

ФИЛОЛОГИЯ

Д.А. Морель Морель

Поисковая оболочка электронного словаря как инструмент лингвистического исследования

Данные лексикографических источников весьма широко используются отечественными (А.А. Поликарпов [1, с. 62–75], З.В. Беркетова [2, с. 69–77]) и зарубежными (J. Fargues [3, p. 112–117], Ph. Trigano [4]) лингвистами как основной материал исследования. Это становится возможным в силу того, что словари и словарные дефиниции являются источниками, с достаточной степенью адекватности и объективности отражающими содержательную сторону языковых единиц (P. Imbs [5, p. XI–XLVIII], Fargues [3], P. Swanepoel [6, p. 160–190]). Будучи результатом методологически обоснованного отвлечения от множества употреблений наименования в реальных высказываниях, словарная дефиниция аналитически представляет и фиксирует облигаторный для всех представителей языкового коллектива инвариант лексического значения [общезыковое (словарное) значение], обеспечивая реализацию коммуникативной функции языка (О.А. Михайлова [7] Поликарпов [1, с. 62–75]).

Таким образом, значимость словарей при проведении лингвистических исследований не должна подвергаться сомнению.

В силу некоторых особенностей (Д.А. Морель Морель [8, с. 194–207]) электронные словари обладают рядом преимуществ перед традиционными бумажными. Основным положительным моментом, с точки зрения исследователя-лингвиста,

является высокая степень эргономичности электронных словарей, обусловливаемая наличием разного рода поисковых интерфейсов.

Целью настоящей работы является анализ особенностей поисковой оболочки французского электронного словаря «Nouveau Petit Robert. Version électronique» (NPR 1997) (далее – NPR) [9] и выявление возможностей, которые она предоставляет при проведении лингвистических исследований разной направленности.

В том что касается непосредственно лексикографической базы и способов ее представления, NPR является точной копией исходного бумажного варианта (PRI 1992[10]), наследуя все его лучшие стороны. Он включает в себя около 60 000 словарных статей, 150 000 перекрестных ссылок, 180 000 примеров и устойчивых выражений, 40 000 цитат [9].

Электронная версия NPR отличается от бумажной следующим:

- гипертекстовой организацией статей с фиксацией «истории переходов»;
- мощной и дифференцированной полнотекстовой системой поиска;
- наличием аудиоданных (9 000 озвученных слов и около 100 озвученных цитат);
- высокой эргономичностью, включая возможность ограничивать вывод на экран избыточной (с точки зрения конкретных задач) информации словарной статьи;
- возможностью создавать и сохранять собственные примечания;
- интегрируемостью в MS Word (наличие специального макроса, позволяющего обращаться к данным NPR из документа, открытого в указанном текстовом процессоре);
- возможностью экспорта выбранных данных через буфер обмена MS Windows в текстовые редакторы и вывода их на печать.

Первая возможность в настоящее время является обязательной для всех сколько-нибудь серьезных информационных ресурсов. Она заключается в приложении гипертекстового метода навигации в информационных массивах к словарной статье. Специфика в данном случае состоит в том, что каждое слово (независимо от его грамматической формы) одной статьи является гиперссылкой на посвященную ему статью. Дополнительная функция, автоматически фиксирующая переходы от одной статьи к другой (так называемая история поиска), позволяет отслеживать образующиеся цепочки развертывания дефиниций и облегчает дальнейшую навигацию.

Последняя особенность делает NPR незаменимым при проведении компонентного анализа значения слова.

Исследователями отмечается существование глубинной связи между лексикографической дефиницией и компонентным анализом [11, p. 301–315]: «... l'information sémantique principale consiste dans la *définition*, qui est la forme lexicographique traditionnelle de l'analyse componentielle. La définition consiste en effet à rendre compte, sous la forme d'un énoncé analytique, des sèmes pertinents qui entrent dans la composition d'un sens» [5, p. XXXVII].

Необходимым дополнением к данному методу, существенно расширяющим его возможности, является метод развертывания словарных дефиниций. Гипертекстовое построение словарной статьи в полной мере отвечает потребностям исследователя, желающего использовать указанные методы.

