

Možnosti korpusů psané češtiny při zkoumání deklinace oikonym

Jana Marie Tušková

Katedra českého jazyka, PdF MU, Poříčí 7/11, 603 00 Brno

Abstract

Possibilities of written Czech corpora when describing declension paradigms of oikonyms

The contribution focuses on the possibilities and limitations of the SYN2005 corpus, and in some cases the SYN2000 corpus, for describing declension paradigms of Czech oikonyms. Attention is given to specific features of oikonym search such as the existence of oikonym forms homonymous with other classes of proper names as a result of the process known as transonymization. The paper analyses which instances of Czech oikonym declension are sufficiently represented in the corpus, and which occur only peripherally. Findings are presented on the basis of complex description of the declension of Czech feminine oikonyms.

Na materiálu synchronních korpusů psané češtiny, zejména SYN2000 a SYN2005, vznikla již řada odborných studií zabývajících se apelativní vrstvou jazyka. Zkoumání jevů propriální vrstvy jazyka stálo však zatím na okraji zájmu lingvistů využívajících možností korpusového materiálu. Svůj příspěvek bych proto chtěla věnovat problematice proprií a zhodnotit míru využití korpusů SYN2005 a SYN2000 pro výzkum deklinace oikonym.

Předmětem mého pozorování byla femininní oikonyma češtiny, tzn. vlastní jména pojmenovávající osídlené objekty, jež se vyskytují na území České republiky bez ohledu na jejich původ. Ze skupiny oikonym byly vyděleny pouze názvy ulic, náměstí, nábřeží, mostů a názvy tzv. staveb nesídlitních, tj. kostelů, božích muk, pomníků, mohyl atd.¹ pro specifičnost jejich formálních struktur. Převážná část dokladů oikonym z korpusů SYN2005 a SYN2000 se váže k názvům osad (měst, vesnic), městských čtvrtí, sídlišť či předměstí. Při vytváření základních souborů oikonym určitého deklinačního typu jsem obvykle pracovala

¹ Je-li východiskem klasifikace Šmilauerova (Šmilauer 1973), byla tedy hodnocena kromě uvedených dvou skupin jmen deklinace všech ostatních souborů tzv. sídlitních jmen (místních), tj. v novější terminologii oikonym, která zahrnují: 1. osady (města, vesnice), čtvrti, sídliště, předměstí; 2. místní části – skupiny domů, samoty, myslivny, hájovny, hrady, obydlené zříceniny atd.; 3. stavby v osadách – věže, zámky apod.

s programem Paradigma,² který je k dispozici na internetových stránkách Českého národního korpusu³ a slouží primárně k vyhledávání lemmat, jež patří k zadanému paradigmatu. Pro účely popisu paradigmat oikonym řazených k jednotlivým deklinačním typům byly využity možnosti korpusového manažeru Bonito. Ten je, jak je známo, uzpůsoben pro kladení jednoduchých, ale i kombinovaných dotazů, přičemž vyhledávky kombinovaných dotazů obsahujících tagy se jeví jako méně přesné. Chybovost značkování (tagování), jak uvádí publikace Statistiky češtiny (Bartoň et al. 2009), se v korpusu SYN2005 pohybuje v průměru kolem 4 %, chybovost v oblasti lemmatizace je o něco nižší,⁴ a proto bylo také spolehlivější vyhledávat oikonyma na základě lemmat než tagů.

Při tvorbě souborů oikonym určitého deklinačního typu a popisu jejich paradigmat jsem se setkala s určitými specifiky typickými pro oikonyma, v některých případech i pro další třídy proprií, která uvádím na příkladech:

1. K vytvoření souboru femininních oikonym češtiny, např. zakončených na *-a* (typ *Ostrava*), byl využit dotaz v programu Paradigma:

```
word="[AÁBCČDEÉFGHIÍJKLMNOÓPRŘSŠTUÚVWZZ].*" lemma=".*a"tag="NN.*",
```

který vyhledal všechna substantiva zakončená v základním tvaru (lemmatu) na *-a* a začínající alespoň v některém svém tvaru velkým písmenem. Dotaz byl záměrně zadán široce, tj. bez další specifikace tagu na ženský rod a singulár, aby vznikl seznam co nejobsáhlejší, a eliminovaly se tak v co největší míře případné chyby v morfologickém značkování. Ze vzniklého seznamu byla manuálně vybrána lemmata domácích⁵ oikonym. Vytvořený soubor byl pak konfrontován s příklady oikonym uváděnými v českých mluvnicích⁶ a v publikaci

² Charakteristiku a některé možnosti využití programu Paradigma uvádí autor programu V. Cvrček v článku Morfologická paradigmatica a korpus (2008).

