

Возможности использования базы данных в двустороннем анализе русско-чешской археологической терминологии

Станислав Оплатек

Университет им. Масарика, Брно, Чешская Республика

В течение 2006-ого года в университете им. Масарика в Брно была выполнена бакалаврская диссертация по анализу использования баз данных в языкознании [Оплатек 2006], целью которой явилось построение лингвистической базы, которая удовлетворяла бы следующим требованиям:

- а) простота использования и возможность адаптации базы к нуждам лингвистов;
- б) простота создания базы в такой степени, чтобы подобную базу мог по предложенному образцу построить любой лингвист;
- в) программа, в которой база данных написана, должна быть общедоступна.

В своей работе мы сосредоточились на лексикографической и лексикологической обработке русско-чешской археологической терминологии. Поскольку мы были намерены использовать базу данных в двух, не совсем совпадающих лингвистических «направлениях» – а) лексикографическом и б) лексико-морфологическом, то одним из критериев при ее разработке явилось стремление построить по возможности наиболее универсальную базу данных, которая бы позволила осуществить работу в этих направлениях.

Целью нашей лексикографической работы являлось расширение и уточнение русско-чешского археологического словаря А. Бранднера [Brandner 1994], а также разработка электронной версии словаря, которую в то же время можно и напечатать. Так как в области терминологии точность перевода является ключевым требованием к словарям, нами также была осуществлена попытка обеспечить возможность обратной проверки правильности перевода каждого термина пользователями словаря.

Мы пытались совместить в базе данных ту информацию, которая позволяла бы использовать нашу базу и в виде особого языкового корпуса. Поскольку нашей целью являлось прежде всего намерение подать общую морфологическую и словообразовательную характеристику русско-чешской

археологической терминологии, мы сосредоточились в первую очередь на тех данных, которые позволяют осуществить указанные виды анализа. Однако, несмотря на эти требования, мы стремились построить такую базу данных, в которую было бы возможно внести не только морфологическую и словообразовательную, но и другую информацию.

Наш проект основан на терминографических базах данных, которые описаны в научных публикациях [Křečková 2001, Jugčácková 2002, Macháček 1995]. Особое внимание мы обратили на способы использования этих баз в других общественных науках, т. е. не только в лингвистике, но и в археологии [Macháček 1997] и библиотечном деле. Поскольку в настоящее время MS ACCESS является наиболее распространённой базовой программой, мы решили выполнить проект именно на ее базе (в частности MS ACCESS 2003). За основной тезис, на котором строится настоящая концепция базы данных, мы принимаем общепринятое в терминологической науке положение о том, что терминам присущи все те же лексические характеристики, что и лексике общеупотребительной, включая многозначность, омонимию, синонимию. Если слово-термин обладает многозначностью, имеет больше двух значений, связанных с разными понятиями, то каждому такому слову в базе следует подобрать определённое количество соответствующих ему эквивалентов. Нами были предложены, проанализированы с точки зрения возможностей и соответствия целям и задачам работы три проекта. Так как первые два проекта, как показало их тестирование с помощью собранных терминологических данных¹, оказались неподходящими для лексикографической и одновременно терминографической обработки материала и сравнительного анализа двух языков, то для нашего базового проекта был выбран третий вариант.

Третий проект – двухступенчатая база данных. Так как она работает отдельно с данными, касающимися понятия, и данными, связанными со словом как морфолого-лексической единицей, то она состоит из двух

¹ В каждый тип базы данных было введено 100 терминов из собранных нами данных, насчитывающих 1955 русских и 1802 чешских лексических единиц. Путём базовых запросов отдельные типы баз были протестированы (определение синонимии, многозначности, принадлежность терминов к тематическим группам, словообразование).

уровней – ступеней таблиц – для каждого из языков (русского и чешского). Первая ступень таблиц обеспечивает работу с лексическими (род, число и т.п.) и словообразовательными характеристиками. Она представляет слово как речевой знак, выражающий понятия. Многозначность термина обеспечивает вторая ступень таблиц базы данных – таблицы, отражающие отношение термина к понятию. На этой ступени каждому значению термина, обозначаемому в языке определённым словом, соответствует одна таблица. Такой «веерообразный» характер отношения (слово – понятие) обеспечивает связь таблиц $1 : n^2$, т.е. одна таблица «Слово» связана с неопределённым, по сути бесконечным количеством таблиц «значение-понятие». Таким способом обеспечена и двусторонность базы, так как одному русскому слову-термину отвечает ровно одно значение слова-термина чешского.

Систему таблиц мы назвали двухступенчатой, так как такое название имеет своё обоснование не только в графической схеме таблиц, но и в схеме распределения данных по отдельным таблицам. При рассмотрении лексики с помощью базовой системы следует различать определяемые категории слова как лексико-морфологической единицы и как единицы лексико-семантической. Так, некоторые из них свойственны слову, а некоторые – его значению. Их правильное распределение в таблицах важно для безошибочного функционирования базы данных.

