

K MOŽNOSTEM A MEZÍM REGISTRACE SLOVOTVORNÝCH VZTAHŮ V LEXIKÁLNÍ DATABÁZI¹

On the possibilities and limitations of registering word-formation relations in a lexical database

Pavla Kochová

Abstrakt: Článek informuje o projektu, jehož cílem je registrovat a graficky prezentovat slovotvorné vztahy v lexikální databázi, která stojí na pozadí Internetového slovníku současné češtiny. Slovotvorné vztahy byly zpracovány zčásti automatizovaně, zčásti manuálně a pro vnější uživatele slovníku jsou zobrazeny v přehledném diagramu.

Klíčová slova: slovotvorba, derivace, slovotvorný vztah, lexikální databáze

Abstract: The article provides information about a project whose aim is to register and graphically present word-formation relations in a lexical database, which is the background of the Internet Dictionary of Contemporary Czech. Word-formation relations were processed partly automatically, partly manually and are displayed via a clear scheme for external users of the dictionary.

Keywords: word-formation, derivation, word-formation relation, lexical database

Cílem článku je představit záměr registrovat a graficky prezentovat slovotvorné vztahy v lexikální databázi, která stojí na pozadí *Internetového slovníku současné češtiny* (dále ISSČ).²

¹ Článek vznikl s podporou projektu *Využití multimediálního výkladového slovníku pro moderní výuku češtiny* v programu ÉTA 5 (TA ČR), reg. č. TL05000546.

² *Internetový slovník současné češtiny* (<https://www.nechybujete.cz/slovník-soucasne-cestiny>) byl poprvé zveřejněn 10. 10. 2013. Ve verzi 3.0 (z r. 2024) zahrnuje 70 000 hesel, 85 000 významů (lexií), 190 000 příkladů, vč. ustálených slovních spojení a frazémů.

Nástroj, který by systematicky registroval slootovorné vztahy mezi lexémy obsaženými v lexikografickém díle, není zatím pro českou slovní zásobu k dispozici. Přesto lze dílčím způsobem navazovat na starší slovníková díla, a to díky lexikografické technice hnízdování, která byla v tradičních výkladových slovnících určitou měrou aplikována. Využita byla zejména ve *Slovníku spisovné češtiny pro školu a veřejnost*.

V poslední době vznikly inspirativní softwarové derivační aplikace, zejména databáze *DeriNet* (Ševčíková, Žabokrtský, Vidra, Straka, 2016). Tento softwarový nástroj však slouží k jiným než lexikografickým účelům, zahrnuje mimo jiné lexémy potenciální.

Bohemistická slootovorná tradice navazuje zejména na teorii odvozování slov M. Dokulila (v prohloubené a doplněné podobě je obsažena v 1. díle *Mluvnice češtiny*, tzv. akademické, Dokulil a kol., 1986). V naší práci jsme vycházeli zejména z navazujících, ale novějších mluvnických příruček. Mezi těmi, z nichž jsme čerpali, uveďme alespoň následující; každá z nich reprezentuje jiný typ. Jsou to *Velká akademická gramatika spisovné češtiny* F. Štíchy a kol. (2018), *Slovník afixů užívaných v češtině* (ed. J. Šimandl, 2017) a průběžně aktualizovaná vysokoškolská učebnice R. Adama a kol. *Gramatické rozborý češtiny* (2023). Všechny uvedené příručky sice poskytují různé obsáhlá pojednání k jednotlivým slootovorným jevům, popř. k jednotlivým slootovorným prvkům, avšak konkrétní slova se v nich z pochopitelných důvodů uvádějí pouze výběrově, tj. jen jako ilustrační příklady k výkladům. V připravovaném lexikografickém slootovorném nástroji je perspektiva opačná, východiskem jsou konkrétní lexémy. Také vzhledem k tomu lze předpokládat využití nástroje v pedagogické praxi.

Ve výuce češtiny se tvoření slov věnuje pozornost na základních i středních školách. Všeobecně se však toto téma považuje za látku obtížnou, a to mnohdy i pro samotné vyučující. Zvláště v případě jazykově nadaných žáků však může tato látka velkou měrou rozvíjet jejich abstraktní myšlení. „Při vhodném didaktickém uchopení lze na ní budovat a rozvíjet zejména tři z poznávacích a řešitelských postupů: analýzu, analogii a hierarchizaci.“ (Kostečka, 2019) Slootovorbě i jejímu využití ve školní praxi bylo věnováno také samostatné číslo *Didaktických studií* (2, 2023).

