

K návrhu, vývoji a funkcím lexikální databáze češtiny¹

Jindra Světlá

Ústav pro jazyk český AV ČR, v. v. i.

On the Specification, Development and Functions of a Lexical Database of the Czech Language

Since 2005, the Department of Lexicography and Terminology (DLT) of the Institute of the Czech Language (ICL) of the AS CR, v. v. i., has focused on the creation of the material and technical preconditions for using modern information technologies when investigating and describing vocabulary, and mainly on the design and development of our own lexicographic software, ranging from initially specifying the requirements, through developing and testing a special lexicographic program called PRALED, to launching a lexical database designed by us on our own server. The programming part involves the development of the relevant software, the empty shell of the database and the interface for data input, which should also serve after 2010 for the development of a new monolingual dictionary named LEXIKON 21. Our department has simultaneously focused on conceptual work linked with a future detailed treatment of Czech vocabulary in the lexical database.

Etapy vývoje a budování lexikální databáze

v Ústavu pro jazyk český AV ČR, v. v. i.

Výzkumný záměr nazvaný *Vytvoření databáze lexikální zásoby českého jazyka počátku 21. století* se na jedné straně zaměřil na vytvoření předpokladů pro využití moderních informačních technologií při výzkumu české slovní zásoby (k tomu viz v tomto sborníku podrobně příspěvek A. Rangelové), na druhé straně na budování vlastní lexikální databáze.

Výzkumný záměr byl zahájen v roce 2005; původní návrh počítal s nákupem, úpravou a zprovozněním programového vybavení pro tvorbu lexikální databáze češtiny během prvních šesti měsíců (v rámci etapy přípravných prací). Avšak vzhledem k tomu, že žádný vhodný lexikografický software specializovaný pro potřeby češtiny dosud neexistuje a nedá se koupit hotový, **přípravná etapa** potřebná k vytvoření našeho vlastního programu a prázdného jádra databáze byla celkově mnohem delší, než se původně předpokládalo; zahrnuje dvě fáze vývoje softwaru.

První fáze vývoje trvala přibližně rok a půl (leden 2005 až červen 2006); během ní bylo nutno konkrétně definovat veškeré požadavky na software včetně detailní specifikace celé požadované struktury databáze i všech jednotlivých polí

¹ Příspěvek vznikl v rámci výzkumného záměru ÚJČ AV ČR, v. v. i. *Vytvoření databáze lexikální zásoby českého jazyka počátku 21. století* (AV0Z90610521).

(rubrik, položek) umožňujících zápis dat do heslového formuláře (o nich bude podrobněji pojednáno dále), postupně je zadávat programátorům, průběžně testovat všechny úpravy a celý program postupně doladovat. Paralelně s tímto úkolem bylo třeba promýšlet i jednotlivé, dílčí části teoretické koncepce nového lexikografického popisu, aby bylo možno vůbec navrhnout metodiku práce, všechny požadované položky formuláře a kompletní nabídky v nich, a to diferencovaně pro jednotlivé slovní druhy, případně pro různé typy hesel (takto byly rámcově připraveny obecné zásady např. pro zpracování termínů, frazémů, sousloví atd. v databázové struktuře). Výsledkem této úvodní fáze byla první verze speciálního programu nazvaného *Praled*.²

Druhá fáze vývoje probíhá od července roku 2006 a je zaměřena na specifikaci nového uživatelského rozhraní a postupné doplňování dalších funkcí zmíněného programu; toto nové rozhraní bylo nazváno *Praled 2*.

Teprve další fáze, v zadání výzkumného záměru označená jako **etapa vlastní výstavby lexikální databáze**, bude zaměřena na postupné naplňování hesláře a jednotlivých položek databáze jazykovými daty v souladu se zadáním výzkumného záměru: pro každou položku databáze (lexikální heslo) mají být vyhledány výskyty v textových zdrojích a poté vyhledávky seřazeny a klasifikovány v konfrontaci se zpracováním ve starších slovnících. Takto má být vytvořeno cca 100 000 až 120 000 položek – záznamů o slovech užívaných mluvčími současného českého jazyka.³ Jako výchozí, pracovní heslář pro první etapu prací byl vybrán Frekvenční slovník češtiny obsahující 50 tisíc lemmat. Tento soubor lemmat je již vložen do našeho hesláře a v další fázi bude postupně doplňován.

Věcný obsah databáze se po skončení stávajícího výzkumného záměru má stát východiskem pro vypracování výkladového slovníku současné češtiny; jednotlivé položky databáze budou sloužit jako materiálový základ pro vypracování příslušných slovníkových hesel. Pro obsah databáze (kvůli jeho odlišení od části programové) jsme zvolili název *Pralex*.

Pro databázové zpracování budoucího výkladového slovníku je určen název *LEXIKON 21* (*L 21* je jeho zkrácené, pracovní označení). Pro rozlišení obsahu těchto názvů používáme též někdy výrazy „lexikální databáze“ (*Pralex* – současná databáze naplňovaná v rámci výzkumného záměru) a „slovníková databáze“ (*LEXIKON 21* – budoucí „nadstavba“, tj. nové slovníkové zpracování).

² *Praled* (*Pražská lexikální databáze*) – program vyvíjený ve spolupráci lexikograficko-terminologického oddělení ÚJČ AV ČR, v. v. i., a Centra zpracování přirozeného jazyka FI MU v Brně (programátor Adam Rambousek).

