ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/374632175

Prolegomena ke kvantitativní onomastice

Article *in* Acta onomastica · October 2023

DOI: 10.58756/a2641351

CITATIONS
3

READS
33

2 authors:

Jaroslav David
University of Ostrava
56 PUBLICATIONS **139** CITATIONS
SEE PROFILE

Michal Místecký
University of Ostrava
43 PUBLICATIONS **91** CITATIONS
SEE PROFILE

All content following this page was uploaded by Michal Místecký on 12 October 2023.

The user has requested enhancement of the downloaded file.

Jaroslav David (Ostrava) – Michal Místecký (Ostrava)

PROLEGOMENA KE KVANTITATIVNÍ ONOMASTICE

PROLEGOMENA TO QUANTITATIVE ONOMASTICS

The aim of this paper is to present the potential of quantitative methods for the research of proper names (mostly personal and place names) as units operating in Czech texts. Three case studies treat particular research topics on the background of the development of Czech onomastics in the 20th century and in the first decades of the 21st century. They are aimed at 1) quantification of relational (semantic) patterns; 2) collocation analysis; 3) morphological categories of proper names. The text also provides an overview of previous attempts to apply quantitative and corpus approaches to proper name research.

Key words

onomastics; quantitative methods; corpus; relational (semantic) patterns; collocations; morphology

Klíčová slova

onomastika; kvantitativní metody; korpus; vztahové modely; kolokace; morfologie

Cíle studie

Cílem této studie je představit a konkretizovat tři oblasti, ve kterých se podle našeho názoru nejvíce nabízí experimentální přístup k vlastním jménům: 1) kvantifikace vztahových modelů; 2) kolokační analýza; 3) morfologické kategorie vlastních jmen. Jsme si vědomi toho, že propria se už stala součástí šíře pojatých kvantitativněLingvistických nebo pomezních výzkumů (Cvrček a kol., 2020; Collins a kol., 2022); zde se proto pokusíme prozkoumat jiné, dosud méně probádané sféry. Jestliže používáme slovo „experiment“, pak ho chápeme ve vědeckém slova smyslu, tj. jako kontrolovaný pokus, který vede k prověření určité hypotézy. Čerpáme tak z pojetí lingvistiky jako experimentální vědy, které načrtl jazykovědný vizionář Luděk Hřebíček (Hřebíček, 2002).

Úvod – geneze a stav aplikace kvantitativních přístupů v onomastickém výzkumu

Dnešní kvantitativní lingvistika je souhrnné označení pro mnoho různě propojených proudů, které se v poslední době slévají do jednoho řečiště (viz např. Uhlířová, 2017). Počátky početního přístupu k jazykovým jevům lze vystopovat zřejmě až do období vědeckého pozitivismu (Schreibman – Siemens, 2008; Sedlačíková, 2013) a kvantifikaci se nebránil ani strukturalismus, i když jeho vztah k parolově pojaté lingvistice byl poměrně složitý (Čech, 2005). Nepřerušené kvantitativní

bádání iniciovala především vědeckotechnická revoluce šedesátých let 20. století (Richta a kol., 1966), která stimulovala komputerizaci potřebnou pro první zpracování velkého objemu dat (Sgall – Hajičová – Piťha, 1982). Technologický růst s sebou přinesl další významný mezník, kterým byl vznik Ústavu Českého národního korpusu v roce 1994, na nějž postupně navazovala implementace textových databází do jazykovědného výzkumu. Matematizace se však realizovala i jinými cestami – prostřednictvím kvantitativní pedagogiky (např. Pluskal, 1989), využíváním kvantitativnělingvistických nástrojů v oblasti medicíny (Yoder, 2006), rozvojem exaktní metodologie v literární vědě (Petrů, 1968; Levý, 1971; Plecháč – Kolár, 2017; Změlík, 2017, 2018; srov. též Hernández-Lorenzo, 2022) a köhlerovsko-altmannovským synergetickým přístupem k jazyku, který se snaží, v návaznosti na průkopnické dílo George Kingsleyho Zipfa (Zipf, 1949), vypracovat pro jazyk teoretickou základnu (Köhler, 1986).

Rostoucí kvantifikace lingvistiky se česká onomastika dosud příliš neúčastnila, protože musela obstát v boji z hlediska její existence důležitějším – z pomocné vědy historické a disciplíny, která se nacházela na pomezí několika oborů, se chystala přerodit v plnohodnotný jazykovědný obor s funkční organizační a metodologickou základnou (srov. Šrámek, 2007). Dějiny lingvistiky si tak do určité míry odžívala se zpožděním, narychlo a částečně i izolovaně od aktuálních trendů jazykovědného bádání: do osmdesátých let 20. století si procházela svou historicko-komparatistickou (etymologickou) fází, na přelomu milénia rozvinula strukturalistická východiska reprezentovaná modelovou teorií (Pleskalová, 1992; Šrámek, 1999) a v desátých letech 21. století jsme svědky jejího pragmatického obratu. Rozvíjí se socioonomastika (Knappová, 1989; viz též Bauko, 2011), literární onomastika (Dvořáková, 2017) a pozornost se postupně upíná k jazykové krajině a společenským a politickým souvislostem vlastních jmen (David, 2011; David – Mácha, 2014; Lábus – Vrbík, 2018; Ptáčníková, 2021; Jordan – Mácha a kol., 2021; viz též David – Klemensová – Místecký a kol., 2022). Pokud se tedy dnes zasazujeme o pěstování onomastiky orientované také kvantitativně, jsme si vědomi toho, že ještě více zrychlujeme vývoj, kterým prochází; věříme však, že právě takový tavicí kotel přístupů nejlépe ukáže, které metodologické rámce jsou pro tuto relativně mladou jazykovědnou disciplínu přínosné.