Отметим, что семантические признаки (компоненты, семы), на основе которых осуществляется включение конкретного наименования в состав исследуемой лексико-семантической группировки, могут иметь различную степень эксплицированности в словарных дефинициях. Так, в случае с лексико-семантическим полем «Nourriture», явное выражение они получают включением в дефиниции исследуемых наименований слова 'nourriture' либо близкого по значению 'aliment'. Во всех

остальных случаях необходимо привлечение метода развертывания словарных дефиниций (ср.: метод ступенчатой идентификации [12, с. 107–108]). Например:

«BIFTECK n. m. Tranche de bœuf grillée ou destinée à l'être» [9].

→ «TRANCHE n. f. A. *Sens propre* 1. Morceau assez mince, coupé sur toute la largeur d'une chose comestible» [9].

→ «COMESTIBLE adj. et n. m. 1. Qui peut servir d'aliment à l'homme» [9].

К тому же использование данного метода позволяет выявлять как иерархическую организацию лексического значения отдельных наименований, так и иерархическую организацию лексики в пределах лексико-семантического поля. Данная возможность обуславливается спецификой организации а) словарной дефиниции, в которой выделяются два блока информации, находящихся в родовидовых отношениях [5, р. XXXVIII; 3; 12, с. 107], и б) лексико-семантической системы и структур представления знаний, характеризующихся древовидным иерархическим построением на основе родовидовых отношений.

Определенную трудность при использовании данного метода представляют случаи «перекрещивания» дефиниций [12, с. 104–106], свидетельствующие о незавершенности процессов дифференцирования лексического значения подобных наименований на данном историческом этапе:

«CONFISERIE n. f. ... 3. ... Friandise à base de sucre cuit...» [9];

→ «FRIANDISE n.f. ... 2. Petite pièce de confiserie ou de pâtisserie...» [9].

Решение данной проблемы заключается в обращении к дефинициям анализируемых наименований, приводимых *иными* словарями:

«FRIANDISE n. f. Préparation sucrée ou salée de petite dimension, d'un goût délicat. – Spécialt. Sucrierie ou petite pièce de pâtisserie» («Bibliorom LAROUSSE» BL 1996 [13]).

Впрочем, обращение к дефинициям одного и того же наименования, приводимым *разными* словарями, желательно в любом случае, поскольку позволяет собрать наиболее объективную и всестороннюю информацию о его семантике [12, с. 106–120; 14, р. 91–110].

Высокая **эргономичность** электронной версии NPR проявляется (помимо отсутствия необходимости постоянно листать весьма тяжелый и толстый фолиант) в удобном оформлении программной оболочки и «дружественном интерфейсе», наиболее важной особенностью которого является довольно мощная и дифференцированная система поиска внутри корпуса словаря. Рассмотрим подробнее ее особенности и возможности с точки зрения использования данного ЭЛР в образовательном процессе и исследовательской работе.

Поисковая система NPR представлена отдельным окном с пятью вкладками, соответствующими различным направлениям поиска.

Поиск по заголовкам статей (вкладка «Entrées»)

Расширенные возможности данного режима незаменимы при рассмотрении различных морфологических явлений.

Данная вкладка делится на три подменю:

- 1а) окно ввода искомого слова;
- 1б) флажковый переключатель режима «formes fléchies»;
- 1в) набор переключателей режима поиска;
- 2) подменю поиска по грамматическим категориям;
- 3) подменю поиска по парадигмам.

Что касается ввода искомого слов, следует особо отметить расширенные возможности использования символов-«джокеров»: символ «_» замещает одну любую букву, «*» – произвольное количество букв, следующих подряд, «#» – гласную, «&» – согласную. В рамках данного режима поиска «джокеры» могут располагаться в любой части вводимого слова. Так, в результате поиска по запросу «#a&t_» (два символа подчеркивания в

конце) были найдены следующие статьи: 'd'author' (статья 'autorité'), 'fauter', 'fautif, ive', 'hautin', 'laiton', 'laitue', 'maître, maîtresse', 'naître', 'paître', 'sauter'.