³ http://ucnk.ff.cuni.cz/hledat_v_cnk.php.

⁴ Podle Mgr. V. Cvrčka, Ph.D., z Ústavu Českého národního korpusu v Praze, nejsou chyby v tagování a lemmatizaci rozprostřeny napříč korpusem rovnoměrně na všechna slova a jejich tvary, ale soustředí se zejména na oblast homonymie.

⁵ Korpus SYN2005 dokládá také cizí oikonyma zakončená na *-a*. Prvními deseti nejfrekventovanějšími byla jména: *Moskva, Bratislava, Varšava/Warszawa, Barcelona, Ženeva, Atlanta, Káhira, Havana, Ankara, Sevilla*. Jsou-li cizí oikonyma v českém kontextu počestěna (*Warszawa* → *Varšava*, *Genève* → *Ženeva* atd.), stávají se tzv. exonymy.

⁶ Mluvnice česká pro školy střední a ústavy učitelské I (Gebauer 1895), Mluvnice česká pro školy střední a ústavy učitelské I (Gebauer, Ertl 1926), Mluvnice spisovné češtiny I (Trávníček 1951), Nauka o českém jazyku (Šmilauer 1972), Česká mluvnice (Havránek, Jedlička 1981), Mluvnice češtiny 2 (Komárek et al. 1986) a Příruční mluvnice češtiny (Karlík et al. 2003).

Naše místní jména a jak jich užívat (Polívková 2007), a to především z toho důvodu, že v něm chyběla část oikonym vzniklých z apelativ. Chybějící oikonyma byla dohledávána / ověřována jednoduchým dotazem typu: [lemma=" "], např. oikonyma *Bučina*, *Horní Bříza*, *Bílá Voda* atd., která mají v korpusu SYN2005 apelativní lemmata „bučina, bříza, voda“. S obdobným jevem je třeba počítat – ve větší či menší míře – také u některých dalších deapelativních oikonym patřících k jiným deklinačním typům, např. oikonymum *Strážnice* má lemma „strážnice, u oikonymuma *Lnáře* je uvedeno lemma „lnář“, oikonymum *Trosky* má lemma „troska“ atd. Jak je patrné, lemmata řady oikonym jsou uváděna v singuláru, bez ohledu na gramatické číslo jmen.⁷ V případě ženských deapelativních oikonym s adjektivní deklinací se zase často vyskytují apelativní lemmata v mužském rodě, např. *Trnová* má lemma „trnový“. Odstranění uvedených nedostatků v lemmatizaci by podstatně přispělo ke zkvalitnění korpusů a zároveň by zjednodušilo práci lingvistů. Přes výše uvedená fakta, tzn. i při současném stavu lemmatizace, vyhledá však program Paradigma u každého deklinačního typu ve sledovaných korpusech takový počet oikonym, který ho zpravidla v dostatečné míře reprezentuje z morfologického hlediska a zachycuje v jeho rámci vývojové tendence při případné volbě dubletních či variantních koncovek.⁸

2. Pro popis paradigmat jednotlivých oikonym a zhodnocení podoby skloňování celých deklinačních typů byla oikonyma vyhledávána prostřednictvím korpusového manažeru Bonito dotazem typu [lemma=„ ... “]. Ve vzniklých konkordančních seznamech hledaných oikonym jsem pracovala s několika nabízenými informacemi. Nejrychleji zjistitelný byl údaj o tzv. absolutní frekvenci oikonyma. Při práci s materiálem psaných korpusů jsem došla k závěru, že tento údaj odpovídající frekvenci lemmatu propria zahrnuje často kromě frekvence oikonyma současně také hodnoty frekvence jiných tříd proprií – homonymních s hledaným oikonymem, např. antroponym, anoikonym nebo chrématonym (oikonymum *Světlá* → antroponymum *Světlá*, oikonymum *Opava* → antroponymum *Opava*; oikonymum

⁷ Oikonyma, na rozdíl od apelativ, existují zpravidla jen jako singularia tantum (*Praha*), nebo pluralia tantum (*Vinohrady*).