³ Koncem roku 2007 byl úkol našeho oddělení dále konkretizován – nyní se zaměřujeme převážně na exemplifikační část hesla, tj. na podrobný popis spojitelnosti heslových slov pracovně členěných na jednotlivé významy podle SSJČ. Práce koncepčního charakteru zaměřené na budoucí slovník byly prozatím zastaveny, resp. omezeny pouze na potřeby současného databázového zpracování. Naše činnost je v současné fázi orientována především na kvantitativní nárůst obsahu a na naplnění databáze požadovaným počtem jednotek.

Struktura lexikální databáze realizovaná v programu *Praled*

Z uživatelského hlediska jsou hlavními součástmi databáze zejména **vstupní obrazovka** s přihlašovacím oknem, hlavní okno aplikace nazvané **Heslář** a okno pro jednotlivé **heslo** s formulářem obsahujícím **nástroje** pro zápis dat a přepínání mezi kartami **Heslo** a **Význam** + zobrazení v **Náhledu**.

Do databáze jsou integrovány dva softwarové nástroje: morfologický analyzátor *AJKA* a slovníkový prohlížeč *DEBDict* pro rychlé vyhledávání ve slovnících (tyto dva nástroje vznikly v Brně na FI MU). Připravujeme rovněž samostatný program *Pramat* pro třídění dokladů (náš vlastní program, jeho vývoj pokračuje – viz podrobněji v příspěvku Z. Opavské a B. Štěpánkové); dále nutně potřebujeme doplnit program pro pokročilé vyhledávání v *Praledu* – i na něm se průběžně pracuje.

Vstup do databáze: Na vstupní obrazovce se oprávněný uživatel v přihlašovacím okně přihlásí jménem a heslem (jméno uživatele se zároveň automaticky zapíše do položky Zpracovatel u zpracovávaného hesla). Tím se dostaneme do hlavního okna aplikace – na **Heslář**, který nám slouží jako východisko všech dalších činností. Z obrazovky Heslář můžeme dále pokračovat několika způsoby: a) přes funkci Nové heslo založíme nový prázdný formulář, b) výběrem z hesláře (dvojitým kliknutím na příslušnou položku) otevřeme heslo generované automaticky ze seznamu (popř. už vyplněné, zpracované), c) zadáme do vstupního pole pro vyhledávání požadované lemma, a tím otevřeme rovnou konkrétní heslo, nebo zadáme řetězec znaků, čímž se v hesláři zobrazí všechna hesla obsahující daný řetězec.

Obecné principy zápisu dat do databáze: Na rozdíl od jiných programů formulářového typu (např. v účetnictví), kde se u každé položky formuláře zapisuje jeden konkrétní, jednoznačný údaj, lexikografický popis slovní zásoby představuje mnohaúrovňový rozhodovací proces, který se projeví jak na složitosti postupu při zápisu dat, tak i na složitosti struktury samotného pracovního formuláře. Proto se údaje do databáze vkládají buď zápisem do textového pole, nebo výběrem z nabídky, nebo automaticky či programově (automaticky se vkládají následující údaje: zpracovatel, datum vytvoření a změny; programově byly např. ke všem lemmatům z FSC vloženy statistické údaje o jejich frekvenci v korpusu SYN2000).

Část informací se bude vkládat formou zápisu do textového pole, tj. metajazykovým popisem v podobě slovního komentáře. U těchto popisných údajů by bylo žádoucí v rámci možností sjednocovat používané formulace. Týká se to jak budoucích výkladů významu podle jednotlivých slovních druhů a lexikálněsémantických skupin, tak i formulací v nejrůznějších poznámkách, komentářích apod., aby se – vedle jednotného zpracování – následně usnadnilo i formalizované vyhledávání nejrůznějších údajů v rámci celé databáze.

Část údajů přiřazovaných ke všem lexémům vznikne jako výsledek klasifikace a zařídování k jednotlivým kategoriím, takže i zde je vhodné od samého počátku budování databáze vycházet z jednotného pojetí jak při určování, tak i při zápisu jednotlivých kategorií do formuláře metodou výběru z pevné nabídky v menu. K tomu bude vypracován Manuál pro zpracovatele, který bude shrnovat celkové pojetí práce s databází, jednotlivé nabídky pro výběr a základní pracovní postupy pro zápis dat v různých formulářích podle typu hesla. Teoretická část zásad zpracování hesel v databázové podobě vzniká postupně a průběžně se upřesňuje a ověřuje na základě analýzy zkušební vzorku, na základě analýzy zpracování lexikálních jednotek v předchozích slovnících, studiích literatury k jednotlivým tématům atd.

Aby byl požadovaný výsledek co nejvíce „uživatelsky přátelský“, bylo třeba rozhodnout, jak bude vypadat podoba všech formulací, které se ve formuláři vybírají nebo vyplňují. V nabídce by se měl objevit úplný výčet možností, které pro danou informaci mohou být relevantní. Výběrem z takovéto nabídky v menu u všech položek formuláře, kde je to možné, zajistíme formalizovanou podobu velkého množství údajů, které se do databáze dostanou, a tím usnadníme jejich vyhledání u všech hesel, kde to bylo jednotně vyplněno. (Takto se postupně mění některá textová pole na základě zkušeností získaných ze zpracování určitého počtu hesel na výběr z pevné nabídky, abychom této jednotnosti dosáhli.)

Také názvy všech položek (rubrik) ve formuláři musí mít promyšlené znění, protože neslouží jen zpracovatelům jako návod pro vyplnění, ale objeví se i v zobrazení jako součást metajazyka popisu, a kromě toho budou sloužit budoucím uživatelům jako filtr nebo atribut pro vyhledávání. Rozhodovali jsme se proto mezi názvy domácími a internacionálními, mezi názvy zkrácenými a plnými, ať již tvoří titulky jednotlivých položek, anebo jsou součástí pevné nabídky (např. *subst. x podst. jm. x podstatné jméno*) apod.