Следует обратить особое внимание, что при использовании «джокеров» поисковая система не дифференцирует наличие/отсутствие и тип диакритических знаков у гласных. Так, результатом поиска по запросу «saut&» стали статьи 'sauté' и 'sauté, ée'.

Переключатель режима «Formes fléchies» (по умолчанию сброшен) ограничивает поиск точным соответствием введенному слову («сброшен») либо позволяет выявить все статьи, включающие любую из возможных (и занесенных в базу словаря) форм введенного слова («установлен»).

Имеется три переключателя режимов поиска: «dans l'ordre» (по умолчанию), «par phonétisation» и «anagrammes». При включении второго режима поисковая система выводит список статей, слова-заголовки которых имеют сходное произношение с введенной в поисковое окно последовательностью символов. Данный режим полезен при поиске омонимов. Возможное задание может быть следующего типа: *найдите слова, произносящиеся как [par]* (выводятся следующие статьи: 'par – grér', 'par – n.m.', 'pare-', '-pare', 'part – n.f.', 'part – n.m.'). Режим поиска анаграмм предоставляет удобную возможность включения игрового момента при изучении лексики (отметим, что «джокеры» в данном режиме использованы быть не могут).

Второе подменю обеспечивает поиск по грамматическим категориям (флажок на переключателе «Inclure la sélection» автоматически устанавливается при выборе хотя бы одной строки из списка предложенных); например, в списке предлагаются следующие категории имен существительных: n., n.f., n.m., n.m. et n.f., n.m. ou n.f., n.pr. Из списка можно выбирать несколько вариантов (щелчок правой клавишей мышки при удерживаемой клавише «Ctrl»), которые при поиске будут объединены логическим оператором «ИЛИ» (нестрогая дизъюнкция).

При этом не следует забывать, что при использовании данного подменю поисковая система ищет статьи, грамматическая информация в заголовках которых *точно* соответствует выбранным строкам списка; то есть если вам необходимо найти все имена существительные, заканчивающиеся на «-re», в списке следует отметить *все* имеющиеся варианты, касающиеся существительных.

Третье подменю обеспечивает поиск по парадигмам и включает два поля списков:

- существительные (по этому выбору будут обрабатываться также и прилагательные), глаголы 1-й, 2-й, 3-й групп спряжения;

- собственно конкретные парадигмы с наиболее типичными примерами, например: «pluriel en -x ou -s [richelieu]»; «féminin en -gu/gue [aigu]»; «céder (6) * v. en é.er (altern é/è)»; «courir (11) * v. en -ir (présent en -s, -s, -t)», etc.

При выборе нескольких вариантов из списков второго и третьего подменю (щелчок правой клавишей мышки при удерживаемой клавише «Ctrl») данные варианты при поиске объединяются логическим оператором «ИЛИ» (нестрогая дизъюнкция).

При одновременном задействовании нескольких подменю выбранные варианты объединяются логическим оператором «И» (конъюнкция). Таким образом, если критерии поиска по подменю обозначить буквами А, В, С, то получится следующая формула поиска: $A \& (B_1 \vee B_2 \vee B_n) \& (C_1 \vee C_2 \vee C_n)$.

Следует обратить внимание, что при работе с глаголами задействовать второе подменю без особой на то нужды нежелательно, поскольку его использование будет заметно сказываться на формируемом списке результатов.

Режим поиска по произношению (вкладка «Phonétique»)

Данный режим по определению предназначен для работы над фонетическим материалом.

Соответствующая вкладка содержит а) окно ввода *транс-*

крипции, б) кнопки ввода знаков международной транскрипции и в) два переключателя режима поиска («dans l'ordre» и «anagrammes»).

Особенности работы с данной вкладкой не отличаются от вышеописанной (в том числе используются те же «джокеры»).