⁸ A. Polívková (2007) při svém popisu deklinace místních jmen vychází rovněž z určitého vytvořeného souboru jmen, který obsahuje celkem 5542 místních jmen. Základním požadavkem pro jejich výběr byla význačnost města či obce, a to nejen rozlohou a počtem obyvatel, ale i správním, hospodářským a kulturně historickým významem. Na příkladě uvádíme počet oikonym jedné deklinační třídy, kterou zahrnuje publikace A. Polívkové i korpus SYN2005, tj. singulárová oikonyma s adjektivní deklinací zakončená na *-á*: Polívková (celkem 238 oikonym) – SYN2005 (147 oikonym s AF 5 a výše).

Olomouc → anoikonimum „okres“ *Olomouc*; oikonimum *Olomouc* → chrématonymum – „sportovní klub“ *DHK Olomouc*, *1. HFK Olomouc*, *Lokomotiva Olomouc* apod.⁹).

Knihovny v obcích Grygov < (okres *Olomouc*) > a *Radonice* (SYN2005, POP)
Početná výprava z interligového klubu < *DHK Olomouc* > se účastnila letošního mistrovství (SYN2005, PUB)

Existence homonymních názvů je důsledkem tzv. transonymizace, která je tradičně chápána jako vztah proprium → proprium, tj. vznik propria na základě mezipropriálně fundujícího vztahu, např. oikonimum *Praha* → chrématonymum „kino“ *Praha*, viz Encyklopedický slovník češtiny (Karlík et al. 2002). Transonymizaci chápeme jak v užším, tradičním pojetí, viz výše uvedený příklad,¹⁰ tak i v širším pojetí, viz Tušková (v tisku), kdy vznikají víceslovná transonyma, jejichž součástí je proprium formálně homonymní s původním jednoslovným propriem, např. oikonimum *Olomouc* → chrématonymum „sportovní klub“ *DHK Olomouc*. V případě některých transonymizovaných chrématonym, zejména jedná-li se o názvy sportovních klubů, dochází při komunikaci k univerbizaci víceslovného oficiálního názvu. Výsledná komunikační univerbizovaná podoba je často variabilní, může obsahovat různé části oficiálního víceslovného chrématonyma. Např. hokejový klub *Kometa Brno* → *Kometa*, nebo *Brno*.

Hokejisté < *Kometry Brno* > ve veřejším dramatickém zápase porazili celek *Prostějova* (SYN2005, PUB)
Rosice ještě < *Kometu* > na jejím ledě neporazily. (SYN2005, PUB)
Doživotní distanc hrozí Jiřímu Titzovi (51), bývalému ligovému hokejistovi < *Brna* > a *Sparty* (SYN2005, PUB)

Vzhledem k častým případům transonymizace oikonimum byla rozlišována dvojí hodnota frekvence – zavedený termín celková absolutní frekvence (CAF), tj. frekvence lemmatu propria, a běžně užívaný termín absolutní frekvence (AF), kterým byla označována jen frekvence oikonimum, tzn. bez hodnot frekvence případných homonymních názvů řazených do jiných propriálních tříd. Hodnoty absolutní frekvence (AF) byly získány manuální kontrolou a vytříděním vyhledaných konkordančních seznamů proprií, a to u oikonimum s CAF 100 a nižší a v případě dubletních či tripletních tvarů u všech oikonimum. Zjišťování absolutní frekvence (AF) oikonimum, byť zatím touto časově náročnou formou, přineslo nejen objektivní statistické

⁹ *DHK Olomouc* – házenkářský klub, *1. HFK Olomouc* – fotbalový klub, *Lokomotiva Olomouc* – ragbyový klub.

¹⁰ Při transonymizaci se nemění forma propria, ale dochází k proměně propriálně druhových příznaků, které zvolený propriální lexém přiřazuje k určité třídě proprií (Šrámek 2007).

údaje vztažené zejména k materiálu korpusu SYN2005, ale i obecnější závěry týkající se distribuce tvarů oikonym a tvarů chrématonym vzniklých jejich transonymizací. Například o distribuci tvarů chrématonym vzniklých transonymizací se dá obecně říci, že v dokladech korpusu SYN2005 užívají místo dublety, která se vyskytuje u oikonym, zpravidla již jen jednu koncovku, a to historicky mladší, např. v Gsg. typu *Příbram* volí novější koncovku *-i*.

Kdyby trefil do < Příbramě >, hned by se tam rozběhl. (SYN2005, NOV)

Staročeskou masopustní tradici si tento týden připomenou obyvatelé < Příbrami > a Struhařova na Benešovsku. (SYN2005, PUB)

Znova se o tom přesvědčili hráči < Příbrami > na ostravských Bazalech (SYN2005, PUB)

Už v sobotu hostí volejbalisté Dukly tým < Příbrami > (SYN2005, PUB)

Pokud by korpusový manažer dokázal rozpoznávat jednotlivé třídy proprií, a to i v případě transonymizovaných proprií, velkou měrou by urychlil a zkvalitnil výzkumnou práci.