Zaměříme-li se na výzkum proprií a aplikaci kvantitativních metod a korpusových nástrojů v současné lingvistice, zjistíme nejen to, že vlastní jména nepředstavují pro tyto disciplíny (přístupy) zásadní téma, ale také to, že onomastika neposkytuje – až na určité výjimky – pro tyto aplikace dostatečný prostor. Analýza gramatických jevů v rozsáhlých textech, která kombinuje kvalitativní a kvantitativní metody, je v rámci současné lingvistiky vnímána jako zcela standardní a obvyklá (srov. např. Cvrček – Kováříková, 2011; Gries, 2013, s. 3–7; Stefanowitsch, 2020). Při obhajobě kvantitativního přístupu lze vyzdvihnout především jeho zaměření na empirická data a intersubjektivní charakter výzkumu, kdy testování hypotéz

umožňuje dosáhnout co nejvyšší úrovně vědeckého poznání. V případě onomastiky je však kvantitativní přístup stále ještě minimálně využívanou metodou nejen při popisu morfologie a slovotvorby, stylistických charakteristik nebo konkurenčních variant, ale také například ve výzkumu syntaktického chování nebo sémantiky vlastních jmen (k využití kvantitativní metody, reprezentované především korpusovými analýzami, v češtině a v angličtině přehledovým způsobem viz Tušková, 2019; nejnověji Motschenbacher, 2020).

Toto tvrzení platí obecně nejen pro světovou, ale i pro českou onomastiku, jak je patrné z publikovaných výzkumů v onomastických časopisech *Onoma* (ICOS), *Names* (USA), *Onomastica* (Polsko), *Voprosy onomastiki* (Rusko) a *Acta onomastica* (Česko) v letech 2010–2022. Tendence využívat korpusový a obecně kvantitativní výzkum se výrazněji prosadila až v poslední třetině sledovaného období, a to pouze v časopisech *Names* (např. Nick, 2017; Rambousek – Horák – Parkin, 2018; Motschenbacher, 2020) a *Acta onomastica* (např. Tušková, 2012; David – Davidová Glogarová, 2020; Drkošová – Místecký, 2022). Důvody lze hledat v etymologicko-klasifikačních kořenech onomastiky a v netextově zaměřeném pohledu na propriální materiál. Máme tím na mysli především orientaci na sběr a výzkum proprií jako izolovaných jazykových jednotek, tvorbu soupisů či lexikonů vlastních jmen apod., které v dané fázi vývoje disciplíny vycházely z potřeby vytvořit materiálový základ pro českou onomastiku. Z hlediska efektivního automatického zpracování textu pak považujeme za klíčové potíže, které s sebou nese kategoriální a formální vymezení propria, popř. jeho druhů, zejména v případě transonymizace nebo homonymie (problematika desambiguace, viz Petkevič, 2017); toto vymezení je přitom pro potřeby počítačové analýzy nezbytné (srov. Tse, 2005; Markert – Nissim, 2007; Giger – Štěpán, 2007; Osolsobě – Žižková 2020, 2021; viz také diskusi o podstatě a pravopisu vlastních jmen Štěpán, 2020). S tím souvisejí i další skutečnosti – propriální jazykové jednotky zpravidla nebývají v korpusech anotovány, a jejich detekce a analýza jsou proto komplikované, což je překážkou širšího zájmu o propria v korpusových studiích.

Výjimku v českých korpusech představuje *Pražský závislostní korpus* (*Prague Dependency Treebank*, dále jako PDT), který pochopitelně nemůže konkurovat *Českému národnímu korpusu* co do množství textů a jednotek. V tomto konkrétním případě jsou však diskutabilní i anotační pravidla zvolená autory PDT, která neodpovídají vnímání víceslovných proprií jako komplexních lexikálních jednotek: „Určování začátku a konce názvu je při současném stavu anotace důležité jen pro zařazení identifikačního výrazu do skupiny A nebo B [A – výrazy se skloňovaným řídícím členem, např. *Hudební divadlo v Karlíně*; B – ostatní identifikační výrazy bez skloňovaného řídícího členu, např. název románu *Proti všem* – pozn. aut.]. Do budoucna je tato otázka důležitá pro zavedení atributu, který bude pro každý uzel určovat, zda výraz, který reprezentuje, je, nebo není součástí názvu“ (PDT – Vlastní jména a názvy).

Naopak z pohledu onomastiky velmi nadějný projekt, byť nebyl dále rozvíjen, představuje *Czech Named Entity Corpus* verze 2.0 (viz Straková – Straka – Ševčíková – Žabokrtský, 2017; k vztahu pojmenovaných entit a vlastních jmen viz Ševčíková – Žabokrtský – Krůza, 2007; srov. též Thielen, 1995). Daný korpus obsahuje 8 993 českých vět s ručně anotovanými 35 220 českými pojmenovanými entitami. Souvisejícím výstupem je také nástroj *NameTag* 2, sloužící k rozpoznávání pojmenovaných entit – v textu identifikuje vlastní jména a zařazuje je do předem definovaných kategorií, jako jsou jména osob, míst, organizací atd.

Kvantitativní přístup je v současné české onomastice především zastoupen korpusovými analýzami, které se zaměřují na následující témata:

1) korpus představuje zdroj dat a poskytuje nástroje (určování frekvence, kolokací, kontextu – jde tedy o corpus-based research) pro analýzu vybraných jednotek (např. David – Klemensová, 2019; David – Davidová Glogarová, 2020);
2) vlastní jména jsou tématem korpusového výzkumu zaměřeného utilitárně na jejich detekci a na problémy jejich desambiguace jako jazykových jednotek (např. Giger – Štěpán, 2007; Osolsobě – Žižková, 2020, 2021; srov. též polemický pohled Dvořáková, 2019);
3) popis gramatických charakteristik vlastních jmen, a to i ve světle procesů transonymizace a apelativizace (např. Tušková 2011, 2012, 2019; Tušková – Žižková, 2019);
4) nejnověji se pozornost zaměřuje na stylistické a stylometrické charakteristiky vlastních jmen. Monografie *Registry v češtině* (Cvrček a kol., 2020) je pokusem o návrh kvantitativně založené stylistiky češtiny, která vychází z multidimenzionální analýzy (MDA; podrobněji k použité metodologii viz recenzi Místecký – Kubát, 2021). Z onomastického hlediska se mezi uvažovanými jazykovými rysy objevují toponyma (PROPT) a antroponyma (PROPA). Propriální rysy se prosazují v dimenzi 4 (polytematický vs. monotematický), konkrétně PROPT jako rys polytematičnosti, a v dimenzi 6 (obecný vs. konkrétní), kde se na pólu konkrétnosti uplatňují jak toponyma, tak antroponyma. Další výzkumy ukázaly, že tendence k této konkrétnosti je jedním z důležitých pohybů ve vývoji polistopadové české žurnalistiky (Cvrček, 2022).

