

Оксана Остапчук (Москва, Россия)

Эквивалентность и парадигматические отношения между ключевыми словами в многоязычной информационно-поисковой системе iSybislaw

Ключевые слова: информационно-поисковый язык, ключевые слова, iSybislaw, эквивалентность, функциональная синонимия

Key-words: information retrieval (IR) query language, key-words, iSybislaw, equivalency, functional synonymy

Многоязычный характер информационно-поисковой системы славянского языкознания iSybislaw выдвигает особые требования к языку ключевых слов, являющимся разновидностью информационно-поискового языка¹. Будучи синонимичными терминам и терминологическим сочетаниям из разных славянских языков и отобранные с учетом традиций национальных терминологий ключевые, в первую очередь, призваны быть эффективным средством информационного поиска. При этом требования формальной межъязыковой эквивалентности терминов уступают место соображениям целесообразности и функциональной значимости ключевых слов, связанных с конкретными текстами – документами в поисковой системе.

Информационно-поисковая система iSybislaw уже давно успела зарекомендовать себя как важный и эффективный инструмент научного поиска, адресованный прежде всего профессионалам, специалистам в области славянского языкознания, который позволяет искать информацию в четко определенном лингвистическом сегменте, не разделяя его на национальные языки/школы, что несомненно, способствует научному диалогу в международной славистике. Цифровая библиография, используемая в научных исследованиях, предоставляет также широкие возможности для ее более прикладного применения, в частности, в процессе обучения будущих специалистов-филологов. Материалом для нашего доклада послужило пилотное исследование, проведенное в фокус-группе российских студентов старших курсов (4 курса бакалавриата и 1-2 курса магистратуры), специализирующихся в области русистики и славистики. Были обработаны все запросы, сформулированные студентами, а также полученные результаты, релевантной считалась такая смена ключевого слова или их комбинации, которая давала другой результат поиска.

Участники фокус-группы явно отдавали предпочтение ключевым словам, способным определить широкую предметную область текста (*языкознание*,

¹ Понятие языка ключевых слов активно развивается коллективом, возглавляемым Э. Рудник-Карват с 1999 г. см., в частности: (Rudnik-Karwatowa, Mikos, 2007).

диалектология, история языка)², тем самым в общих чертах ограничивая сегмент поиска. Особую роль играли здесь лингвонимы. Единичное ключевое слово рус. язык (и эквивалентные ему пол. *język*, укр. *мова*) обладает крайне низкой информативностью³, иначе обстоит дело с сочетаниями, где оно является опорным: ключевые слова *русский язык, украинский язык, церковнославянский язык* и т.п. использовались абсолютно всеми участниками фокусной группы как такие, которые способны отразить тематическую область текста, прояснить характер используемого материала. В то же время осознание коммуникативной недостаточности лингвонимов, использованных в качестве ключевых слов, приводило к параллельному использованию нескольких сочетаний: *русский / польский / украинский язык* или парами *русский / украинский язык, русский / польский язык* (заметим, что в сочетании с одним и тем же предметным классификатором это давало разные результаты поиска). Стремление приблизиться к содержанию документа (в частности, при желании выявить тексты, в которых проводится сопоставительный анализ) вызвало также использование дополнительного компонента поискового сочетания, прямо указывающего на сопоставительный характер искомой работы (*в сопоставлении с русским, польским и пр.*).

При поиске необходимой лингвистической информации участниками фокусной группы предпочтение по очевидным причинам отдавалось русскому языку как родному, однако знание других современных славянских языков (польского, болгарского, украинского, сербского) обеспечило возможность обращения пользователей к ключевым словам, оформленным на других языках. Для поиска, осуществляемого с использованием ключевых слов одновременно на нескольких языках, особое значение приобретает их эквивалентность, в условиях информационно-поискового языка тесно связанная с проблемой установления функционального соответствия. Для целого ряда ключевых слов, прежде всего входящих в ядро соответствующей семантической подгруппы поискового языка, пользователи обращались к функциональным эквивалентам во всех интересующих их языках: *история языка – історія мови – historia języka; фонетика – фонетика – fonetyka; историческая фонетика – історична фонетика – fonetyka historyczna*.