Характеристики термина в базе данных были подобраны и распределены по таблицам следующим образом: *таблице «Слово»* были приписаны общие характеристики термина, а также характеристики 1) морфологические: род, число, грамматические варианты термина (в том числе ударения), количество слов, образующих термин, количество корней в термине; 2) словообразовательные: происхождение (собственное, заимствованное слово), способ словообразования (морфологический, семантический способ), словообразовательные аффиксы (префикс, суффикс, интерфикс), характер словообразовательных аффиксов (специальные, общелитературные),

2 Знаки „1 : 1“, „1 : n“ выражают соотношение, связь между базовыми таблицами. „1 : 1“ означает, что одной конкретной таблице соответствует в базе данных одна и только одна таблица другого характера, с другими данными. Связь „1 : n“ означает, что одной конкретной таблице соответствует неопределённое количество однородных таблиц.

исходное производящее слово (его форма и принадлежность к части речи); для составных терминов были классифицированы разложимость термина и его принадлежность к определённой словообразовательно-грамматической модели; 3) другие характеристики: источник, из которого был взят термин, лексические варианты термина (сокращения и т.п.), количество значений, присущих термину, примечания.

В *таблице, относящейся к значению*, мы приводим следующие характеристики: термин (значение), энциклопедическое значение (толкование), тип термина (т.е. его принадлежность к определённой логической категории, например 'процесс', 'орудие' и т. п.), синонимы термина, антонимы, гипонимы, гиперонимы, способ образования (морфологический или семантический), сфера и характер употребления (общелитературный язык, диалект, архаизм, исконно специальное слово).

С помощью такого распределения характеристик термина можно получить большое количество различных данных, дающих исчерпывающую характеристику рассматриваемых терминов с точки зрения как словообразования, так и терминографии. Одновременно эту базу можно использовать как археологический русско-чешский – чешско-русский словарь.

Список литературы

- Винокур 1939 – *Винокур, Г. О.* О некоторых явлениях словообразования в русской технической терминологии / Г. О. Винокур // Тр. МИФЛИ. – 1939. – Т. V.
- Ильина 1981 – *Ильина, С. В.* Возможности использования фактографической ИПС в качестве инструмента лексикографической обработки научно-технических терминов / С. В. Ильина // Термин и слово. – Горький : Изд-во ГГУ им. Н. И. Лобачевского, 1981. – С. 75–81.
- Оплатек 2006 – *Оплатек, С.* Специальная терминология и её лексикографическая переработка : Бакалаврская работа / Станислав Оплатек. – Brno: Filozofická fakulta Masarykovy univerzity, 2006.
- Поликарпов – *Поликарпов, А. А.* Славянская компьютерная лексикография: задачи и проблемы [Электронный ресурс] / А. А. Поликарпов. – Режим

доступа : http://www.philol.msu.ru/~lex/articles/slav_lex.htm, свободный. –
Загл. с титул. страницы.

Brandner 1994 – *Brandner, A.* Rusko-český slovník pro archeology. – Brno :
Masarykova univerzita, 1994.

Čermák 1995 – *Čermák, F.* Komputační lexikografie / F. Čermák // Manuál
lexikografie. – Praha : H & H, 1995. – P. 50–71.

Křečková 2001 – *Křečková, V.* Informácie v terminologických databázich a ich
využívanie při preklade odborných textov / V. Křečková // Slovenčina a čeština
v počítačovom spracovaní / Ed. A. Jarošová. – Bratislava : Veda, 2001. –
P. 142–149.

Jurčacková 2002 – *Jurčacková, Z.* Terminológia. Základné zásady, metódy a ich
aplikácia / Z. Jurčacková. – Bratislava : Centrum vedecko-technických
informácií SR, 2002.

Kubálek, and et. 2003 – *Kubálek, T.* Manažerská informatika. Databázový systém
Microsoft Access verze 2002 CZ / T. Kubálek, L. Švecová. – Praha :
Oeconomica, 2003.

Macháček 1997 – *Macháček, K.* Počítačová podpora v archeologii / K. Macháček. –
Brno : Masarykova univerzita v Brně, 1997.

Machová 1995 – *Machová, S.* Terminografie / S. Machová // Manuál lexikografie. –
Praha : H & H, 1995. – P. 137–157.

Podborský, and et. 1977 – *Podborský, V.* Numerický kód moravské malované
keramiky. Problémy deskripce v archeologii / V. Podborský, E. Kazdová,
P. Koštuřík, Z. Weber. – Brno : Univerzita Jana Evangelisty Purkyně, 1977.

Poštolková, and et. 1983 – *Poštolková, B.* O české terminologii / B. Poštolková,
M. Roudný, A. Tejnor. – Praha : Academia, 1983.

Possibilities for using databases in a two-sided analysis of Russian-Czech
archaeological terminology

Stanislav Oplatek

Masaryk University, Brno, Czech Republic

In 2006 a linguistic database was set up at Masaryk University in Brno as a
part of an effort to extend an existing Russian-Czech archaeological dictionary. The

database is to be utilized not only as an electronic dictionary but also as a particular language corpus, which makes it easy for the collected data to be edited as well as further exploited as a terminological language corpus for general linguistic analysis.