V úvodním přehledu současného stavu, kde jsme konstatovali spíše rezervovaný přístup ke kvantitativním metodám ze strany onomastiků, je však nezbytné připomenout i starší vazby mezi onomastikou a kvantitativními přístupy k jazyku. Nelze říci, že by na ně současná kvantitativně pojatá onomastika přímo navazovala – jsou spíše argumentem proti občas zaznívajícím tvrzením, že kvantifikace nemá v onomastickém výzkumu místo. Společným jmenovatelem kvantitativních kořenů české onomastiky je osobnost Vladimíra Šmilauera. Ve vědeckém živo-

topisu zakladatele moderní onomastiky se několikrát objevuje reflexe významu exaktního přístupu k jazyku. Ten je zmiňován nejen v textech jeho spolupracovníků a žáků (Jelínek – Bečka – Těšitelová, 1961, s. 10; Těšitelová, 1965, s. 260; Uhlířová, 1975, s. 133), ale svědčí o něm především jeho vlastní dílo. Kromě Šmilauerovy vedoucí role v práci na frekvenčním slovníku češtiny v didaktickém odboru Kruhu přátel českého jazyka v době 2. světové války (viz Šmejkalová, 2015, s. 378–380; slovník *Frekvence slov, slovních druhů a tvarů v českém jazyce* vyšel až roku 1961) jde především o frekvenční charakteristiky tvarů, pádů nebo slovních druhů v *Novočeské skladbě* (2. vyd. 1966) a *Nauce o českém jazyku* (1972). Z hlediska onomastiky je ovšem průkopnická úvodní studie ke Kopečného *Průvodci našimi jmény* (1974). V ní Šmilauer podává přehled vývoje české antroponymie a ve výkladu systematicky odkazuje na frekvenční charakteristiky osobních jmen (např. pořadí výskytu rodných jmen v konkrétních obdobích, procentuální zastoupení jmen v písňových textech, proměny nejčastěji dávaných jmen).

V následující části chceme na třech konkrétně zaměřených příkladech představit možnosti využití kvantitativních a korpusových metod pro zkoumání toponymie.

Příklad 1 – Kvantifikace vztahových modelů

Aplikace modelové teorie a práce se vztahovými modely (dále jako VM) je v české onomastice po průkopnických textech Jany Pleskalové (1992) a Rudolfa Šrámka (1976, 1999, s. 36–50) standardem, i když své uplatnění našla především v popisu zeměpisných jmen. Na základě počtů názvů přináležejících k jednotlivým modelům lze však statisticky usuzovat i na obecné tendence české toponymie, například na osách město – vesnice či standardizovaná – nestandardizovaná jména.

Zde prezentovaný výzkum se formou sondy zaměřuje na druhý zmíněný kontrast, tedy standardizovaná versus nestandardizovaná toponymie. Standardizovanou toponymií v tomto kontextu rozumíme názvy oficiální a úředně stanovené (většinou jde o jména ulic a náměstí zaštítěné místní autoritou); názvy nestandardizované (živé) byly získány na základě terénního výzkumu – řízených rozhovorů s místními obyvateli. Jako vzorek byly vybrány názvy z katastrů tří sídelních jednotek – měst Hlučín a Vratimov a neměstské obce Řepiště (Moravskoslezský kraj) –, které byly shromážděny v rámci kvalifikačních prací obhájených na Katedře českého jazyka Filozofické fakulty Ostravské univerzity (Dušková, 2021; Neuwirthová, 2020; Vajtrová, 2021). Jsme si vědomi faktu, že studentské práce mohou obsahovat chyby, ale lidský faktor je fenomén, který nelze zcela odstínit v žádném výzkumu. Podobně bereme v potaz skutečnost, že naše výsledky jsou provizorní a závislé na poměrně malém a specifickém vzorku.

Součty jmen obou skupin, které jsou přiřazeny k daným modelům, souhrnně zachycuje tabulka 1.

Tabulka 1: Výsledky vztahověmodelové analýzy vybraných sídel (příklady dle Dušková, 2021; Vajtrová, 2021)

	Standardizovaná jména (absolutní frekvence)	**Nestandardizovaná jména (absolutní frekvence)**
VM A – kde, odkud (např. ulice *Cihelní*, hrobka *Synagoga*)	117	67
VM B – kdo, co (např. ulice *Důlky*, autobusové nádraží *Busák*)	17	62
VM C – jaký, za jakých okolností (ulice *Jarní*, jezero *Štěrkáč*)	45	16
VM D – čí (např. ulice *Antonína Dvořáka*, kopec *Balgarovec*)	41	11

Z tabulky 1 jsou patrné rozdíly v modelových preferencích, které jako statisticky signifikantní vyhodnocuje i chí-kvadrátový test ($\chi^2 = 61$, $p << 0{,}05$). Cramerovo V, které určuje, jak relevantní je daný rozdíl, charakterizuje vztah mezi standardizací/nestandardizací jména a vztahovými modely jako středně silný ($V = 0{,}403$). Nejvýraznějšími diferencemi jsou nízká zastoupení vztahových modelů C a D v nestandardizované toponymii, což odpovídá očekávání i závěrům výzkumu Jany Pleskalové (1992, s. 120). Objekt, který je místním obyvatelům známý, není třeba charakterizovat co do vlastnosti; vyjádření posesivity, které je spjato s VM D, je vlivem poválečných změn v zemědělství a obecně ve společnosti oslabeno a manifestuje se spíše jako honorifikace (tedy nepravá posesivita), která je typická pro více či méně politicky motivovaná standardizovaná urbanonyma. V nestandardizované toponymii jsou naproti tomu velmi frekventované VM A a VM B, které vyjadřují polohu objektu a pojmenovávají jeho podstatu, a mluvčím tudíž vyhovují svou precizností a jasnou identifikační funkcí.

Příklad 2 – Kolokační analýza

Kolokace chápeme jako slovní spojení, jejichž asociační síla nabývá určité, předem stanovené hodnoty (viz Čermák – Cvrček, 2017; Geeraerts, 2019, s. 175–183; Stefanowitsch, 2020, s. 215–259). Pro kvantitativně zaměřený onomastický výzkum se nabízejí ve dvou rovinách: jednak je možné zkoumat ustálená sousloví, která se pojí s osobami, místy, firmami apod., a odhalovat tak konotace, jež v daném kontextu tato vlastní jména mají, včetně proměn jejich sémantiky; jednak se lze zaměřit na problematiku proprií obecně a zjišťovat, zdali je pro ně určitý soubor kolokátů typický. Zatímco výzkumy z první oblasti se hodí pro zacílené případové studie, které v současné době už vznikají (srov. David – Semian, 2015;

David – Klemensová, 2019; David – Davidová Glogarová, 2020; Drkošová ‿ Místecký, 2022), zde se zaměříme na průzkum druhé, obecnější sféry.

