

# Zápis morfologických dat – návrh řešení pro lexikální databázi LEXIKON 21<sup>1</sup>

Milada Voborská  
Ústav pro jazyk český AV ČR, v. v. i.

## Tools for Inputting Morphological Data – A Proposal for the Lexical Database LEXIKON 21

The morphological tools developed within PRALED conform to our endeavour to describe the morphological features of a word (while taking into consideration also the vagueness of word classes and their subcategories). By these means, it will be possible to capture and describe i.a. the fluctuation in the gender, the fuzziness of using the number of nouns, or to discover a word with similar characteristics but belonging to another word class.

Popis morfologických charakteristik heslových slov v počítačové databázi se odlišuje od klasického slovníkového zpracování zejména využitím technických možností počítačového programu a rovněž dostatku místa pro záznamy o nejuplněnějších a přehledných informacích. Exemplifikace slova v konkrétním tvaru v širším kontextu pak umožní uživateli snazší pochopení jazykových situací, v nichž se daný tvar vyskytuje, i explicitní demonstraci toho, jaký příznak může nabývat užití určitého slovního tvaru v některých textech.

Pracovní nástroje pro zápis morfologických dat v lexikální databázi byly navrženy tak, aby bylo možné postihnout kolísání mezi slovními druhy a jejich kategoriemi, průniky a přechody mezi nimi a aby byl prostor pro popis těch jevů, které nejde jednoznačně zařadit do jednotlivých mluvnických kategorií. Tato snaha byla vedena jak různými možnými interpretacemi z hlediska lingvistického, tak i přihlédnutím k současným technickým možnostem (např. vyhledávání určitých jevů). Chtěli jsme se vyhnout nutnosti jednoznačného zařazení např. u **výrazů s významovým rysem kvantifikace**, kdy slova *dvojice*, *moře* je možné v určitých kontextech interpretovat jako substantivum nebo číslovku, *mnohamiliónový* jako adjektivum i číslovku atd., nebo u **deiktických slov**, která se řadí k různým slovním druhům, ale vyznačují se totožnými rysy, jako např. *něco*, *nějaký*, *několik*, *někde*.

<sup>1</sup> Příspěvek vznikl v rámci výzkumného záměru ÚJČ AV ČR, v. v. i. *Vytvoření databáze lexikální zásoby českého jazyka počátku 21. století* (AV0Z90610521). LEXIKON 21 – lexikální databáze češtiny navrhovaná v lexikograficko-terminologickém oddělení ÚJČ AV ČR, v. v. i., pro zpracování budoucího výkladového slovníku; L 21 – současné pracovní označení i pro tento budoucí výkladový slovník, který se v ÚJČ AV ČR začne zpracovávat až po roce 2011. Jednotlivé části koncepce postupně vznikají od roku 2005 v rámci současného výzkumného záměru *Vytvoření databáze lexikální zásoby českého jazyka počátku 21. století* (AV0Z90610521), viz též v tomto sborníku příspěvek J. Světlé *K návrhu, vývoji a funkcím lexikální databáze češtiny*.

Slova, která mají v různých užitích platnost různých slovních druhů, budou zpracována jako samostatná hesla, tj. na různých heslových kartách. Na volbě slovního druhu totiž závisí nabídka nástrojů pro popis morfologických kategorií, která se objeví po kliknutí na příslušné tlačítko ve spodní části obrazovky editovaného hesla. Zapsané údaje se pak zobrazí v řádce pod záhlavím hesla (nevyplněné položky zůstanou skryty). Stejně nástroje je možné použít jak pro celé heslo, tak i pro jednotlivé významy mnohovýznamových slov. Další možností pro uvedení gramatické informace by měla být poznámka vztahující se k paradigmatu, tak aby bylo možné komentovat příznaková užití některých tvarů slov, variantních koncovek apod., s níž by, jak už bylo řečeno, měl korespondovat exemplifikační blok, v němž budou zachycena jejich kontextová užití.