K návrhu a vývoji heslového formuláře v souladu s připravovanou koncepcí lexikografického popisu

Na základě zahraničních zkušeností (Kanada, V. Británie aj.) jsme se v úvodní etapě vývoje rozhodli, že souběžně s naplňováním databáze podle požadavků stanovených ve výzkumném záměru budeme na vybraných databázových položkách („heslech“) zároveň testovat možnosti podrobného zpracování heslové statí pro potřeby budoucího elektronického výkladového slovníku (slovníkové databáze L 21). Proto – tedy již s výhledem do budoucna, ale zároveň se zřetelem k české lexikografické tradici – jsme se rovnou zaměřili na navržení mnohem podrobnější struktury heslového formuláře, než bylo požadováno výzkumným záměrem, aby byla zajištěna návaznost prací i po roce 2010. Protože z hlediska této slovníkové databáze bude důraz kladen na významové členění každého hesla, na nové výklady významů a jejich dokládání dostatečným počtem autentických dokladů, navrhli

jsme specifické formuláře pro budoucí podrobnější lexikografický popis jazykových jednotek různého typu (různou strukturu budou mít hesla jednoslovná, více-slovná, podslovní části a zkratky); formulář bude obsahovat i rubriky, u kterých předpokládáme zpracování v rámci kooperace s dalšími pracovišti.

Jak již bylo výše naznačeno, v první fázi prací bylo třeba navrhnout celou strukturu hesla, postupně specifikovat jednotlivé položky, které se budou v pracovním formuláři u každého hesla vyplňovat, určit jejich pořadí ve formuláři, způsob jejich zápisu, stanovit požadavky na zobrazení a vyhledávání zapsaných dat i na propojení mezi různými hesly, uvažovat o návrhu výstupních formátů atd. K realizaci tohoto úkolu vznikly v rámci našeho oddělení dílčí pracovní skupinky, kterým byly ke zpracování prvních návrhů přiděleny jednotlivé okruhy témat (např. sekce morfologické informace; expresivita a další typy příznakovosti; zpracování variant; pojetí polysémie a homonymie; pojetí synonym a antonym; uvádění exemplifikace; poznámkový aparát atd.). Každá pracovní skupinka tak navrhovala svou část heslového formuláře (tj. příslušnou sekci/informační zónu a položky, které se v ní budou vyplňovat) – první návrh viz na obr. 1.

Při testování prvního formuláře realizovaného programátorem v první polovině roku 2006 se však ukázalo, že takto navržená podoba není dostatečně pružná pro

Obr. 1. Vzhled původního formuláře s pevným pořadím sekcí i jednotlivých položek

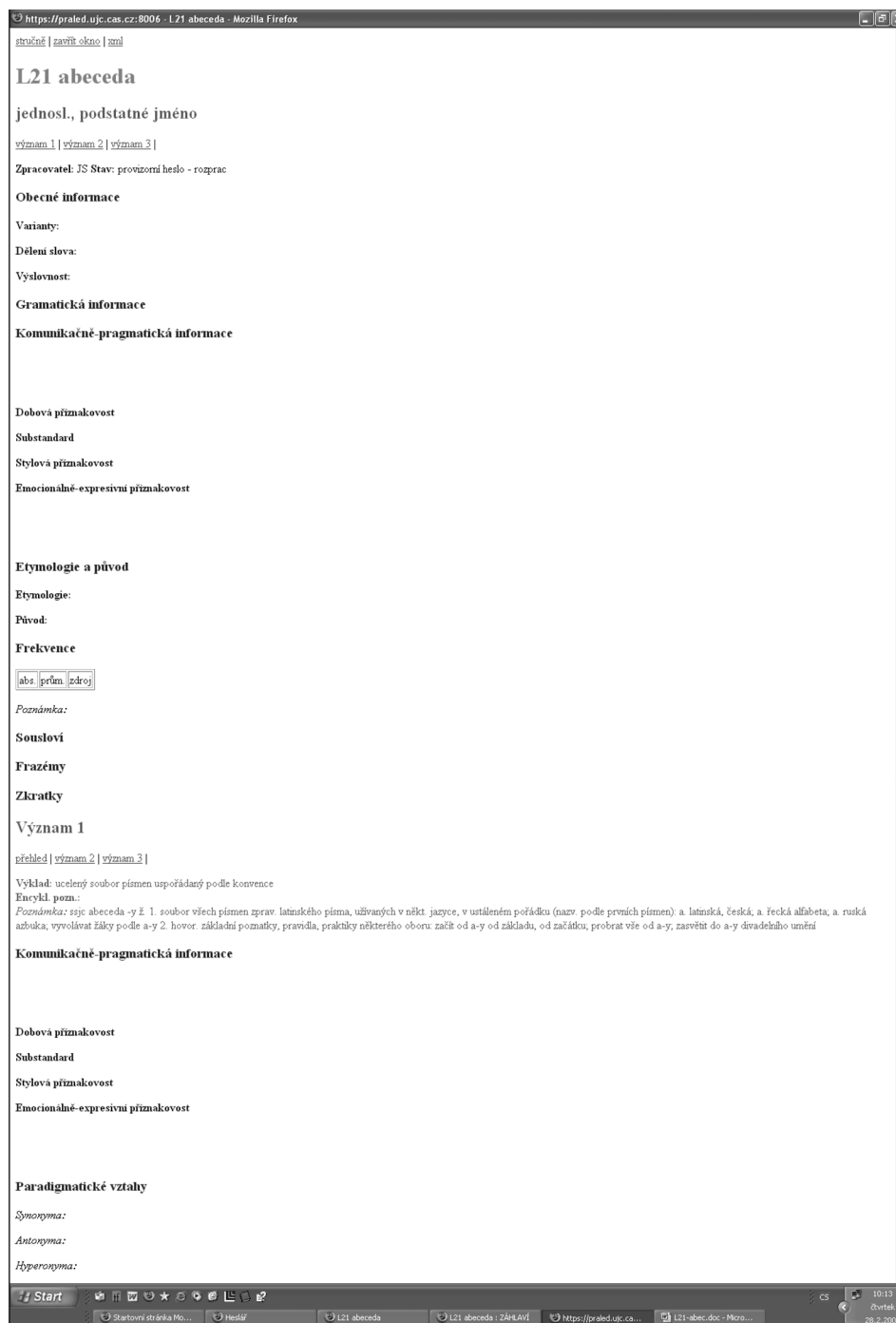
zápis složitě strukturovaných údajů a že formulář s pevným pořadím jednotlivých sekcí a vyplňovaných položek v každé z nich není pro zápis lexikografických dat příliš vhodný, mj. je příliš dlouhý a nevejde se celý najednou na monitor počítače. Při zpracování zkušebního vzorku cca 1000 hesel se potvrdilo, že navržený formulář by mohl stačit pro jednoduché zápisy a ukládání dat do databáze v intencích původních požadavků výzkumného záměru, ale z hlediska budoucího zpracování elektronického výkladového slovníku příliš nevyhovuje speciálním požadavkům uvažovaného komplexního lexikografického popisu, který by se měl realizovat na různých hierarchických úrovních (rovina celého hesla, rovina jednotlivých významů, rovina exemplifikace).

