциальных университетах и институтах.

¹Термин этот скорее из бизнеса, когда крупные интернациональные компании “локализируют” свою продукцию и подходы на местном рынке.

РОССИЙСКАЯ ЛИНГВИСТИКА (RUSLING)	
<ul style="list-style-type: none"> • Новости • Организации • Научные исследования и публикации • Журналы • Персоналии • Лингвистические ресурсы • Обучение и преподавание • Ссылки • Научный сайт ВИНТИ • Для сообщений 	<p>Задача раздела — предоставить лингвистам “информационную карту” области: где можно получить ту или иную полезную информацию. В качестве прототипа выбран известный сайт <i>LINGUIST List</i> (http://www.linguistlist.org/). Сайт этот существует уже около 15 лет и пользуется заслуженной популярностью среди лингвистов-профессионалов. Структура его тщательно продумана и, главное, за много лет “обкатана” в работе.</p> <p>Работа над разделом только начинается, и разработки будут благодарны за любые замечания и пожелания.</p> <p><i>Раздел создается в Отделении лингвистических исследований ВИНТИ РАН.</i></p> <p><i>Разработчики:</i> <i>Контакт — Е. В. Рагилина, Т. И. Резникова, В. Б. Борцев</i> <i>Web-мастер — Е. Н. Хасина</i> <i>Координатор раздела — В. Б. Борцев</i></p>
<p>На главную страницу Новости Организации Исследования и публикации Журналы Персоналии Ресурсы Обучение Ссылки </p>	

Рис. 1. Рубрики портала RUSLING

<p>Омск <i>Омский государственный университет</i> Филологический факультет <u>Факультет иностранных языков</u></p> <p><i>Омский государственный педагогический университет</i> Филологический факультет</p> <p>Орел <i>Орловский государственный университет</i> Филологический факультет <u>Факультет иностранных языков</u></p>
--

Рис. 2. Высшие учебные заведения и факультеты филологического профиля в России и СНГ (фрагмент списка)

В настоящее время в той или иной мере заполнены рубрики **Организации**, **Журналы**, **Лингвистические ресурсы** и **Персоналии**. Ниже мы кратко остановимся на принципах организации и наполнении каждого из этих разделов.

Рубрика **Организации** включает в себя пять подразделов:

- *Высшие учебные заведения и факультеты филологического профиля в России и СНГ;*
- *Зарубежные кафедры и институты славистики и русистики;*
- *Научно-исследовательские институты*
- *Научные общества*
- *Центры по подготовке специалистов по теоретической и прикладной лингвистике*

Перечень вузов и факультетов филологического профиля дает достаточно полную картину высшего филологического образования в России: в списке, организованном по названиям городов (в алфавитном порядке) представлены расположенные в этих

городах вузы и даны ссылки на сайты филологических и близких к ним по профилю факультетов соответствующих вузов (всего около 90 ссылок), ср., например, фрагмент списка (см. рис. 2, подчеркнутые строки соответствуют гиперссылкам, нажав на которые, пользователь попадает непосредственно на сайт выбранного учреждения/факультета):

Аналогичным образом организован подраздел *Зарубежные кафедры и институты славистики и русистики*. Здесь ссылки (всего их свыше 150) сгруппированы на верхнем уровне по странам, а внутри страны – по названиям городов (т. е. пользователь может вначале выбрать интересующую его страну из алфавитного списка стран и затем просмотреть отсортированный по городам список университетов, в которых есть соответствующие кафедры, и по ссылке перейти на сайт выбранной кафедры), ср. фрагмент списка славистических отделений университетов Германии (рис. 3):

Universität zu Köln — Slavisches Institut
 Universität Konstanz — Slavistik Fachgruppe Sprachwissenschaft
 Universität Leipzig — Institut für Slavistik
 Otto-von-Guericke-Universität Magdeburg — Institut für fremdsprachliche Philologien — Slavistik
 Johannes Gutenberg-Universität Mainz — Institut für Slavistik
 Universität Mannheim — Slavisches Seminar
 Philipps-Universität Marburg — Institut für Slawische Philologie

Рис. 3. Зарубежные кафедры и институты славистики и русистики (фрагмент списка)

РОССИЙСКАЯ ЛИНГВИСТИКА (RUSLING)		
<ul style="list-style-type: none"> • <u>Новости</u> • <u>Организации</u> • <u>Научные исследования и публикации</u> • <u>Журналы</u> • <u>Персоналии</u> • <u>Лингвистические ресурсы</u> • <u>Обучение и преподавание</u> • <u>Ссылки</u> • <u>Научный сайт ВИНТИ</u> • <u>Для сообщений</u> 	Лингвистические ресурсы в Интернете КОРПУСА <u>Славянские языки</u> <u>Другие языки</u> СЛОВАРИ <u>Одноязычные словари</u> <u>Двуязычные и многоязычные словари</u> ДРУГИЕ РЕСУРСЫ <u>Информационные сайты и рассылки</u> <u>Архивы</u> <u>Блоги</u> <u>Отдельные проекты</u>	
	На главную страницу <u>Новости</u> <u>Организации</u> <u>Исследования и публикации</u> <u>Журналы</u> <u>Персоналии</u> <u>Ресурсы</u> <u>Обучение</u> <u>Ссылки</u>	