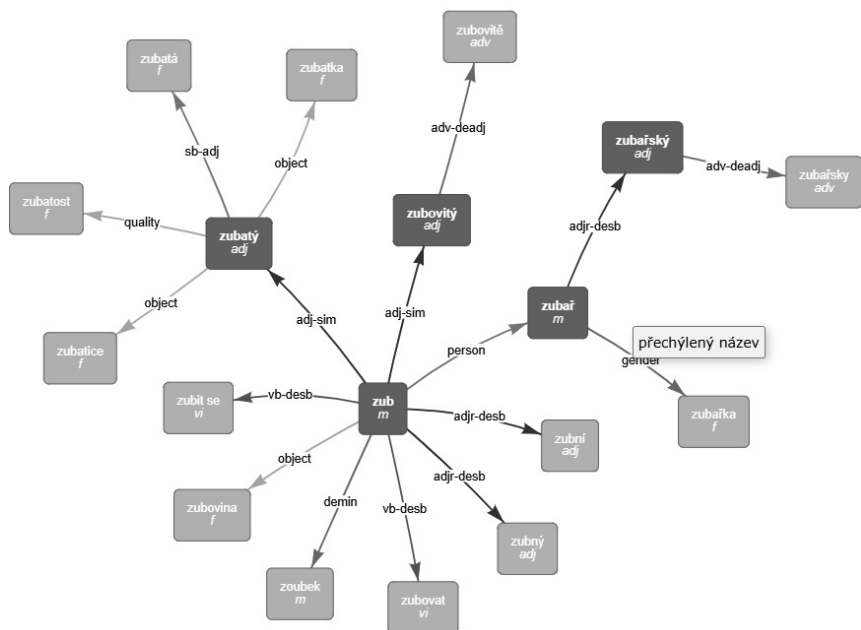
V slootovorném nástroji ISSČ jsou využita zejména slovníková slootovorná hnízda (díky nim lze maximalizovat vloženou lexikografickou práci). V ISSČ se totiž společně se základovými podstatnými jmény systematicky registrují přechýlené názvy, zdrobněliny, vztahová adjektiva ap., s adjektivními základy se v hnízdech sdružují derivovaná příslovce a názvy vlastností na *-ost*, se slovesnými základy verbální substantiva na *-ní*, *-tí*, pasivní verbální adjektiva na *-ný*, *-tý*, procesuální adjektiva na *-cí*, účelová i další adjektiva derivovaná ze slovesných základů. Polo-

automatizovaně a zčásti manuálně je zpracována zbývající relevantní část hesel. Slovo tvornými hranami jsou tedy na základě zadaných regulárních výrazů propojeny další lexémy, které bylo možné identifikovat jako dvojice slov základových a odvozených (např. z názvů osob jsou ke slovesným základům připojeny činitelské názvy, k substantivním základům konatelské názvy a obyvatelská jména, k adjektivním základům nositelé vlastnosti ap.). Koncové řetězce, kterými se tyto dvojice lišily, zčásti odpovídaly příponám, zčásti – zejména kvůli alternacím – zahrnovaly i část slovo tvorného základu.

Z celkového objemu ISSČ (cca 70 000 lexémů) bylo poloautomatickými metodami vytvořeno cca 30 000 hran, které odpovídají prioritně vztahu odvození (mezi slovem základovým a slovem odvozeným); registrována je většina příponových hran. V aktuální verzi dat jsou jako základová slova zpracována podstatná jména, přídavná jména a slovesa, a dokonce i číslovky. Se základovým slovem jsou spojena i slova, při jejichž odvozování dochází k hláskovým alternacím, včetně vznikových a zánikových (např. *dar* – *dárek*; *čtverec* – *čtvercový*; *hroch* – *hroší*).

(Polo)automatizovaně se podařilo podchytit relativně spolehlivě propojení základového a odvozeného slova po stránce formální (fundaci), při manuální kontrole se však zvažují i vztahy významové (motivace), které jsou často komplikovanější. Na pozadí slovo tvorného nástroje tedy registrujeme i vztahy polymotivace, např. *zvoník* – jednak k *zvonit* (činitelský název), jednak ke *zvon* (konatelský název). Polymotivaci je třeba zvažovat rovněž u slov, jejichž morfematické složení zahrnuje jak prefixy (popř. prefixoidy nebo radixoidy), tak sufixy (resp. sufixoidy). Např. názvy osob (odborníků) se zpravidla považují za deriváty k názvům oborů – tj. výraz *hydrobiolog* je utvořen desufixací k základovému slovu *hydrobiologie*. Možná je však i interpretace, že slovo bylo vytvořeno předřazením komponentu *hydro-* před slovní základ *-biolog*. Manuálně jsme se zabývali rovněž homonymií, jednak homonymií lexikální, jednak homonymií přípon (např. přípony *-ka*, morfému pro přechylování, zdrobňování, tvoření názvů prostředků i dějových substantiv).