Ukázalo se totiž, že původně navržené pevné, lineární uspořádání formuláře, kdy byly všechny položky umístěny po sekcích napevno pod sebou, nám bude u delších, zejména polysémních jednotek často dělat problémy s možností pohybu mezi jednotlivými částmi zpracovávaného hesla (v podstatě bylo možné pouze rolování nahoru/dolů, anebo přepínání mezi kartami a okny bez možnosti vidět jich více najednou).

Nelze předpokládat, že budeme schopni při zpracování hesla postupovat ve formuláři vždy lineárně od začátku do konce – takový postup je zřejmě možné zachovat např. při závěrečném přepisu zpracovaného hesla z předem připravených podkladů, jak se to dělalo dříve, ale ne při tvorbě heslové stati jako celek prostřednictvím zápisu dat různého druhu rovnou do strukturovaného formuláře. Bylo tedy třeba klasické metody lexikografie, jako např. ruční třídění lístkových dokladů, postupný zápis zpracovaných podkladů na samostatné karty podle jednotlivých významů apod., promítnout do možností a metod elektronického zpracování hesla v databázové podobě.

Zkušenosti získané na základě zpracování zkušebního vzorku jasně prokázaly, že v databázovém formuláři a jeho struktuře je třeba více zohlednit náročné pracovní postupy lexikografického zpracování hesel, protože se v rámci daného hesla potřebujeme pohybovat všemi směry a mít možnost rychle se vracet k jednotlivým částem formuláře a doplňovat či revidovat údaje na různé úrovni hesla a v různých fázích zpracování, např. budeme potřebovat přecházet mezi jednotlivými významy a postupně vzájemně upřesňovat výkladové definice, průběžně vkládat doklady k různým významům, atd.

Proto se od poloviny roku 2006 začalo na základě našeho nového zadání vyvíjet **nové uživatelské rozhraní, nazvané *Praled 2*** (programátor Adam Rambousek za Centrum zpracování přirozeného jazyka FI MU v Brně, nově ve spolupráci s pražským programátorem Pavlem Žikovským za ÚJČ), které mělo jednak odstranit nedostatky první verze, jednak bylo třeba mimo formulář navrhnout také samostatné zobrazení celého hesla, které více odpovídá našim potřebám i tradiční tištěné podobě slovníkových hesel (původní náhled hesla byl pro naše potřeby nevyhovující, viz obr. 2). V souvislosti s dalším upřesňováním našich požadavků se tak



Obr. 2. Původní náhled hesla navržený programátorem

nakonec změnil celý design formuláře, princip používání jednotlivých „nástrojů“ v něm obsažených i zobrazení dat v novém Náhledu, který v nové verzi zůstává stále na pracovní ploše celé obrazovky.

Nové uživatelské rozhraní: *Praled 2*

Hlavní rozdíl v práci s novým rozhraním je ten, že se původní formulář „rozpadl“ na řadu dílčích částí – samostatných oken, která se otevírají pomocí tlačítek jen v případě potřeby (a tudíž nezabírají místo na pracovní ploše, pokud je aktuálně nepotřebujeme). Tato okna označujeme jako jednotlivé „nástroje“; většinou odpovídají původnímu členění formuláře na dílčí sekce podle typu informací, které se do nich vkládají. Každý nástroj obsahuje různé množství rubrik – o nich mluvíme jako o „položkách“ formuláře. Abychom nezaměňovali označení „položka databáze“ a „položka formuláře“, v rámci makrostruktury (celé databáze) mluvíme o **heslech**, popř. položkách hesláře, kdežto v rámci mikrostruktury (struktury hesla) mluvíme o jednotlivých **položkách** konkrétního nástroje (ať již z hlediska heslového formuláře a jeho nástrojů, nebo z hlediska zobrazení vyplněných položek v Náhledu).