S patřičnou obezřetností je nutné přistupovat k vyhodnocování velmi nízkých výskytů. Na základě práce s psanými korpusy za ně považují doklady oikonym do absolutní frekvence (AF) 5, které zpravidla nelze považovat za relevantní pro objektivní závěry týkající se případné distribuce variantních či dubletních tvarů. Korpusy SYN2000 a SYN2005 však u mnohých oikonym právě doklady s tak nízkou frekvencí poskytují, např. SYN2005 u řady dubletních tvarů Gsg. oikonym typu *Příbram: Libčevsi (1)/Libčevse (1), Kněževsi (1)/Kněževse (3), Bedihošti (1)/Bedihoště (1)*. V případě nulových výskytů je třeba počítat také s tím, že nulová frekvence nemusí vždy znamenat, že gramatický tvar oikonyma neexistuje, spíše nebývá v korpusu z různých důvodů doložen, např. SYN2005: *Velké Černoce (0)/Velké Černoci (2)*. Na výše popsany stav má vliv několik faktorů – je dán samotnou **propriální povahou zkoumaného materiálu** (v komunikaci jsou propria ve srovnání s apelativy zastoupena podstatně méně), ovlivňuje ho také **geografická velikost a aktuální společenská významnost oikonymického objektu** (nízké či nulové frekvence lze očekávat u oikonymických názvů menších a společensky méně významných obcí či měst). Významnou měrou se na výskytu tvarů oikonym podílí jejich **pád** (nulovou či velmi nízkou frekvenci mají zejména vokativní tvary oikonym,¹¹ velmi nízkou frekvenci má také dativ, a to jak u singulárových, tak u plurálových oikonym). Vliv na výskyt oikonym a jejich tvarů má

¹¹ Tvary vokativu se v korpusu SYN2005 vyskytly jen u vysoce frekventovaných oikonym, např. *Praha* a *Brno*. V případě oikonyma *Praha* je doloženo celkem 19 výskytů pocházejících z beletrie nebo publicistiky, okrajově i populárně-naučné literatury, u *Brna* 2 doklady, které pocházejí z beletrie. Jedná se o vokativy ve funkci poetizační personifikace či jen personifikace.

bezesporu také **skladba korpusových textů**. Obecně se dá říci, že tvary oikonym dokládají např. v korpusu SYN2005 převážně texty publicistické, podstatně méně pak beletrie, naučná či populárně naučná literatura, a to i přes žánrové složení korpusu SYN2005 (33 % publicistika, 40 % beletrie, 27 % odborná literatura).

Na pódiu vystoupí různé hudební skupiny z < Českých Budějovic >. (SYN2005, PUB)

A přece se staly < Pardubice > zastávkou na mé cestě (SYN2005, NOV)

Goethe po prvním setkání s Beethovenem v < Teplicích > poznamenal (SYN2005, ENC)

Lesk salonu zeslábl odchodem Němcové do < Domažlic > a manželů Čelakovských (SYN2005, POP)

Domnívám se, že by pro účely morfologického výzkumu oikonym bylo vhodnější, aby korpusy obsahovaly jednak vyšší procento regionálních textů, ať již z publicistiky, administrativy či jiných žánrů, případně více odborné či populárně-naučné literatury tematicky orientované na oblast oikonym či šířeji toponym. Tento požadavek však tematicky široké korpusy SYN2000 a SYN2005 nemohou splnit, a ani to není jejich cílem. Spíše by mohl v budoucnu vzniknout korpus specializovaný přímo na výzkum proprií.