Как отмечают сами авторы и разработчики NPR, «la recherche par la phonétique proposée dans le CD-ROM du Petit Robert vous permet d'effectuer des recherches absolument inédites sur la prononciation des mots du dictionnaire. Le grand intérêt de ces recherches est de fournir en général des listes de mots souvent très éloignées graphiquement, mais partageant des caractéristiques communes de prononciation... Avec un peu d'imagination, les résultats vous surprendront!» [9].

Режим поиска в этимологических данных (вкладка «Étymologie»)

Данный режим по определению предназначен для использования в этимологических исследованиях, при рассмотрении лексико-семантических (и отчасти морфологических и фонетических) явлений в диахронии.

Данная вкладка включает в себя три подменю:

- 1) выбор языка,
- 2) выбор датировки,
- 3) поиск конкретных слов из этимологической части статей.

Первое подменю позволяет осуществлять поиск указаний на тот или иной язык (рассматривается около ста шестидесяти языков) в этимологической части статей. При этом можно ограничиться списком групп языков (около тридцати) либо сузить поиск, выбрав один или несколько (щелчок левой клавишей мышки, удерживая нажатой клавишу «Ctrl») языков конкретной группы. Можно осуществлять поиск по нескольким группам языков одновременно (отобранные критерии поиска будут объединены поисковой системой нестрогой дизъюнкцией).

Как правило, данный режим используется для выявления происхождения слов и отслеживания путей их заимствования.

Второе подменю позволяет найти все статьи, слова-заголовки которых впервые зафиксированы в письменных источниках определенного периода. Для этого в соответствующие окна нужно ввести граничные даты рассматриваемого периода. При этом нужно учитывать, что период не может превышать 99 лет. Кроме того, часть датировок представлена не в виде точных дат, а указаний на век или его часть. Для учета подобных довольно многочисленных данных необходимо задействовать флажковый переключатель «Inclure les siècles» (по умолчанию снят). Для частей веков поисковой системой приняты двадцатилетние интервалы: «déb. XVIIIe» соответствует интервалу 1701–1720 гг., «mil. XVe» – 1440–1460 гг., «fin XIXe» – 1880–1900 гг.

Особо отметим, что датировки, приводимые *внутри* статей (как правило, годы появления новых семем), не учитываются в данном режиме поиска, поскольку он распространяется только на этимологический блок информации статей.

Весьма продуктивным использование данного поискового режима может оказаться при *отслеживании динамики заимствований из определенного языка в диахронии*. При этом важным представляется сочетание подобной работы с исследованием экстралингвистических факторов, процессов во внутриполитической, экономической, духовной, внешнеполитической сферах рассматриваемого исторического периода. Подобный подход позволяет в полной мере говорить о выходе в сферу лингвокультурологии.

Третье подменю позволяет искать в этимологических разделах статей следующие виды данных: а) слово (будет осуществляться поиск либо точных соответствий, либо всех флективных форм – см. ниже), б) часть слова (для замены произвольного количества символов *на конце слова* использует-

ся единственный «джокер» – символ «*»), в) словосочетание, г) комбинация слов и служебных операторов (для ограничения поиска).

Режим «lemmatisation» позволяет находить все статьи, в которых встречаются все включенные в базу словаря флективные формы введенного слова (слов, входящих во введенное словосочетание). Если флажок данного режима не установлен (по умолчанию) поисковая система выводит список статей со словоформами, *точно* повторяющими введенную.

Для ограничения поиска предусмотрено использование двух логических операторов [«ET» («И») и «OU» («ИЛИ»)] и оператора «PRES/n», указывающего на удаленность друг от друга (n – число слов) двух введенных слов.

Отметим, что возможно *комбинированное* использование различных критериев поиска в пределах рассматриваемой вкладки.

Следует упомянуть также, что данный режим очень полезен при исследовании динамики концептов и систем языковых средств их репрезентации.

Поиск в тексте цитат (вкладка «Citations»)

Существенное количество цитат в сочетании с имеющимися поисковыми возможностями делают рассматриваемый ЭЛР достаточно полезным подспорьем в филологических и литературоведческих исследованиях.

Данная вкладка включает два подменю: выбор типа источника цитат и поиск конкретных слов в цитатах.