K základním funkcím programu *Praled 2* patří tedy **vkládání dat** do konkrétních *položek* pomocí samostatných *nástrojů*, **zobrazení** vložených dat jak v jednotlivých nástrojích, tak i v celkovém náhledu, **tisk** vybraných sestav; důležitá je též funkce **vyhledávání** požadovaných dat podle různých parametrů (filtrů): např. podle jednotlivých slovních druhů, podle data a/nebo zpracovatele, podle kombinace požadovaných údajů, podle zadaného řetězce znaků – abecedně nebo retrográdně atd. Nové uživatelské rozhraní nám umožnilo vkládat údaje do různých nástrojů v libovolném pořadí, otevírat více nástrojů najednou, vložená data okamžitě zobrazovat ve výstupním formátu na pracovní ploše, informace stejného druhu v zobrazení jednotně formátovat a odlišit jednotlivé informační zóny (sekce) a typy informací různou barvou, typem a velikostí písma apod., uspořádat některé informace do seznamů a využívat hypertextové odkazy, atd.

Struktura heslového formuláře a funkce programu *Praled 2*

Struktura hesla je rozčleněna v souladu s požadavky budoucího lexikografického popisu na několik úrovní, neboť jsme tradiční strukturu slovníkového hesla museli převést do nové, databázové podoby.

Nabídka nástrojů a položek (odpovídající příslušnému formuláři pro zápis dat) se v příslušném hesle konkretizuje na základě výběru slovního druhu nebo typu hesla. Každá budoucí heslová stať má hierarchickou strukturu a z hlediska významového členění může mít jednu až čtyři úrovně popisu – údaje platné pro celé

heslo, údaje vztahující se k jednotlivým významům, popř. tzv. podvýznamům, a údaje uváděné v Exemplifikaci⁴.

Vzhledem k databázové struktuře je i monosémní lexikální jednotce přiřazeno číslo významu (v nástroji se uvádí jen č. 1, v zobrazení jako kombinace písmene a čísla, tedy V1).

Systém „karet“ (karta Heslo – karta Význam – karta Podvýznam): Výchozí pracovní plochou pro zpracování nového hesla je obrazovka nazvaná **karta Heslo**, hierarchicky nižší úroveň zpracování představuje **karta Význam**. Mezi kartami různé úrovně (karta Heslo – karta Význam) i mezi jednotlivými významy (z jedné karty na druhou, např. Význam 1 – Význam 2 atd.) se přepíná pomocí názvů karet umístěných na horní liště obrazovky. Karty pro jednotlivé významy se přidávají/odebírají podle potřeby pomocí tlačítek Přidat význam nebo Odebrat význam umístěných rovněž na horní liště; přidané významy se okamžitě automaticky očíslovají, ale pořadí významů lze měnit zatím jen ručně.

Pro první význam je rovnou připravena karta **Význam 1**, na kterou se přepíná z karty Heslo kliknutím. Pro otevření karty pro další význam je třeba postupovat přes tlačítko Přidat význam – tím se přidá a pak otevře další karta (s názvem Význam 2 – v zobrazení V 2 atd.).

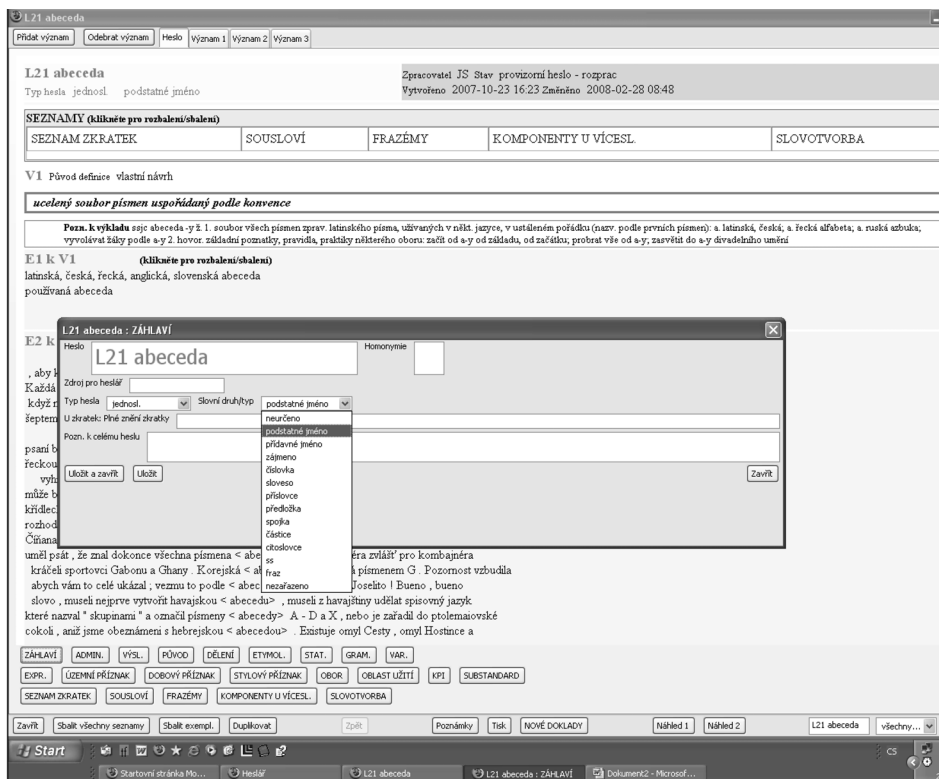
Každý význam lze dále členit hierarchicky na nižší úroveň popisu, která je pracovníčně označena jako „podvýznam“. Zatím není přesně vymezeno, k čemu bude systém podvýznamů používat – zda k naznačení hierarchie a vazeb mezi jednotlivými významy (v souvislosti s pořadím významů v hesle), anebo ve funkci významového odstínu (v souvislosti se zpřesňováním výkladu). Výchozí návrh pro programové řešení počítal s tím, že podvýznam by mohl odpovídat významovému odstínu, přičemž jeden význam jich může mít víc. Karta **Podvýznam** má proto všechny nástroje totožné s kartou Význam, což při zpracování usnadní pozdější rozhodování, zda jde o samostatný význam nebo podvýznam (stačí změnit označení). Aby bylo možné měnit pořadí jednotlivých významů včetně tohoto hierarchického členění, Podvýznamy by měly být programově spárovány s daným Významem (P1 k V1, P2 k V1) tak, aby i po změně pořadí zůstaly podřazeny témuž významu. Konceptně je třeba možnosti využití podvýznamů dále analyzovat na základě určitého vzorku zpracovaných hesel a výhledově rozhodnout, jak se bude tato možnost pro slovník využívat.