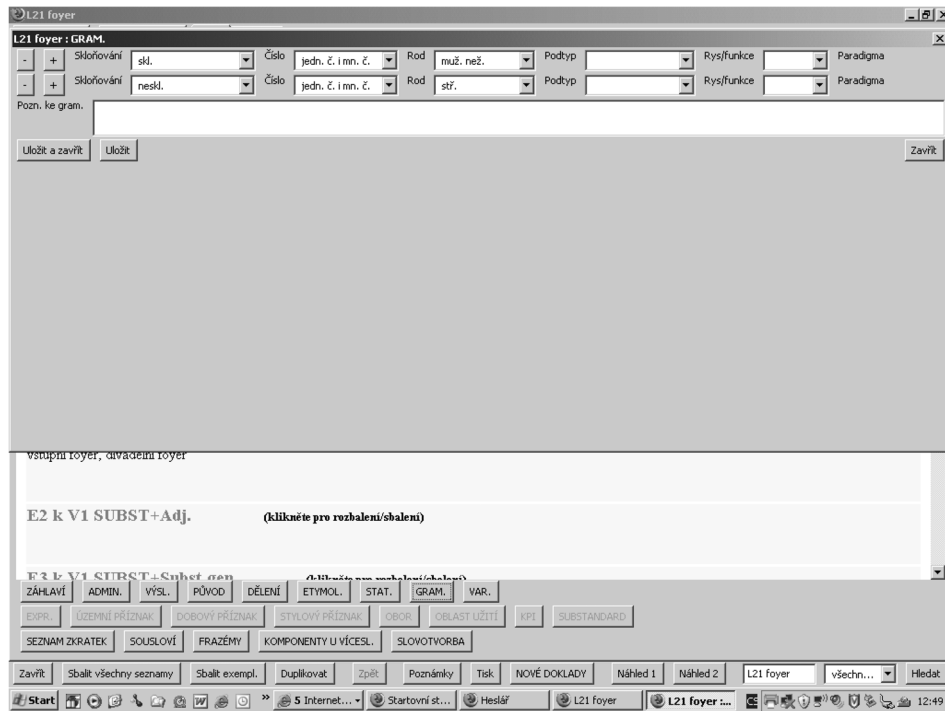
**Volba nástrojů pro zápis morfologických dat u jednotlivých slovních druhů a možnosti jejich využití v budoucím výkladovém slovníku – slovníkové databázi LEXIKON 21** (dále L 21) (nejvíce jsou tyto nástroje a nabídky v nich připraveny pro substantiva a adjektiva, pro další slovní druhy bude třeba návrh řešení dále rozpracovat):

## Substantiva

Skloňování heslového slova je možné popsat díky výběru připravených položek, které vystihnou **flektivnost** popisovaného hesla. Zvlášť je možné označit situaci, kdy se slovo skloňuje i zůstává neskloňné – položka „skl. i neskl.“ (*foyer*), anebo u substantivizovaných adjektiv zvolit položku naznačující způsob sklonění „jako přídavné jméno“ (*pracující*). Položky je možné **duplikovat**, takže pokud je nutné slovo zařadit ke dvěma skloňovacím typům, např. ke skloňování podle mužského rodu životného i neživotného (*slaneček*), druhou možnost naznačíme **přidáním další řádky** s totožnými položkami (pomocí tlačítka +), na jejímž konci by mělo následovat úplné paradigma (viz obr. 1).

Tento postup nám umožní popsat případy, kdy např. slovo existuje v rodě mužském i středním, avšak v mužském rodě pouze v čísle jednotném (*kníže*), nebo kdy se slovo skloňuje v čísle jednotném podle rodu středního, ale v čísle množném podle rodu ženského (*dítě*). **Každý typ skloňování bude mít svou vlastní řádku** (duplikuje se stisknutím tlačítka +), kde je možné specifikovat, pro jaké číslo dané skloňování platí.

Možné **odlišnosti užití jména v určitém čísle** budou zaznamenány buď v gramatické poznámce – např. pokud plurál u jmen látkových signalizuje druh (*různá vína*), míru (*dát si dvě piva*), intenzitu (*dlouhé pochody*) apod.; tato spojení pak budou doložena a vysvětlena v rámci gramatické informace v samostatném exemplifikačním bloku. Na úrovni jednotlivých významů, pokud užitím slova v plurálu vzniká nový význam (např. *paměť x paměti*) můžeme na kartě daného konkrétní-



Obr. 1. Nabídka v nástroji Gramatická charakteristika (pro substantiva) s možností přidání další řádky

ho významu jak v podheslí, tak v sekci gramatické informace specifikovat, že jde pouze o plurálové užití.