Первое подменю располагает набором переключателей, определяющих типы источников цитат:

- 1) писатели;
- 2) видные деятели;
- 3) авторы лексикографических, справочных, учебных изданий;
- 4) источники разного рода (известные религиозные,

юридические, политические тексты, цитаты из песен, опер, фильмов);

5) пресса.

В дополнение имеется список *конкретных* авторов и источников, содержимое которого меняется в зависимости от положения переключателя. Отметим, что в данном списке можно выбирать несколько позиций.

Подмену поиска слов полностью аналогично третьему подмену предыдущей вкладки. Как и в предыдущем случае, в случае одновременного использования критериев поиска обоих подменю, они объединяются поисковой системой отношением конъюнкции.

Кроме этого на данной вкладке имеется флажковый переключатель «Citations prononcées» (по умолчанию снят), ограничивающий поиск только *озвученными* цитатами.

Поиск во всей совокупности текста словаря (вкладка «Texte»)

Данный режим, который является наименее специализированным, позволяет выявить все словарные статьи, в которых встречается введенный текст. Как показывает личный опыт авторов, использование данного режима становится неоптимальным подспорьем при проведении лексикологических и семантических исследований.

Соответствующая вкладка содержит а) окно ввода искомого текста, б) флажковый переключатель режима «lemmatisation» и в) набор переключателей, ограничивающих поиск внутри статей [поиск во всей статье (по умолчанию) только в дефинициях, только в примерах и среди наиболее употребляемых словосочетаний, только в отсылках и антонимах].

Ввод искомого текста осуществляется по тем же правилам, что и в предыдущих двух вкладках. Однако именно здесь наиболее явно проявляется нехватка оператора отрицания «НЕ». Так, при поиске статей, включающих однокоренные с 'magie' слова, запрос «magi*» не позволяет автоматически отфильтровать слова типа 'magistrat', 'magistral', что вполне ре-

шаемо в поисковой системе «Bibliorom LAROUSSE», где можно задействовать следующую формулу запроса: «magi* поn magis*».

Использование переключателя «lemmatisation» позволяет существенно расширить список найденных статей. Так, форма 'aller' встречается в 775 статьях. При установке соответствующего флажка поисковая система обнаруживает 1949 статей, в которых встречаются различные видовременные и неличные формы глагола 'aller'.

Эта функция также полезна и при запросе на поиск словосочетаний (при поиске определенных конструкций, фразеологизмов, устойчивых словосочетаний, сложных родовых признаков лексического значения). Так, для поиска значения выражения «il y a anguille sous roche» значительно проще ввести три последних слова, чем самостоятельно открывать и просматривать нужные статьи.

Однако в случае с поиском словосочетаний установка флажка «lemmatisation» может привести к довольно неожиданным результатам. Так, при поиске словосочетания «action de faire» среди прочих была выведена статья 'déroulement', включающая пример «Le déroulement de l'action, des faits, des événements dans une pièce de théâtre, un film». Хотя в целом при четком представлении ожидаемых результатов несложно правильно сформулировать запрос, и использование данного режима может оказаться полезным. Так, при вводе запроса «le bien qui» со снятым флажком было выявлено две статьи при снятом и семь – при установленном флажке (за счет формы множественного числа «les biens»).

Очень полезным рассматриваемый режим может оказаться за счет возможности поиска помет (точку в сокращениях можно не ставить), что делает его ценным при проведении стилистических и лексикографических исследований.

Отметим также, что поскольку данный ЭЛР является аналогическим, а не идеографическим словарем, то его оптималь-

но использовать при первичном анализе фреймов.

Существенной недоработкой данного программного продукта, на наш взгляд, является невозможность комбинирования результатов поиска по разным вкладкам (допустим, включение функции «искать в найденном», используемой в поисковой системе «Bibliorom LAROUSSE» [13]). Решение данного вопроса сделало бы NPR поистине незаменимым инструментом исследователя-лингвиста.

В заключение подведем некоторые итоги.