Рис. 4. Рубрика Лингвистические ресурсы

Подраздел *Научно-исследовательские институты* ориентирован как на специализированные лингвистические научные учреждения (как, например, Институт языкознания РАН), так и на институты, в которых лингвистика является одной из областей научного исследования наряду с различными другими (например, ВИНТИ РАН). Следующий подраздел — *Научные общества* — пока не заполнен. Наконец, последняя подрубрика — *Центры по подготовке специалистов по теоретической и прикладной лингвистике* — ориентирована на особый вид высшего филологического образования, наиболее тесно связанный с лингвистической научной деятельностью. На сегодняшний день здесь имеются ссылки к московским центрам данного профиля.

В рубрике *Журналы* отражены основные российские и многие западные периодические лингвистические издания. Ссылка на российский журнал снабжена краткой справкой, дающей сведения об истории издания (год первого выпуска), его тематике, выпускающей организации, периодичности и доступности его материалов в электронном виде. Ссылка на западный журнал включает информацию о выпускающем его издательстве. Оба списка организованы по алфавиту.

Рубрика *Лингвистические ресурсы* включает подразделы *Корпуса*, *Словари* и *Другие ресурсы* (см. рис. 4).

В подразделе *Корпуса* собраны ссылки на коллекции текстов, предназначенные для лингвистических исследований. Ссылки сгруппированы по языкам, каждая ссылка снабжена краткой справкой, дающей сведения о разработчиках корпуса, его объеме, типе составляющих его текстов и видах разметки, применявшейся к этим текстам (см. фрагмент списка корпусов русского языка, рис. 5). В настоящий момент наиболее полно в системе отражены корпуса славянских языков (всего — 29 сылок), однако представлены и некоторые корпуса других языков.

В подразделе *Словари* содержатся ссылки на одноязычные (толковые, словари синонимов) и двуязычные/многоязычные (переводные) электронные словари различных языков. Ссылки снабжены краткой информацией об объеме словаря и типе доступа к нему (т. е. свободный или осуществляющийся на платной основе).

В числе *Других ресурсов* в системе собраны ссылки на некоторые информационные сайты и электронные рассылки, архивы лингвистических статей, основные лингвистические блоги — как российские, так и англоязычные, а также на отдельные лингвистические проекты, представленные сайтами в Интернете. Каждая ссылка снабжена краткой информационной справкой о ресурсе.

РОССИЙСКАЯ ЛИНГВИСТИКА (RUSLING)	
<ul style="list-style-type: none"> • Новости • Организации • Научные исследования и публикации • Журналы • Персоналии • Лингвистические ресурсы • Обучение и преподавание • Ссылки • Научный сайт ВИНИТИ 	<p>КОРПУСА, СЛАВЯНСКИЕ ЯЗЫКИ</p> <p>Русский язык</p> <p><u>Национальный корпус русского языка</u></p> <p>Корпус был подготовлен в рамках программы РАН “Филология и информатика” специалистами из Института русского языка РАН, ВИНИТИ РАН, Института языкознания РАН при участии других институтов и университетов Москвы и Санкт-Петербурга.</p> <p>Планируемый объем корпуса 200 млн. слов. Корпус включает метатекстовую, морфологическую, семантическую и акцентную разметку. В настоящее время на сайте размещен корпус современных текстов (второй половины XX — начала XXI века) объемом свыше 100 млн. слов; размещение остальных текстов планируется в ходе дальнейшей работы над Корпусом.</p> <p><u>Ханко-Хельсинкский аннотированный корпус</u></p> <p>Планируемый объем -- 100 тыс. слов (из журнала “Итоги”)</p> <p>Проект по созданию Хельсинкского аннотированного корпуса русских текстов ХАНКО осуществляется на Отделении славянских и балтийских языков и литератур Хельсинкского университета. Предполагается, что корпус будет содержать морфологическую, синтаксическую и функциональную информацию о текстах общим объемом около 100 тыс. текстоформ, извлеченных из журнала “Итоги”. Проект осуществляется под руководством профессора Арто Мустайоки. Одним из основных принципов создания корпуса является направленность на максимальный охват грамматической информации, а не на объем материала.</p>

Рис. 5. Подрубрика Корпуса. Славянские языки (фрагмент)