Однако для ключевых слов, занимающих место на периферии поискового языка (в различных его национальных сегментах), пользователи сталкивались с отсутствием функциональных эквивалентов для них. Это хорошо заметно на примере такой узко специальной сферы, как изучение церковнославянского языка и древней (славянской)

² О функциях ключевых слов см., в частности (Москвитина, 2009, с. 282).

³ Язык – малоинформативное ключевое слово, термин, «не отражающий специфику статьи, поскольку может использоваться в разных областях знания» (Камшилова, 2013, с. 113).

книжности. Для поиска документов по данной проблематике наиболее часто использовались следующие (русские) ключевые слова: *книжность*, *книжная традиция*, а также *книжно-письменная речь* / *книжный язык*. Заметим, что названные ключевые слова, не всегда являясь терминами в узком значении слова (как *книжная традиция*) не только часто используются в заглавии документов, но и составляют важную часть русского научного дискурса, описывающего языковые особенности памятников письменности вплоть до XVIII в. При этом перечисленные ключевые слова, как правило, имеющие статус дескриптора, оказываются изолированными в поисковой системе, поскольку для них отсутствуют параллели в сегментах, оформленных на других славянских языках. Формально пол. *piśmiennictwo* и укр. *писемність* являются функциональными синонимами рус. *книжность*, однако в базе документов с такими дескрипторами нет. Сложность поиска функционального соответствия для таких ключевых слов, как *книжно-письменная речь* или *книжный язык* связана также с тем, что само их вычленение в русской научной традиции не находит параллелей в других национальных школах, так, пол. *styl artystyczny* явно не покрывается с ним ни по объему, ни по функционалу значения. Аналогичным образом обстоит дело также с элементами поискового языка, находящимися на его периферии в силу относительно недавнего их появления, как например, ключевые слова, связанные с описанием функционирования языка в ряде сфер, ограниченных специфическими условиями коммуникации. Так, рус. *религиозный язык* соответствует укр. *релігійний стиль*, в польском же сегменте поискового языка параллельно возможно использование в качестве функциональных синонимов ключевых слов *język religii*, *język religijny*, а также *styl biblijny*, которые лишь отчасти могут быть признаны функциональными соответствиями для перечисленных выше ключевых слов.

Общепризнанным является тот факт, что специфика языка ключевых слов состоит в принципиальной однозначности его элементов, а в идеале также в их семантической «изолированности». Это одно из основных отличий ключевых слов на фоне единиц естественного языка, связанных с другими языковыми элементами разнообразными парадигматическими отношениями (синонимии, антонимии, паронимии). Однако это утверждение справедливо лишь применительно к отдельным разрядам ключевых слов, например, соотносимых с однозначными терминами (типа *грамматика* или *фонема*), а также с прецедентными именами и персоналиями (*Максим Грек*, *Библия*, *Matka Boska*). Однако, как оказалось, даже ключевые слова, соотносимые с отдельными географическими названиями, в разных национальных подсистемах поискового языка могут оказаться функционально не эквивалентными в силу различия семантико-

коннотативных компонентов значения, как, например, *Буковина* в русском, польском и украинском сегментах поиска.