Pro slova, která se v určitém významu užívají **převážně** nebo **pouze** v **čísle jednotném** nebo **množném**, jsou připraveny položky „zprav. jedn. č.“, „zprav. mn. č.“, „pouze jedn. č.“, „pouze mn. č.“.

**Kolísání v rodě** při shodném významu slova i shodné podobě v nominativu singuláru bude řešeno uvedením obou (více) paradigmat, a to **přidáním celé řádky** (pomocí tlačítka +) s gramatickými informacemi (*smeč, kápo*). U jmen typu *držgrešle, miláček* bude vysvětlivka o užití pouze jednoho rodu pro pojmenování představitelů obou pohlaví uváděna v **gramatické poznámce** (a bude jednotná pro všechna slova tohoto typu). Užití životných koncovek u jmen neživotných (*najít hřiba, dát góla* ap.) bude zaznamenáno v **paradigmatu**, kde budou mimo to uváděny všechny spisovné alternativní koncovky; pokud užití některé z nich bude příznakové, v **gramatické poznámce** pak bude následovat vysvětlivka o tomto typu užití. Jména typu *choť, bačkora, naivka*, která mají shodný tvar nominativu singuláru pro mužský i ženský rod, ale následně jiné skloňování, budou řešena jako **samostatná homonymní hesla**.

V exemplifikaci pak budou poznámky uvedené u gramatické informace doloženy na konkrétních příkladech; budeme se snažit uvádět kontexty pro všechny varianty (případný příznak bude uveden v příslušné kolonce vedle pole pro kontexty).

Jako **Podtyp** je pracovně nazvána kolonka sloužící k uvedení informace o typu podstatného jména (jméno pomnožné, hromadné, abstraktní, látkové, vlastní jméno).

Kolonka **Rys/funkce** nabízí prozatím dvě charakteristiky, a to pro podstatná jména s rysem kvantifikace, pro něž je připravena zkratka **kvant.** (bude i u dalších slovních druhů), tak aby bylo možno po zadání zkratky vyhledat všechna slova vyznačující se tímto příznakem (*polovina, stovka, moře*). Zkratka **predik.** bude využita u toho významu polysémního lexému, v němž bude podstatné jméno v predikátu (*zima, hanba, radost, škoda*).

**Jako samostatná hesla** budou zpracována podstatná jména slovesná, substantivní pojmenování vlastností (-ost), jména přechýlená a zdobněliny (ty podle stupně lexikalizace a frekvence v korpusu).

## Adjektiva

První informace bude sloužit k upřesnění **způsobu skloňování** přídavných jmen. Pokud existují významové rozdíly v užití složených nebo jmenných tvarů, v příslušné gramatické informaci na kartě daného významu bude uvedeno, které z tvarů se k danému významu vztahují: např. zatímco složeného tvaru *hodný* i jmenného *hoden* může být užito ve významu „zasluhující si něčeho“ – *je hodný naší pozornosti / je hoden naší pozornosti*, pouze složený tvar přídavného jména *hodný* pak můžeme užít ve významech „dobrý, laskavý“ nebo „velký“ (kus cesty).

Další nabídka u adjektiv slouží k zaznamenání podrobnější informace, o jaký **typ** přídavného jména jde, případně jak bylo vytvořeno (jména přivlastňovací, adjektiva vzniklá z přechodníků a příčestí, např. *synův, tekoucí, pečený, ustrnulý*). Podobně jako u substantiv i zde bude využita zkratka **kvant.** k zaznamenání rysu kvantifikace (*mnohý, pětimilionový*).

Jako poslední v sekci gramatické informace uvedeme na konci řádky údaj o **stupňování** přídavných jmen.

O **zařazení** přídavných jmen odvozených od slovesných tvarů **do hesláře** rozhodne jejich lexikalizace a frekvence lexému. Budou vyložena, nebo u nich bude jen odkaz, který navede uživatele na příslušné sloveso.

## Zájmena

Vymezení jednotlivých druhů zájmen bude konzultováno s odborníky z oddělení gramatiky. Vycházíme z teorie, že zájmena spolu s dalšími výrazy podobné

povahy tvoří **skupinu deiktických slov**. Kromě zájmen tvoří skupinu deiktických slov především zájmenná adverbia (viz dále) a zájmenné číslovky. Všechna by měla mít v L 21 označení **deikt.** = deiktické slovo.