Na pozadí slovo tvorného modulu, v lexikální databázi, jsou slovo tvorné vztahy podchyceny textově (součástí je i sémantická značka derivačního vztahu, v řadě případů včetně polymotivačních vztahů), pro vnější uživatele jsou však prezentovány graficky. Pro veřejnou prezentaci je rozsah rozlišujících sémantických značek omezen a zobecněn do několika nejvýraznějších skupin (např. názvy osob, v jejich rámci přechýlené názvy, dále zdrobňelé názvy, verbální substantiva ap.). Postupně jsou derivační vztahy sémanticky (re)anotovány, resp. některé automaticky vytvořené spoje jsou ručně revidovány.



Slovotvorný modul je začleněn do webového rozhraní slovníku společně s dalšími nástroji a funkcionalitami, jejichž cílem je přispět k vytvoření moderního, studijně orientovaného výkladového slovníku současné češtiny a zároveň slovníku multi-mediálního. Zapojeny jsou nahrávky výslovnosti heslových slov, dva syntetické modely hlasů (mužský i ženský) k ozvučení výkladů významů a exemplifikační části heslové stati. Obrazová složka je prezentována jednak ve formě obrázkového slovníku (v samostatné záložce), jednak v heslových statích příslušných heslových slov. Inovativní je rovněž zapojení značného množství poznámek, z nichž podstatná část vychází z anglosaské tradice tzv. usage notes („poznámek k užití“), další poznámky mají ráz etymologický.

Hlavní motivací pro rozšiřování ISSČ o další funkce a nástroje je snaha postupně proměnit uživatelskou praxi týkající se běžně využívaných jazykových příruček, a to směrem od pravopisných ke slovníkovým, tj. takovým, které poskytují nejen informaci o grafice slova, ale zejména o jeho významu, způsobu a vhodné situaci užití.³

³ Slovotvorný nástroj je přístupný na portálu *Nechybujte.cz*, v záložce *Slovník současné češtiny*. U slov, která jsou v této fázi přípravy zapojena do slovotvorné sítě, se v levé části obrazovky zobrazuje

Literatura

- ADAM, R. a kol. (2023): *Gramatické rozbory češtiny. Výklad a cvičení s řešeními*. Praha: Karolinum.
- DeriNet. *Lexical Network of Word-Formation Relations in Czech*. Dostupné z: <https://ufal.mff.cuni.cz/derinet>. [cit. 2024-1-28]
- DOKULIL, M. a kol. (1986): *Mluvnice češtiny 1. Fonetika, fonologie, morfonologie a morfeematika, tvoření slov*. Praha: Academia.
- KOSTEČKA, J. (2019): K výuce tvoření slov na středních školách. Asociace češtinářů. Dostupné z: <https://www.ascestinaru.cz/jiri-kostecka-k-vyuce-tvoreni-slov-na-strednich-skolach/>. [cit. 2024-1-28]
- Slovník spisovné češtiny pro školu a veřejnost* (1978; 2., opr. a dopl. vyd. 1994; 3., opr. vyd. 2003). Praha: Academia.
- ŠEVČÍKOVÁ, M.; ŽABOKRTSKÝ, Z.; VIDRA, J.; STRAKA, M. (2016): Lexikální síť DeriNet: elektronický zdroj pro výzkum derivace v češtině. *Časopis pro moderní filologii*. Ročník 98, s. 62–76.
- ŠIMANDL, J. (ed.) (2018): *Slovník afixů užívaných v češtině* [online]. Ústav pro jazyk český AV ČR, v. v. i. Dostupné z: <http://www.slovníkafixu.cz/>. [cit. 2024-1-28]
- ŠTÍCHA, F. a kol. (2018): *Velká akademická gramatika spisovné češtiny 1. Morfologie: Druhy slov / Tvoření slov*. Praha: Academia.

PhDr. Pavla Kochová

Lingea, s. r. o.

pavlak@lingea.net

ORCID: 0000-0002-0348-5715

tlačítko *slovo*tvorba, na které je třeba pro vizualizaci slovo

tvorných vztahů kliknout. Neveřejnou částí tohoto nástroje je funkce, díky které mohou lexikografové a spolupracující osoby postupně vyznačovat další, zatím neregistrovaná spojení, a to zejména v případech, kdy se deriváty z hlediska podoby morfémů výrazně odlišují a kvůli tomu nebyly dosud zahrnuty do (polo)automatizovaného propojování.