Pro účely naší experimentální sondy budeme pracovat na jedné straně se všemi substantivy (apelativy i proprii) a na druhé straně místními jmény pětice největších měst České republiky (Praha, Brno, Ostrava, Plzeň a Liberec) v materiálu korpusu SYN2020 (ČNK), neboť jsme předpokládali, že jejich jména budou v korpusu dostatečně zastoupená (např. lemma „Liberec" se v něm objevuje s frekvencí 4 300). Tyto dva soubory jazykových dat budou podrobeny kolokační analýze za užití tří asociačních měr – skóre MI, které selektuje exkluzivní a nízkofrekvenční kolokáty, t-skóre, které vyhledává kolokace gramatické (koligace), a míry logDice, která stojí mezi těmito extrémy (Brezina, 2018; Stefanowitsch, 2020; Brezina – Pořízka, 2021). Výsledky analýz (vždy prvních deset kolokátů u každé míry) jsou prezentovány v tabulce 2. Minimální frekvence kolokátu byla stanovena na 3 a zkoumáno bylo pouze bezprostřední levé okolí slova (rozpětí -1/0).

Rozdíly mezi charakteristickými kolokáty u všech substantiv a proprií je možné klást do souvislosti s výsledky dalších analýz. Příkladně kolokát *město*, který vyhodnotily jako charakteristický míry logDice a t-skóre, může poukazovat na užívání nominativu jmenovacího u vybraných toponym (viz příklad 3 v této studii; srov. Kolářová, 2019, 2020). Podobně charakteristické je užití názvů měst v kontextu poštovních směrovacích čísel a domicilů novinových článků (kolokáty *0* a */*). Na druhé straně se jména měst nepojí s interpunkčním znaménkem tečky, což může poukazovat na fakt, že pro ně není typická pozice na počátku věty. Užití lokalizačních předložek (např. *v*, *z*, *u*, *do*) kolokační analýzy jako charakteristické pouze pro názvy měst nevyhodnotily – stejné předložky se pojí i se substantivy obecně. Pro ně je však typická též předložka *na*, která má také abstraktní význam (např. *myslet na někoho*) a mezi desítkou nejsilnějších kolokátů proprií se neobjevuje. Kolokáty nerozlišených substantiv jsou i slovnědruhově pestřejší, vyskytují se zde zájmena, adjektiva a slovesa, zatímco vybraná propria se spíše navazují na podstatná jména a předložky.

Idiosynkratický seznam kolokátů vyprodukovala metrika MI. V případě podstatných jmen ho tvoří úzkoprofilová, oborová adjektiva, převážně z oblasti medicíny a přírodních věd (*umbilikální*, *fetomaternální*), a některé nezvyklé, ale ústrojně utvořené novotvary (*oreganový*, *tučňáčí*). U zkoumaných toponym se jako kolokáty objevily další názvy – chrématonyma, která označují firmy se sídlem v daném městě. Tato jména dále navyšují nominativní výskyty místních jmen, např. ve spojeních *Promedica Praha* nebo *PEMM Brno*.

Příklad 3 – Morfologické charakteristiky vlastních jmen

Vlastní jména se v textu chovají specificky, což se obecně projevuje i v jejich morfologických charakteristikách, které jsou odlišné od apelativ (Tušková, 2011,

Tabulka 2: Kolokáty typické pro všechna substantiva a vybraná propria podle tří asociačních měr

Všechna substantiva			Místní jména – města		
Kolokát	**Absolutní frekvence kolokace**	**logDice**	**Kolokát**	**Absolutní frekvence kolokace**	**logDice**
v	1 346 440	10,47	v	27 790	8,57
.	1 381 895	10,30	město	1 337	8,09
na	1 014 852	10,08	centrum	906	8,09
a	761 222	9,61	do	4 840	7,68
,	790 577	9,46	/	882	7,62
s	448 554	8,95	0	417	7,20
do	422 372	8,87	z	4 022	7,13
z	407 386	8,81	u	1 009	6,85
být	425 224	8,71	Viktorie	256	6,62
k	286 323	8,32	část	421	6,56
Kolokát	**Absolutní frekvence kolokace**	**t-skóre**	**Kolokát**	**Absolutní frekvence kolokace**	**t-skóre**
v	1 346 440	689,05	v	27 790	157,13
na	1 014 852	566,99	do	4 840	62,72
do	422 372	398,54	z	4 022	54,11
svůj	269 034	344,99	město	1 337	35,12
tento	234 501	341,30	centrum	906	29,53
s	448 554	322,32	/	882	28,24
z	407 386	321,38	u	1 009	27,34
k	286 323	292,62	0	417	20,12
jeho	169 775	282,40	část	421	18,44
velký	142 878	257,93	Viktorie	256	15,94
Kolokát	**Absolutní frekvence kolokace**	**MI**	**Kolokát**	**Absolutní frekvence kolokace**	**MI**
Fanconi	3	2,08	Promedica	3	10,5
umbilikální	21	2,08	TOPSYS	3	10,5
albuminový	3	2,08	ARCDATA	9	10,5
Beaufortův	9	2,08	Geoservis	4	10,5
deacetylázový	3	2,08	PEMM	14	10,5
vodoměrný	5	2,08	MVIL	4	10,5
oreganový	3	2,08	GeoNet	4	10,5
tučňáčí	7	2,08	Dvořákův	3	10,5
fetomaternální	4	2,08	Osona	5	10,5
Onkogynekologické	3	2,08	SKUŘ	3	10,5

s. 51–56; Smejkalová, 2013). Vedle tendence k bohaté hypokoristické slovotvorbě je však možné se zaměřit též na tradiční gramatické kategorie, např. na pád jako kategorii tvarově bohatou a funkčně diferencovanou (srov. Štícha a kol., 2013, s. 289–377; Štícha a kol., 2021, s. 539–635).

Příklad 3A – Morfologické charakteristiky: pádová distribuce

Následující výzkum prověřuje hypotézu, zdali mají propria svou charakteristickou pádovou distribuci. K shromáždění dat jsme využili korpus SYN2020 (ČNK), protože odráží nejaktuálnější stav češtiny. Jako sonda do situace v propriální sféře byla opět vybrána pětice názvů největších českých měst (viz příklad 2). Tabulka 3 zachycuje deklinační preference všech substantiv v korpusu a vybraných místních jmen.