Zároveň jsou zájmena nositeli rysu kvantovosti, která je podle MČ relevantním, popř. konstitutivním komponentem negativ, totalizátorů a některých indefinit. Jde o následující řadu zájmen a obdobné řady deiktických výrazů: *nikdo, žádný – sotvakdo – zřídka – málokdo – kdekdo – každý, všechen*. Tyto výrazy by tedy měly kromě označení deikt. ještě značku **kvant.**

Naším záměrem je naznačit i syntaktickou funkci zájmena ve větě (substantivní, adjektivní).

## Číslovky

Je třeba dořešit jak rozdělení druhů číslovek, tak způsob popisu, aby byly zaznamenány všechny podstatné rysy. V akademické mluvnici je pojetí číslovek velmi široké, jsou mezi ně řazena i taková pojmenování, která slovníky hodnotí jako substantiva nebo adjektiva (*čtvrt, polovina, dvojice, stovka; nulový, tisícový, mnohamilionový*). S tím souvisí i rozlišování některých netradičních skupin číslovek (dílových, skupinových, velikostních). Pro připravovanou databázi navrhuje tyto skupiny výrazů považovat v souladu s dosavadními slovníky za substantiva a adjektiva. Dodatečně by byly na příslušném místě označeny zkratkou pro rys kvantifikace (tato charakteristika by se přidávala ke všem číslovkám a ke slovům, jež jsou na přechodu mezi číslovkami a substantivy, adjektivy, zájmeny nebo příslovci a vyznačují se stejnými rysy). Zároveň bychom chtěli naznačit, že po formální a syntaktické stránce jsou některé číslovky povahy substantivní (*milion*), adjektivní (*pátý*) nebo adverbialní (*dvakrát*). Druhy číslovek budou rovněž ještě konzultovány.

## Slovesa

Cílem budoucího komplexního lexikografického popisu v rámci slovníkové databáze L 21 by mělo být popsat slovesa z hlediska všech jazykových plánů (rovin): foneticko-fonologického, morfologického (tvaroslovného a slovtvorného), syntaktického, lexikálního a stylisticko-pragmatického.

Naším záměrem je u každého slovesa uvést v sekci pro zápis morfologických dat jednak úplné paradigma jeho tvarů včetně tvarů dubletních (a tak budou informace o gramatických kategoriích slovesné osoby, čísla, způsobu, času a rodu uvedeny v tomto přehledu), jednak jsou u tohoto slovního druhu navrženy položky, které budou obsahovat nabídku (menu) pro výběr adekvátních charakteristik vystihujících popisované heslo. Jedna z nich bude určena pro zápis **slovesného vidu**

a **vidového protějšku**, následovat budou **syntaktické a sémantické charakteristiky** slovesa, jako např. valence, valenční pole, gramatický větný vzorec, sémantický větný vzorec a další. Vybrané charakteristiky se promítnou i v exemplifikaci.

## Příslovce

Tradiční **rozlišení významových skupin** adverbii na příslovce místa, času, způsobu, příčiny bude obsaženo **ve výkladu významu** (srov. *SSJČ: kam – zájm. přísl. 1. táz. ptáme se jím po místním cíli děje*), ani u adverbii však nebude vždy zařazení jednoznačné, protože významově se mohou jednotlivá užití lišit (např. ve spojení *neobyčejně krásný* je význam prvního slova „velmi“, zatímco ve spojení *neobyčejně oblečený* je význam adverbia „neobvyklým způsobem“).

Ve formuláři je připravena nabídka pro to, aby bylo možné označit stavová a modální predikativa, příslovečné deiktické výrazy (*jak, kdy, tam, všude* apod.), které budou mít označení deikt. (společné pro všechny deiktické výrazy), a označení kvant. pro příslovečné výrazy vyjadřující kvantum. U **zájmenných příslovečných výrazů** je možnost označit, zda jde o výrazy ukazovací, tázací, vztažné, neurčité nebo záporné.

**Stupňování** některých adverbii bude podobně jako u přídavných jmen uvedeno na konci gramatické informace.

## Literatura

MLUVNICE češtiny 2, 3. Praha, Academia 1986.