Systém samostatných „nástrojů“: Zápis dat do databáze se v novém uživatelském rozhraní realizuje prostřednictvím jednotlivých nástrojů. Všechny nástroje se ote-

⁴ Přesto (a nebo právě díky tomu), že se moderní technologie elektronického zpracování obsáhlých materiálových souborů neustále zdokonalují, výzkum současného lexika i nadále zůstává časově náročnou záležitostí, protože nové možnosti vyhledávání LJ a třídění dokladů v korpusech zároveň zvětšují rozsah a objem zkoumaného materiálu. Proto bude dále podrobněji strukturována i dokladová část hesla, nazvaná Exemplifikace – práce na její struktuře průběžně pokračují.

vírají pomocí tlačítek na spodní liště jako samostatná okna (viz nástroj Záhlaví na obr. 3). Ke každé kartě (karta Heslo, resp. karta Význam) se bude nabídka relevantních nástrojů v podobě tlačítek na dolní liště obrazovky automaticky měnit (či opakovat) v závislosti na jejich funkci ve struktuře heslové stati. Nabídka nástrojů pro kartu Heslo a pro kartu Význam tedy není totožná, např. nástroj Záhlaví je jen na kartě Heslo, kdežto nástroj Gramatické informace je k dispozici na různých úrovních hesla – jak na kartě Heslo, tak na kartě Význam.

Změn oproti původnímu provedení pracovního formuláře pro zápis dat je více: v prvním návrhu formuláře se všechno řešilo pouze přepínáním karet (Heslo/Význam), všechno ostatní bylo na pevném místě ve formuláři, pouze byla možnost některé části „stáhnout“; i náhled hesla se musel otevírat zvlášť, byl příliš dlouhý, protože se v něm zobrazovaly nadpisy všech položek, i nevyplněných (srov. obr. 2) – toto řešení bylo nevyhovující. Nyní se všechny nástroje otevírají jen v případě potřeby, aby zbytečně nezabíraly pracovní plochu obrazovky. Program též umožňuje otevírat více nástrojů zároveň a podle vlastního uvážení si je rozmístit na ploše



Obr. 3. Celá pracovní plocha na obrazovce včetně zobrazení zpracovávaného hesla, s otevřeným nástrojem Záhlaví a otevřenou dílčí nabídkou Slovní druh/typ, z níž se vybírá kliknutím

obrazovky. Každý nástroj si „pamatuje“ svou poslední pozici a formát (velikost) okna, vše se dá ovládat a měnit myší. Toto řešení je nové a bylo časově a programově náročné. Důležité pro nás také je, aby se jednotlivé otevřené nástroje samy nezavíraly (či nestáhly na lištu) při otevření dalšího nástroje, ale aby jich mohlo zůstat otevřeno více najednou. Každé okno si zpracovatel zavře sám podle potřeby, případně si některé otevřené okno prostřednictvím funkce Minimalizovat stáhne na lištu, pokud je jich otevřeno mnoho najednou, ale chce s nimi dále pracovat.

Po vyplnění jedné či více položek v příslušném nástroji data uložíme do databáze a okno nástroje zavřeme. Po opětovném otevření daného nástroje jsou zde vyplněné údaje vidět – to je základní, **přímé zobrazení** uložených dat.

System „položek“: Každý nástroj (okno) obsahuje určitý počet položek, které se vyplňují buď výběrem z nabídky v menu (např. slovní druh, viz obr. 3), anebo formou ručního zápisu (např. lemma) do textového pole (někdy jde i o delší slovní popis, např. v poznámkách), případně se vkládají automaticky (např. údaje o frekvenci v korpusu *SYN2000*).

Při vyplňování 1. nástroje (ZÁHLAVÍ) se vybere v příslušné položce typ hesla (jednoslovné – víceslovné – podslovní část – zkratka) a slovní druh, případně dílčí typy (sousloví, frazémy apod.); podle toho se uplatní formulář s požadovanou strukturou. Např. u jednoslovných hesel se některé dílčí nabídky (gram. aj.) ve formuláři liší podle slovních druhů, takže bez výběru a označení slovního druhu nemůžeme pokračovat dál.

Hodnoty zapsaných nebo z nabídky vybraných atributů lze vyhledávat (zatím jen některé) pomocí Filtrů, tj. prostřednictvím funkce Pokročilé vyhledávání na obrazovce Heslář.