Tabulka 3: Pádové distribuce u substantiv a vybraných proprií v korpusu SYN2020

	Všechna substantiva		Místní jména – města	
Pád	**Absolutní frekvence**	**Procenta**	**Absolutní frekvence**	**Procenta**
nominativ	8 351 043	28,96 %	30 580	36,33 %
genitiv	7 431 853	25,77 %	17 700	21,03 %
dativ	1 076 638	3,73 %	1 018	1,21 %
akuzativ	6 001 510	20,81 %	3 186	3,79 %
vokativ	128 121	0,44 %	17	0,02 %
lokál	3 422 814	11,87 %	29 537	35,09 %
instrumentál	2 428 132	8,42 %	2 129	2,53 %

Předtím, než provedeme statistické testování, je možné detekovat rozdíly mezi distribucemi i opticky. U proprií je patrná především dominance lokálu (35 % proti 12 % u substantiv obecně), nízké zastoupení instrumentálu (2,5 % oproti 8 % u všech jmen), ale i kvaziabsence vokativu a dativu a překvapivě nízké výskyty akuzativu. Některé z těchto jevů jsou očekávané a snadno vysvětlitelné (lokalizační funkce místních jmen, jejich archetypálně neživotný charakter), ale u instrumentálu a dativu by bylo možné uvažovat o vyšších frekvencích (srov. konstrukce *jet* / *toulat se* / *projíždět Ostravou*, *blížit se k Ostravě* apod.). Nízké četnosti těchto pádů lze však vysvětlit jejich charakterem – dativ se typicky vyskytuje jako pád druhého předmětu, kterým je často životný objekt (např. *dát* / *darovat* / *ukázat někomu něco*), kdežto instrumentál sémanticky odkazuje k prostředku. Ani jeden z nich tak nepokrývá oblasti pro toponyma typické.

U akuzativu, kde je rozdíl poměrně dramatický (21 % u všech podstatných jmen proti necelým 4 % u proprií), se projevuje faktuální zaměření publicistiky a odborné literatury; je pravděpodobné, že se čtvrtým pádem se častěji pracuje např. v reklamních textech (např. v obratech *navštívit* / *zažít* / *milovat Ostravu*).

Za pozornost stojí také vyšší frekvence prvního pádu u proprií (36 % oproti necelým 27 % u všech jmen), která může souviset s užíváním nominativu jmenovacího (typ *ve městě Ostrava*) a s výše zmíněnými domicily.

V dalším kroku přistoupíme ke statistickému testování. Chí-kvadrátový test vyhodnotil rozdíly mezi pádovými distribucemi jako signifikantní (χ^2 = 57 189, p << 0,05), podle Cramerova V však vztah není příliš relevantní (V = 0,044). Tento výsledek může souviset se skutečností, že pořadí pádů u nerozlišených substantiv i proprií zůstává až na lokál stejné (viz tabulka 4). Vysoká frekvence lokálu tak zůstává základním rysem zkoumaných proprií.

Tabulka 4: Pořadí pádů podle frekvence u substantiv a proprií

Všechna substantiva	Místní jména – města
nominativ	nominativ
genitiv	lokál
akuzativ	genitiv
lokál	akuzativ
instrumentál	instrumentál
dativ	dativ
vokativ	vokativ

Příklad 3B – Morfologické charakteristiky: variantnost pádových koncovek

Druhý výzkum prověřuje hypotézu, zda variantnost koncovek u vybraných proprií vyjadřuje rozdíly v jejich gramatickém a lexikálním významu. Vycházeli jsme z materiálu korpusu SYN v10 (ČNK2) a analyzovali maskulinní toponyma (konkrétně místní jména zakončená na neměkčený konsonant) náležející k typu *hrad*. Místní jména představují „specifickou deklinační kategorii" a v případě existence variantních koncovek v genitivu a lokálu singuláru „jasná a striktní pravidla pro volbu jedné či druhé koncovky u kteréhokoli názvu obce […] neexistují a mnohé jednotlivé názvy se řídí svým individuálním územ" (Štícha a kol., 2013, s. 334). Pracovali jsme se dvěma soubory. První tvořila cíleně zvolená toponyma *Mnichov* a *Nymburk*, u nichž jsme předpokládali variantnost v deklinaci danou metaforickým používáním (*Mnichov*) a souhláskovou alternací v zakončení (*Nymburk*). Ve druhém případě nás zajímalo sémantické využití variantních koncovek. Druhý soubor tvořilo 25 názvů českých měst daného typu *hrad*, která měla v roce 2020 více než 20 tisíc obyvatel.

Příklad 3Ba – Morfologické charakteristiky: *Mnichov*, *Nymburk*

Předmětem první analýzy byla toponyma *Mnichov* a *Nymburk*. Podoba *Mnichov* je české exonymum k německé formě *München* (viz tabulka 5); v češtině navíc

označení *Mnichov* / *mnichov* funguje jako metafora aktu politické zrady s odkazem na tragické události roku 1938 (srov. Horálek, 2002, s. 240–242; SSJČ). Tato referenční specifičnost vlastního jména se projevuje ve výskytu apelativizovaných plurálových forem typu *vypočítávat naše Mnichovy, Únory a Srpny* (*Reflex*, 1998), *zabránit nejen novým krizovým cyklům, ale i repríze možných Mnichovů* (*Právo*, 2000). Tvary, které byly automatickou analýzou vyhodnoceny jako vokativy, jsou skutečnými vokativy jen v několika málo případech, jinak jde především o lokály zapsané bez diakritiky (*v Mnichove*), a tedy chybně tagované.

Analýza prokázala, že dubletní koncovky v lokálu singuláru slouží k sémantickému rozlišení. Koncovka *-ě* je užita v lokalizačním významu, což potvrzují i nejsilnější kolokáty (okno -1/0, min. frekvence kolokátu 10, míra logDice) – *německý*, *bavorský* a vůbec nejfrekventovanější předložka *v* –, zatímco koncovka *-u* se vyskytuje v případě metaforického použití toponyma pro označení konference mocností v roce 1938, popř. zrady obecně; dominantním kolokátem je zde *po*. Pokud jsme použili širší kolokační okno, a to -3/0, sémantický rozdíl mezi koncovkami v lokálu se ještě více zvýraznil a zpřesnil.