РОССИЙСКАЯ ЛИНГВИСТИКА (RUSLING)	
<ul style="list-style-type: none"> • Новости • Организации • Научные исследования и публикации • Журналы • Персоналии • Лингвистические ресурсы • Обучение и преподавание • Ссылки • Научный сайт ВИНИТИ 	<p>ПЕРСОНАЛИИ. ПОИСК ПО ОБЛАСТИ НАУЧНЫХ ИНТЕРЕСОВ</p> <p>диалектология</p> <p>дискурс</p> <p>история лингвистики</p> <p>количественная лингвистика</p> <p>когнитивная лингвистика</p> <p>компаративистика</p> <p>компьютерная лингвистика</p> <p>корпусная лингвистика</p> <p>лексикография</p> <p>морфология</p> <p>ономастика</p> <p>полевая лингвистика</p> <p>прагматика</p> <p>просодия, акцентология</p> <p>психолингвистика</p> <p>семантика</p> <p>семиотика</p> <p>синтаксис</p> <p>словообразование</p> <p>социоллингвистика</p> <p>терминоведение</p> <p>типология</p> <p>философия языка</p> <p>фонетика и фонология</p> <p>этнолингвистика</p> <p>язык художественной литературы</p>

Рис. 6. Рубрика Персоналии. Поиск по области научных интересов

Апресян Юрий Дереникович

Основное место работы: **Институт русского языка РАН**

Должность: **Зав. отделом, академик РАН**

homepage: <http://cl.iitp.ru/Personal/Apresjan/index.html>

Области научных интересов: **семантика, синтаксис, теоретическая лексикография, создание словарей, формальные модели языка, компьютерная лингвистика, машинный перевод.**

Кандидатская диссертация: **“Фразеологические синонимы в современном английском языке” (1958)**

Докторская диссертация: **“Лексическая семантика: синонимические средства языка” (1984)**

Педагогическая деятельность: **Читал курсы в МГУ, Институте иностранных языков, университете Канберры (Австралия), UCLA (США), в университетах Австралии и Германии.**

Проекты: **Новый объяснительный словарь синонимов русского языка**

Рис. 7. Пример справочной анкеты

По замыслу рубрики **Персоналии**, в ней в идеале должны быть представлены все работающие в России лингвисты-профессионалы. Предусмотрен просмотр списков по алфавиту (русскому и латинскому), а также создание выборок по двум параметрам — *область научных интересов* и *язык* (тот или иной язык, изучением которого занимается данный лингвист). Тем самым пользователь, интересующийся, скажем, определенной сферой лингвистических исследований, может выбрать данный раздел из алфавитного списка (см. рис. 6) и получить список всех занесенных в систему исследователей, занимающихся этой проблематикой.

Каждый элемент в получаемом пользователем списке лингвистов (алфавитном или сформированном по заданному поисковому параметру) представляет собой гиперссылку: выбрав определенную фамилию, пользователь получает краткую анкету с информацией о научной деятельности данного человека (см. рис. 7):

Пока рубрика заполнена очень слабо, в ней всего около сотни имен, причем для многих из них справочная информация отсутствует.

Не заполнены пока и многие другие рубрики (**Научные исследования и публикации, Обучение и преподавание** и др.).

6. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Повторим, что пока наш портал — это только пилотный проект, очень простая система, не сопоставимая с LINGUIST List'ом ни по объему, ни по функциональным возможностям.

Чтобы превратить его в систему, действительно обслуживающую российское лингвистическое сообщество, необходима большая работа. Желательно прежде всего реализовать “принцип самообслуживания”, позволяющий пользователям самим

вносить в рубрики системы записи о существующих и создаваемых “информационных объектах” (конференциях, журналах, корпусах, проектах и т. п.).

Для этого нужно тщательно разработать структуру баз данных, реализующих рубрики портала (аналогичных базам данных LINGUIST List'a), программы и технологию пополнения этих баз данных, позволяющую пользователям комфортные средства работы с этими базами данных. Но реализация описанного метода ведения портала будет возможной только тогда, когда он завоеует признание в сообществе российских лингвистов.

Одним из будущих путей развития системы RUSLING может стать ее объединение с другим российским порталом — “Лингвистика в России: ресурсы для исследователей” (<http://uisrussia.msu.ru/linguist>), возможности которого обсуждаются в настоящее время.

СПИСОК ЛИТЕРАТУРЫ

1. Борщев В. Б., Резникова Т. И., Хасина Е. Н. Интернет-портал “Российская лингвистика” (RUSLING): Тезисы доклада на 7-й Международ. конф. “НТИ-2007”. — М.: ВНИТИ, 2007. — С. 64.
2. Борщев В. Б. Система LINGUIST List — основной мировой лингвистический ресурс // НТИ. Сер. 2. — 2008. — № 6. — С. 22-36.
3. Жижченко А. Б., Изаак А. Д. Информационная система Math-Net.Ru. Применение современных технологий в научной работе математика // УМН. — 2007. — Т. 62, вып. 5 (377). — С. 107-132.
4. Ершов Ю. Л., Клименко О. А., Матвеева И. И., Пикалов В. В. Математическая информационная система MathTree // УМН. — 2007. — Т. 62, вып. 5 (377). — С. 133-142.

Материал поступил в редакцию 26.05.08.