Programové provázání celé heslové struktury slouží pro snadnou orientaci v nástrojích heslového formuláře. Např. je-li v heslovém formuláři vyplněno lemma (buď se zobrazí automaticky při spuštění z hesláře, nebo se u nového hesla zapíše ručně), jednak se ihned automaticky objeví v celkovém zobrazení daného hesla na ploše obrazovky (v tzv. **Náhledu**), a jednak se zároveň při vkládání dalších dat do jednotlivých nástrojů toto lemma zobrazuje jak na horní liště obrazovky na kartě Heslo, tak i na horní liště rámečku každého otevřeného nástroje jako název zpracovávaného hesla (srov. na obr. 3 údaj „L21 abeceda – ZÁHLAVÍ“. V každém okamžiku své práce tak víme, které heslo a který nástroj používáme, protože můžeme mít otevřených jak více hesel, tak více nástrojů u jednoho hesla najednou. Kromě toho se dále automatickým značením diferencuje úroveň celého hesla a rovina jednotlivých významů, takže, otevřeme-li např. nějaký nástroj na kartě Význam, označení lemmatu na horní liště nástroje má navíc přiřazeno ještě označení příslušného významu.

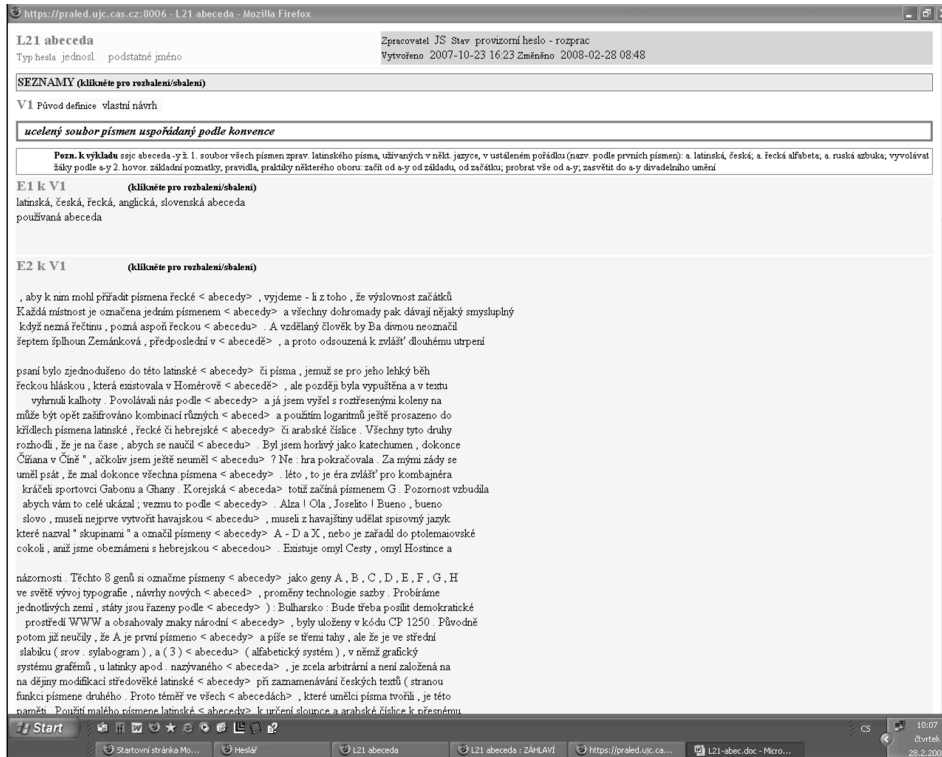
Po vyplnění Záhlaví (tj. po zapsání lemmatu a určení typu hesla a slovního druhu, které je výchozí, protože generuje výběr jednotlivých formulářů) je proto

možné zápis všech dat pro dané konkrétní heslo provádět v libovolném pořadí, výběrem kteréhokoli nástroje. Po zápisu do příslušných položek ve zvoleném nástroji je třeba vše uložit a okno zavřít.

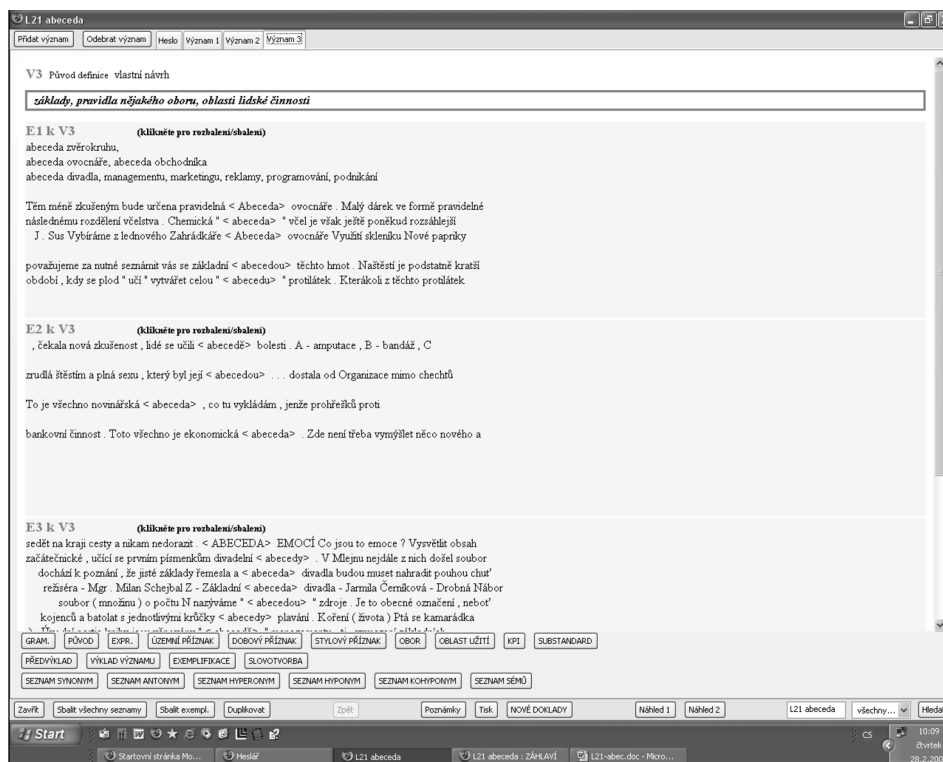
Přímé zobrazení vložených dat v jednotlivých nástrojích a v Náhledu celkově:

V každém nástroji jsou vyplněné položky vidět pouze tehdy, když je tento nástroj otevřen, což je při následném prohlížení celého hesla nepraktické. Proto jsme navíc navrhli samostatnou funkci zobrazování všech zapsaných údajů v tzv. Náhledu hesla, který se ihned po zapsání a uložení dat do databáze objeví na pracovní ploše obrazovky, a to již ve výstupním formátu v podobě budoucí „slovníkové stati“, včetně různých barev a typů písma definovaných pro vložená data podle jednotlivých nástrojů (např. lemma se zobrazí tučně červeně, výklad významu je v červeném rámečku kurzivou, exemplifikace má světle modrý podklad, blok poznámek se uvádí v tenkém černém rámečku petitem).

Pořadí tlačítek nástrojů na dolní liště i pořadí položek uvnitř nástrojů jsme navrhli tak, aby vše odpovídalo logické strukturaci heslové stati a bylo pokud možno v souladu s českou lexikografickou tradicí; ve stejném pořadí se vyplněná data



Obř. 4. Celkový náhled zpracovaného hesla



Obr. 5. Náhled jednoho významu (V3 u hesla abeceda)

budou většinou zobrazovat i ve zmíněném Náhledu (viz obr. 4). Vedle celkového zobrazení všech vložených dat na kartě Heslo slouží pro kontrolu zápisu také dílčí zobrazení na kartě Význam, kde se zobrazují **pouze údaje vztahující se k danému významu** (viz obr. 5).

V Náhledu se tedy ihned po uložení zobrazí **všechny zapsané údaje**. Programově jsme zajistili, aby se zobrazovaly jen položky skutečně vyplněné (platí tedy zásada, že nevyplněné položky ani jejich titulky se nezobrazují). Zobrazení v Náhledu se bude na jedné straně blížit tradičnímu zobrazení heslové stati ve výkladových slovnících, na druhé straně bude přihlížet k možnostem elektronického databázového zpracování (bude obsahovat i hypertextové odkazy, některé informace jsou prezentovány v podobě seznamů atd.).

Zpracovatel tak může jednotlivá data zapisovat v libovolném pořadí – tj. po vyplnění povinných položek v nástroji Záhlaví může pokračovat na kterékoli kartě a kteroukoli položkou v libovolném nástroji, přičemž uložená data se budou v Náhledu automaticky řadit na správnou pozici a zobrazovat v předem určeném formátu.

Pokud chce někdo otevřít heslo jen pro prohlížení, vidí zobrazení všech vyplněných údajů ihned na obrazovce, ale nemůže v tomto náhledu nic měnit (k tomu je třeba otevřít příslušný nástroj a úpravy provést v něm; přístup do nástrojů mají jen zpracovatelé).

Závěr

Tím, že jsme se již od samého počátku přípravných prací zaměřili na to, aby námi navržené softwarové nástroje umožňovaly jak průběžné naplňování všech položek navržených pro jednotlivé formuláře bez ohledu na pořadí ve formuláři, tak i zobrazení všech vyplněných údajů okamžitě po uložení v grafické podobě blízké tradičnímu tištěnému slovníku, jsme vytvořili předpoklady pro to, abychom mohli po roce 2010 plynule přejít na další úkol – zpracování nového výkladového slovníku v elektronické podobě (slovníkovou databázi).

Námi navržený program, realizovaný ve spolupráci s brněnským pracovištěm FI MU, bude sloužit pro ukládání relevantních jazykových dat do lexikální databáze a vyhledávání požadovaných informací různého typu pomocí zadaných atributů a jejich hodnot. Po naplnění lexikální databáze bude možné budoucím uživatelům poskytnout komplexní zobrazení každého hesla nebo dílčího souboru jak v elektronické podobě, tak i v podobě tištěné.

Literatura

- FREKVENČNÍ slovník češtiny. Nakladatelství Lidové noviny, Praha 2004. (FSČ)
- PALA, K.: Návrh české lexikální databáze, In: Slovenčina a čeština v počítačovom spracovaní, Bratislava 2001, s. 155.
- PALA, K. – HORÁK, A. – RAMBOUSEK, A. – RANGELOVA, A.: Nové nástroje pro českou lexikografii – DEB2. In: Gramatika a korpus. UJČ AV ČR, Praha 2007, s. 190–196.
- Využívané korpusy a programy:
- Český národní korpus – SYN2000. Ústav Českého národního korpusu FF UK, Praha 2000. Dostupný z WWW: <<http://ucnk.ff.cuni.cz>>.
- Český národní korpus – SYN2005. Ústav Českého národního korpusu FF UK, Praha 2005. Dostupný z WWW: <<http://ucnk.ff.cuni.cz>>.
- Český národní korpus – SYN2006PUB. Ústav Českého národního korpusu FF UK, Praha 2006. Dostupný z WWW: <<http://ucnk.ff.cuni.cz>>.
- Word Sketch Engine, dostupný z WWW: <<http://ucnk.ff.cuni.cz/corpora>>.