Tabulka 5: Pořadí pádů a zastoupení koncovek u lemmatu *Mnichov*. Kolokace uvádíme pouze pro tvary lokálu

Tvar místního jména	Absolutní frekvence	Pád	Kolokáty (-1/0)	Kolokáty (-3/0)
Mnichově	44 076	lokál	*bavorský*, *německý*, *v*, *metropole*, *daleký*	*bavorský*, *1972*, *OH*, *Oktoberfest*, *regata*
Mnichov	37 676	nominativ		
Mnichova	22 470	genitiv		
Mnichov	5 091	akuzativ		
Mnichovem	3 221	instrumentál		
Mnichovu	2 258	lokál	*po*, *o*, *v*, *na*	*Sudety*, *1938*, *pohraničí*, *bezprostředně*, *Beneš*
Mnichovu	967	dativ		

V případě místního jména středočeského města *Nymburk* (viz tabulka 6) korpus ukázal variantní koncovky ve dvou pádech – v genitivu a lokálu. Pokud se zaměříme na jejich sémantickou distribuci – sledujeme ji prostřednictvím kolokací (okno -1/0, min. frekvence kolokátu 10, míra logDice; vyloučili jsme interpunkční znaménka a spojky) –, nejsilnějšími kolokáty s genitivem *Nymburku* jsou *basketbalista*, *mistrovský* a *palubovka*. Situace je tematicky téměř shodná v případě pádové varianty *Nymburka* – *basketbalista*, *košíkář* a *hokejista*. Kolokační analýza zde tedy rozdílnou významovou distribuci koncovek nepotvrzuje.

V případě lokálu *Nymburku* jsou nejsilnější kolokace s výrazy *mistrovský*, *v*, *středočeský*; pro variantu *Nymburce* je to *Domav* (defektní tvar s 10 výskyty –

náležitě *Doma v*, vyloučeno z analýzy), *v*, *sousední* a *daleký*. Z hlediska absolutní frekvence jsou v obou pádech a pro všechny koncovkové varianty nejvíce zastoupeny kolokace s předložkami – v genitivu *z* a v lokálu *v*. I v tomto případě je tak sémantická diference mezi oběma tvary spíše méně patrná.

Specifický je ovšem fakt, že koncovka lokálu *-e* je – navzdory alternaci, kterou způsobuje – základní formou pro vyjádření lokálnosti, což potvrzuje i srovnání absolutních frekvencí variant lokálu *Nymburce* (37 872 výskytů) a *Nymburku* (16 683 výskytů). Že jde o místní úzus, respektive tvary vnímané jako „tradičnější nebo tradicionalistické“ (Šimandl, 2000, s. 74–75), může potvrzovat nejvyšší zastoupení typu *v Nymburce* v regionálních *Denících Bohemia*, kde předpokládáme větší koncentraci autorů používajících oblastní morfologické formy (srov. Balhar a kol., 2002, s. 131–132; viz též dále formy místního jména *Žďár*).

Tabulka 6: Pořadí pádů a zastoupení koncovek u lemmatu *Nymburk*. Kolokace uvádíme pouze pro tvary lokálu a genitivu

Tvar místního jména	Absolutní frekvence	Pád	Kolokáty (-1/0)	Kolokáty (-3/0)
Nymburk	120 682	nominativ		
Nymburka	47 705	genitiv	*basketbalista*, *košíkář*, *hokejista*, *volejbalista*, *košíkářka*	*basketbalista*, *košíkář*, *Nymburk*, *basketbalový*, *Sadská*
Nymburce	37 872	lokál	*Domav* (!), *v*, *sousední*, *daleký*, *basketbalový*, *středočeský*	*soustředění*, *Hálkův*, *Přemyslovec*, *Komenský*, *Jiljí*
Nymburku	16 683	lokál	*mistrovský*, *v*, *středočeský*, *daleký*, *rodný*	*Hálkův*, *soustředění*, *vlastivědný*, *Žid*, *Chanchong*
Nymburkem	9 996	instrumentál		
Nymburk	9 228	akuzativ		
Nymburku	5 588	dativ		
Nymburku	4 650	genitiv	*basketbalista*, *mistrovský*, *palubovka*, *skalp*, *basketbalistka*	*basketbalista*, *suverén*, *palubovka*, *Ježdík*, *mistrovský*

3Bb – Morfologické charakteristiky: česká města typu *hrad*

V druhém analyzovaném souboru (25 domácích místních jmen – názvů měst – zakončených na neměkčený konsonant, typ *hrad*, viz tabulky 7a, 7b) jsme se zaměřili na variantnost koncovek ve funkčně zásadním pádu toponym, a to v lokálu singuláru (viz výše). Jejich sémantické rozdíly jsme sledovali na základě kolokací (okno -1/0, min. frekvence kolokátu 10, míra logDice). Současně jsme museli řadu výrazů filtrovat, abychom eliminovali částečné shody s jinými toponymy obsahují-

Tabulka 7a: Variantní koncovky lokálu singuláru, místní jména typu *hrad* (frekvence)

Město (místní jméno)	Počet obyvatel	Lokál -ě (absolutní frekvence)	Lokál -u (absolutní frekvence)	Poměr koncovek -ě/-u
Zlín	74 935	199 313	1 180	168,91
Havířov	71 200	46 635	252	185,06
Most	66 034	112 587	5 623	20,02
Frýdek-Místek	55 557	**0**	43 882	0,00
Chomutov	48 635	71 324	424	168,22
Děčín	48 594	67 050	381	175,98
Prostějov	43 651	97 377	412	236,35
Přerov	42 871	98 205	296	331,77
Tábor	34 277	71 954	191	376,72
Cheb	31 977	**0**	72 155	0,00
Kolín	31 973	61 911	311	199,07
Písek	30 415	**0**	76 620	0,00
Trutnov	30 234	36 145	124	291,49
Vsetín	25 974	55 022	441	124,77
Šumperk	25 836	1	48 879	0,00
Hodonín	24 512	47 358	182	260,21
Český Těšín	24 297	17 247	8	2155,88
Litvínov	23 661	48 673	413	117,85
Havlíčkův Brod	23 442	52 561	1 434	36,65
Nový Jičín	23 260	35 043	66	530,95
Krnov	23 257	40 432	102	396,39
Sokolov	23 033	62 628	298	210,16
Vyškov	20 807	39 758	107	371,57
Žďár nad Sázavou	20 717	14 610	15 450	0,95
Bohumín	20 518	20 904	78	268,00

cími sledovaný výraz (např. *Černý Most*, *Soumarský Most*), nebo jsme ze stejného důvodu používali v dotazu pouze celé formy toponyma (např. [word="Havlíčkově"] [word="Brodě" & tag="…S6.*"]). Jsme si také vědomi možného zkreslení závěrů, které je dáno chybami v tagování a lemmatizaci (např. lemma *Sokolov* zahrnuje i tvary *Sokolovo*, referující k bojišti 2. světové války a stejnojmennému filmu).