Существующее в современном информационном пространстве разнообразие типов электронных словарей обуславливает наличие у них широкого спектра возможностей, а их природа – преимуществ перед традиционными бумажными.

NPR является, с одной стороны, обширным, разноплановым и эргономичным источником языкового (и в определенной мере лингвокультурологического) материала, с другой – удобным, многофункциональным инструментом лингвистических исследований различной направленности.

Различные поисковые режимы NPR могут использоваться для работы (как исследовательской, так и учебной) по следующим направлениям: морфология, орфография; фонетика; семантика, лексикология, стилистика; теория и практика перевода; история литературы французского языка, аналитическое чтение; история языка, лингвокультурология; когнитивная лингвистика, лингвоконцептология.

Список литературы

1. *Поликарпов А.А.* Стилистика, семантика, грамматика: опыт анализа системных взаимосвязей (По данным толкового словаря) [Текст] / А.А. Поликарпов, В.Я. Курлов // Вопросы языкознания. 1994. № 1.

2. *Беркетова З.В.* Мотивационные связи в лексике современного русского языка [Текст] // НДВШ. Филологические науки. 2000. № 1.

3. *Fargues, J.* Synthetising a Large Concept Hierarchy from French Hyperonyms [Text] / J. Fargues, A. Perrin // *Coling'90*. Helsinki: University of Helsinki, 1990. V. 2.

4. *Trigano Ph.* La définition: un outil pour l'acquisition de concepts nouveaux dans un lexique sémantique [Ressource électronique] / Ph. Trigano, B. Talon ; [Papier proposé aux 12-èmes Journées internationales sur les systèmes experts et leurs applications, Avignon, juin 1992]. – Режим доступа: <http://www.hds.utc.fr/~ptrigano/avi92.html>.

5. *Imbs, P.* Préface [Texte] / P. Imbs // *Trésor de la langue française. Dictionnaire de la langue du XIX^e et du XX^e siècles*. P.: CNRS, 1971. V. 1. P. XI–XLVIII.

6. *Swanepoel, P.* Dictionary Quality and Dictionary Design: A Methodology for Improving the Functional Quality of Dictionaries [Text] / P. Swanepoel // *Lexikos*. 2001. V. 11.

7. *Михайлова О.А.* Ограничения в лексической семантике русского слова [Текст]: автореф. дис. ... д-ра филол. наук. Екатеринбург, 1998.

8. *Морель Морель, Д.А.* Электронные лексикографические ресурсы в дистанционном обучении лингвистическим дисциплинам [Текст] // *Гуманитарные науки: сб. науч. трудов*. Белгород: СГУ, 2006. Вып. 13.

9. *NPR – Nouveau Petit Robert. Dictionnaire alphabétique et analogique de la langue française* [Ressource électronique]: Version électronique. Bruxelles: Bureau van Dijk, Éditions électroniques, 1997. – 1 электрон. опт. диск (CD-ROM).

10. *PR1 – Le Petit Robert 1. Dictionnaire alphabétique et analogique de la langue française* [Texte] / P. Robert, A. Rey, J. Rey-Debove. P.: Dictionnaires Le Robert, 1992. XXXI, 2174 p.

11. *Wooldridge, T.R.* Procédés lexicographiques, tant anciens que modernes [Text] / T. R. Wooldridge // *Computational Lexicology and Lexicography: Special Issue Dedicated to B. Quemada*. Pisa: Giardini, 1990. T. 2.

12. *Попова З.Д.* Лексическая система языка [Текст]:

Внутренняя организация, категориальный аппарат и приемы изучения / З.Д. Попова, И.А. Стернин. Воронеж: ВГУ, 1984.

13. *BL* – Bibliorom LAROUSSE [Ressource électronique]. P.: Larousse, 1996. – 1 электрон. опт. диск (CD-ROM).

14. *Fillmore, Ch. J.* Describing Polysemy: The Case of «Crawl» [Text] / Ch.J. Fillmore, B.T.S. Atkins // Polysemy: Theoretical and Computational Approaches. Oxford: O.U.P., 2000.