Závěrem lze říci, že korpusy SYN2005 a SYN2000 poskytují informace o současném psaném úzu češtiny včetně oblasti zahrnující deklinaci femininních oikonym češtiny. Dobře jsou zde zachyceny centrální jevy této deklinace a působící vývojové tendence. Převážná část femininních oikonym češtiny je doložena v kodifikovaných tvarech, nekodifikované podoby zaujímají periferní postavení, např. tvary na *-ma* v Ipl. pomnožných oikonym. Okrajově jsou doloženy jevy typické především pro mluvený jazyk, např. v korpusu SYN2005 kolísání v gramatickém rodě jmen typu *Aš, Bystrc, Zliv* apod. Přestože je rozsah obou korpusů značný, nemohou obsahovat a ani neobsahují všechna oikonyma češtiny. Při zamyšlení, ve kterém druhu textů se tvary oikonym nejběžněji a v největším rozsahu vyskytují, se dá konstatovat, že jsou to především texty publicistické, které oba korpusy poskytují v dostatečné míře. Možnosti, které korpusy nabízejí, nespočívají tedy v ověření tvarové soustavy každého jednotlivého oikonyma, ale v tom, že umožňují popsat chování frekventovaných jmen, tzn. názvů měst, městských čtvrtí, sídlišť a některých obcí, případně názvů hradů, zámků, omezeně pak jiných skupin oikonym. Domníváme se, že na zkoumaném materiálu je možné prezentovat určitý model deklinace oikonym, stanovit jednotlivé deklinační typy či podtypy a poměrně přesně zachytit současné vývojové tendence v deklinaci femininních oikonym češtiny, viz Tušková (2010). Pro účely detailnějšího zpracování deklinace oikonym všech rodů by však bylo třeba v budoucnu vytvořit specifický korpus psané češtiny podle výše

uvedených návrhů a rozšířit korpusy mluvené češtiny do té míry, aby mohly být rovněž využity pro výzkum deklinace proprií.

Literatura

- Bartoň T., V. Cvrček, F. Čermák, T. Jelínek, V. Petkevič, 2009, *Statistiky češtiny*. Nakladatelství lidové noviny, Praha.
- Cvrček V., 2008, Morfologická paradigmatica a korpus. In *Gramatika a korpus. Grammar & Corpora. Sborník příspěvků ze stejnojmenné konference. 25.–27. 9. 2007, Liblice*. eds F. Štícha, M. Fried, Academia, Praha, 151–159.
- Encyklopedický slovník češtiny*, 2002. Eds P. Karlík, M. Nekula, J. Pleskalová. Nakladatelství lidové noviny, Praha.
- Gebauer J., 1985, *Mluvnice česká pro školy střední a ústavy učitelské I. Nauka o slově*. Praha–Viedeň.
- Gebauer J., V. Ertl, 1926, *Mluvnice česká pro školy střední a ústavy učitelské I*. Nákladem Grafické unie a. s., Praha.
- Havránek B., A. Jedlička, 1981, *Česká mluvnice*. SPN, Praha.
- Mluvnice češtiny 2 – Tvarosloví*, 1986. Red. M. Komárek, J. Kořenský, J. Petr, J. Veselková. Academia, Praha.
- Polívková A., 2007, *Naše místní jména a jak jich užívat*. Euromedia Group, Praha.
- Příruční mluvnice češtiny*, 2003. Eds P. Karlík, M. Nekula, Z. Rusínová. Nakladatelství Lidové noviny, Praha.
- Šmilauer V., 1972, *Nauka o českém jazyku*. SPN, Praha.
- Šmilauer V., 1973, Třídění vlastních jmen. In *Zbornik pedagogickej fakulty v Prešove Univerzity P. J. Šafárika v Košiciach. V. zasedanie Medzinárodnej komisie pre slovanskú onomastiku a V. slovenská onomastická konferencia (Prešov 3.–7. mája 1972)*, zv. 3, *Slavistika*, eds M. Blichá, M. Majtán, Slovenské pedagogické nakladateľstvo, Bratislava, 109–111.
- Šrámek R., 2007, Transonymizace v propriální nominaci. In *Folia onomastica Croatica*, red. D. Brozović Rončević, Hrvatska akademija znanosti i umjetnosti. Razred za filološke znanosti, Zagreb, s. 499–508.
- Trávníček F., 1951, *Mluvnice spisovné češtiny. Část I. – Hláskosloví, tvoření slov, tvarosloví*. Slovanské nakladatelství, Praha.

Tušková J. M., 2010, *Deklinační systém femininních oikonym v češtině. Synchronní pohled na základě Českého národního korpusu*. Habilitační práce na Filozofické fakultě Masarykovy univerzity, Brno.

Tušková J. M., v tisku, *Transonymizace a chrématonyma*. Olsztyn.

Český národní korpus – SYN2000, 2000. Ústav Českého národního korpusu FF UK, Praha.

Dostupný z WWW: <<http://ucnk.ff.cuni.cz>>.

Český národní korpus – SYN2005, 2005. Ústav Českého národního korpusu FF UK, Praha.

Dostupný z WWW: <<http://ucnk.ff.cuni.cz>>.