V historickém pohledu je variantnost koncovek v lokálu singuláru u tvrdých neživotných maskulin (o-kmeny, typ *hrad*), kde byla původně koncovka *-ě*(*-e*), interpretována jako výsledek přejetí koncovky *-u* od u-kmenů (Gebauer, 1960, s. 38–39; Vážný, 1964, s. 27; srov. Lamprecht – Šlosar – Bauer, 1986, s. 141–142),

Tabulka 7b: Variantní koncovky a kolokace lokálu singuláru, místní jména typu *hrad* (okno -1/0)

Město (místní jméno)	Lokál -ě	Lokál -u
Zlín	*v*, *rodný*, *baťovský*, *extraligový*, *druholigový*	*po*, *o*, *na*, *v*
Havířov	*prvoligový*, *v*, *rodný*, *severomoravský*, *osmdesátitisícový*	*po*, *o*, *na*, *v*
Most	*severočeský*, *v*, *druholigový*, *sousední*, *sedmdesátitisícový*	*druholigový*, *severočeský*, *prvoligový*, *play*, *v*
Frýdek-Místek	**0**	*v*, *palubovka*, *třetiligový*, *šedesátitisícový*, *druholigový*
Chomutov	*severočeský*, *prvoligový*, *v*, *třetiligový*, *sousední*	*po*, *o*, *na*
Děčín	*v*, *basketbalový*, *severočeský*, *sousední*, *daleký*	*po*, *o*, *na*
Prostějov	*Wolkerův*, *v*, *prvoligový*, *rodný*, *druholigový*	*Wolkerův*, *po*, *o*, *na*
Přerov	*v*, *rodný*, *daleký*, *prvoligový druholigový*	*po*, *o*, *na*
Tábor	*jihočeský*, *v*, *rodný*, *husitský*, *prvoligový*	*po*, *o*, *na*, *v*
Cheb	**0**	*v*, *západočeský*, *divizní*, *daleký*, *rodný*
Kolín	*Kmochův*, *rodný*, *v*, *středočeský*, *daleký*	*Kmochův*, *po*, *o*, *na*, *a*
Písek	**0**	*Šrámkův*, *jihočeský*, *rodný*, *v*, *třetiligový*
Trutnov	*podkrkonošský*, *třiatřicetitisícový*, *v*, *východočeský*, *daleký*	*po*, *o*, *na*
Vsetín	*v*, *extraligový*, *rodný*, *daleký*, *sousední*	*po*, *o*, *na*
Šumperk	**0**	*prvoligový*, *rodný*, *v*, *severomoravský*, *domovský*
Hodonín	*rodný*, *sousední*, *v*, *daleký*, *druholigový*	*po*, *o*, *na*
Český Těšín	*v*, *daleký*, *po*, *o*	**0**
Litvínov	*extraligový*, *severočeský*, *sousední*, *v*, *daleký*	*po*, *o*, *na*
Havlíčkův Brod	*v*, *prvoligový*, *rodný*, *daleký*, *druholigový*	*prvoligový*, *v*, *po*
Nový Jičín	*v*, *daleký*, *divizní*, *druholigový*, *samotný*	*po*
Krnov	*severomoravský*, *v*, *rodný*, *slezský*, *daleký*	*o*, *po*
Sokolov	*druholigový*, *Sokolov*, *v*, *daleký*, *Sokolovsko*	*po*, *o*, *v*, *na*, *a*
Vyškov	*v*, *třetiligový*, *divizní*, *rodný*, *moravský*	*po*, *o*, *na*
Žďár nad Sázavou	*v*, *rodný*, *sousední*, *město*, *po*	*v*, *rodný*, *druholigový*, *daleký*, *samotný*
Bohumín	*severomoravský*, *starý*, *v*, *rodný*, *bohatý*	*o*, *po*

nebo nověji jako koncovka dativního původu (Lamprecht – Šlosar – Bauer, 1986, s. 144). Kromě předložkového charakteru lokálu, kdy byla funkce patrná z prepozice a nebylo třeba jedinečné pádové koncovky, tu hrály roli i obecnější tendence k dativu-lokálu a fakt, že koncovka *-u* nevyvolávala alternace.

Náš výzkum, založený na korpusových datech, potvrdil teze, které se objevují v nejnovějších, korpusově založených gramatikách češtiny (viz tabulky 7a, 7b). Jedná se především o převahu koncovky *-ě*(*e*) v platnosti určení místa, zatímco v jiné větněčlenské funkci (nejčastěji předmětu) se uplatňuje koncovka *-u* (Cvrček a kol., 2015, s. 202–203; Štícha a kol., 2021, s. 605–606), např. *městská policie v Havlíčkově Brodě* versus *totéž platí o Havlíčkově Brodu*.

Z hlediska sémantické funkce variantních koncovek jsou však zajímavé některé další trendy a případy, které se v materiálu objevily, respektive neobjevily:

1) absence systémově živých a produktivních forem se zakončením *-e* u některých konkrétních jmen, typ **Frýdce-Místce*, **Písce*, **Chebě*; pouze jeden doklad – *v Šumperce* – se objevil v memoárech historika Augusta Sedláčka z počátku 20. století (srov. lexikografické lakuny, Kováříková – Škrabal – Cvrček – Lukešová – Milička, 2020);
2) takřka vyrovnaná situace v případě lokálových podob toponyma *Žďár nad Sázavou* – *Žďáře* i *Žďáru*, kterou potvrzují i shodné kolokace s oběma variantními formami. Zde se zřejmě lokální úzus (srov. Balhar a kol., 2002, s. 131–132) kombinuje s tendencí k hláskové nealternaci;
3) kolokační analýza naznačuje, že koncovka *-u* se dominantně váže s předložkovými vazbami, z nichž některé pracují s názvem města v metonymickém smyslu, např. v kontextu sportovního zpravodajství – typ *Celek z Vysočiny si tak po Hodonínu připsal skalp dalšího okresního zástupce* (*Regionální deník*, 2010). Koncovka *-ě* se naproti tomu objevuje v adjektivních frázích, které město určitým způsobem charakterizují – typ *Za pomoci publika v rodném Hodoníně se pokusí stolní tenistka Iveta Vacenovská zaútočit na domácí titul* (*Mladá fronta DNES*, 2010).