Как показывает опыт работы пользователей из нашей фокус-группы в поисковой системе ISYBISLAW, между ключевыми словами также могут устанавливаться парадигматические отношения, обусловленные близостью функции элементов поискового языка. Весьма распространенными здесь оказываются гипер-гипонимические отношения, предполагающие функциональную связь ключевых слов по линии «часть-целое», что, собственно, и обуславливает их использование в одном запросе в качестве вариантов: ср. *растения – цветок, церковно-обрядовая терминология – церковно-религиозный термин*, в том числе ситуативные: *диалектология – говоры, фонетика – звуки*. Весьма прозрачными оказываются также отношения функционального сходства, основанные на логической синонимии, причем степень близости таких функциональных синонимов может варьироваться от фактических дублетов (ср. *фитонимы – названия растений*) до синонимического ряда, члены которого связаны лишь ассоциативно (Sacharnyj 1989), будучи включенными в общий тематический круг благодаря наличию единого стержневого слова (*język familijny – język rodzinny – język w rodzinie – język w domu*).

Как показывает наш материал, участники фокусной группы, исходя из собственной языковой компетенции (Sadowska, 2010), в ходе поиска активно использовали комбинации из различных типов ключевых слов, стремясь максимально приблизиться к семантическому ядру искомого текста. При этом различный порядок ключевых слов, а также варианты их сочетаний давали разные результаты поиска: *говоры + русский* и *диалектология + русский*, *историческая фонетика + русский* и *история + фонетика + русский*. Сама возможность варьирования ключевых слов, использование логических синонимов, по мнению пользователей, является важным преимуществом системы iSybislaw и позволяет расширить и нюансировать результаты поиска, задавая необходимую глубину поиска. При этом между ключевыми словами, используемыми в вариативных поисковых цепочках, также возникают отношения функциональной синонимии.

Перспективы использования информационной базы как современного средства научного предметно ориентированного поиска связаны, в частности, с развитием таких важных ее свойств, как интертекстуальность, что позволило бы обеспечить доступ пользователям из разных стран не только к библиографическим данным, но также к полнотекстовым файлам. Это направление развития представляется особенно важным в условиях низкой доступности современной литературы по славянскому языкознанию за пределами страны, в которой издается конкретная научная работа. В то же время этот

вектор развития базы ставит ее создателей перед новыми вызовами: необходимостью оптимизации функционирования системы в условиях неизбежного усиления информационного шума. Как показало наше пилотное исследование, для нынешних студентов, хорошо ориентирующихся в современной информационной среде, особое значение имеет целенаправленность и предметная ориентированность поиска, заложенная в системе iSybislaw. В то же время важной представляется интеграция сведений, содержащихся в системе, в общий поток научной информации, соотнесение их с другими доступными внешними источниками такой информации (в том числе связанными с библиотечными собраниями и электронными ресурсами важнейших национальных библиотек, таких, как Российская государственная библиотека в Москве или Национальная библиотека (Biblioteka Narodowa) в Варшаве).

Камшилова, О. Н. (2013). *Малые формы научного текста: ключевые слова и аннотация (информационный аспект)*, *Известия Российского государственного педагогического университета им. А.И. Герцена*, № 156, сс. 106-117. Электронный ресурс: http://lib.herzen.spb.ru/text/kamshilova_156_106_117.pdf, дата обращения – 15.04.2016.

Москвитина, Т.Н. (2009). *Ключевые слова и их функции в научном тексте*, *Вестник ЧПГУ. Филология и искусствоведение*, 11, сс. 277–283. Электронный ресурс: <http://cyberleninka.ru/article/n/klyuchevye-slova-i-ih-funktsii-v-nauchnom-tekste>, дата обращения – 09.04.2016.

Rudnik-Karwatowa, Z., Mikos Z. (2007) *Nowoczesny system informacji slawistycznej: zagadnienia, dotychczasowe wyniki i perspektywy*, “Zagadnienia Informacji Naukowej”, nr 2 (90), s. 19-40.

Sadowska J. (2010). *Kompetencja językowa twórców i użytkowników języków informacyjno-wyszukiwawczych*. [W:] *Informacja i język*.

Sacharnyj L.V. (1989). *Struktury kojarzeniowe w języku naturalnym i języku informacyjno-wyszukiwawczym (na przykładzie języka haseł przedmiotowych)*, “Zagadnienia informacji naukowej”, nr 1 (54), s. 75-102.