Závěry

Cílem těchto prolegomen je poukázat na některé výzkumy, v nichž může být statistická práce s empirickými daty základním kamenem heuristiky i ověřování hypotéz. Strukturální modely, které je možné srovnávat na základě statistických testů, specificky a obecně zaměřené analýzy souvýskytu vlastních jmen s dalšími lexémy či morfologická specifika proprií, jež je rovněž možné sledovat v kontextu korpusových kolokací, jsou jen třemi směry, jimiž se může vydat současná i budoucí generace onomastiků. Kombinace empirismu a interpretační zdatnosti, podložená znalostí onomastické teorie a materiálu, je zárukou replikovatelnosti výzkumu a otevírá cestu jak k solidně založeným případovým studiím, tak k vybudování nové teoretické základny proprií.

V obecné rovině se potvrdila morfologická, lexikologická i syntaktická specifičnost toponym (zde konkrétně místních jmen): typicky se objevují v lokálu, spojují se s charakteristickými kolokáty (v případě názvů měst např. s názvy místních

firem, institucí a sportovních klubů) a syntagmata či metafory, jichž jsou součástí, ovlivňují jejich tvaroslovné preference (viz část 3B). Důležité je také zjištění o signifikantních rozdílech v distribuci vztahových modelů u standardizovaných a nestandardizovaných zeměpisných jmen, které může přispět k projasnění debaty o těchto dvou složkách toponymického materiálu.

Studie vznikla s podporou grantového projektu GAČR č. 22-09310S *Kvantitativní onomastika: východiska, koncepty, aplikace* (poskytovatel Grantová agentura České republiky).

PRAMENY A LITERATURA

BALHAR, J. a kol. (2002): *Český jazykový atlas* 4. Praha: Academia.

BAUKO, J. (2011): Výskumné oblasti socioonomastiky. In: M. Majtán (ed.), *Jazykovedné štúdie* 29. *Život medzi apelatívami a propriami*. Bratislava: Veda, s. 38–45.

BREZINA, V. – POŘÍZKA, P. (2021): Kolokační grafy a sítě s použitím nástroje #LancsBox: aplikace v angličtině a češtině. *Časopis pro moderní filologii*, 103, 1, s. 36–59, https://doi.org/10.14712/23366591.2021.1.2.

BREZINA, V. (2018): *Statistics in Corpus Linguistics*. Cambridge: Cambridge University Press.

COLLINS, L. – BREZINA, V. – DEMJÉN, Z. – SEMINO, E. – WOODS, A. (2022): Corpus linguistics and clinical psychology: Investigating personification in first-person accounts of voice-hearing. *International Journal of Corpus Linguistics*, 28, 1, s. 28–59, https://doi.org/10.1075/ijcl.21019.col.

CVRČEK, V. – KOVÁŘÍKOVÁ, D. (2011): Možnosti a meze korpusové lingvistiky. *Naše řeč*, 94, 3, s. 113–133.

CVRČEK, V. a kol. (2015): *Mluvnice současné češtiny* 1. Praha: Karolinum.

CVRČEK, V. a kol. (2020): *Registry v češtině*. Praha: Nakladatelství Lidové noviny.

CVRČEK, V. (2022): Proměny registrů české žurnalistiky 1995–2018. *Časopis pro moderní filologii*, 104, 1, s. 7–34, https://doi.org/10.14712/23366591.2022.1.1.

Czech Named Entity Corpus. Praha: ÚFAL [on-line]. <https://ufal.mff.cuni.cz/cnec/cnec2.0> cit. 2022-10-10.

ČECH, R. (2005): Limity (nejen jazykovědného) strukturalismu. *Slovo a slovesnost*, 66, 1, s. 19–31.

ČERMÁK, F. – CVRČEK, V. (2017): Kolokace. In: P. Karlík – M. Nekula – J. Pleskalová (eds.), *CzechEncy – Nový encyklopedický slovník češtiny*. Brno: Ústav českého jazyka FF MU [on-line]. <https://www.czechency.org/slovnik/KOLOKACE> cit. 2022-08-21.

ČNK: *Český národní korpus – SYN2020*. Praha: Ústav Českého národního korpusu FF UK [on-line]. <http://www.korpus.cz> cit. 2022-08-21.

ČNK2: *Český národní korpus – SYN v10*. Praha: Ústav Českého národního korpusu FF UK [on-line]. <http://www.korpus.cz> cit. 2022-08-21.

DAVID, J. (2011): *Smrdov, Brežněves a Rychlonožkova ulice. Kapitoly z moderní české toponymie*. Praha: Academia.

DAVID, J. – SEMIAN, M. (2015): Bohemian Manchester, Moravian Klondike and Silesian Versailles: Attributive structures of foreign place names in contemporary Czech journalism. *Slovo a slovesnost*, 76, 2, s. 129–147.
DAVID, J. – DAVIDOVÁ GLOGAROVÁ, J. (2020): Český Brad Pitt, moravská Mata Hari – antroponymické variování v současné české publicistice. *Acta onomastica*, 61, 2, s. 287–298.
DAVID, J. – KLEMENSOVÁ, T. (2019): Still Having a Conflict Potential? German and Hungarian Toponyms in the Czech and Slovak National Corpora Texts. *Miscellanea Geographica – Regional Studies on Development*, 23, 3, s. 158–162, https://doi.org/10.2478/mgrsd-2019-0005.
DAVID, J. – KLEMENSOVÁ, T. – MÍSTECKÝ, M. a kol. (2022): *Od etymologie ke krajině. Onomastika pro 21. století*. Brno: Host.
DAVID, J. – MÁCHA, P. (2014): *Názvy míst. Paměť, identita, kulturní dědictví*. Brno: Host.
DRKOŠOVÁ, V. – MÍSTECKÝ, M. (2022): Od patvaru k SuperStar: Frekvenční a kolokační analýza názvu Česko v korpusové publicistice z let 1990–2018. *Acta onomastica*, 63, 1, s. 99–115.
DUŠKOVÁ, N. (2021): *Toponymie Hlučína* [bakalářská práce]. Ostrava: Ostravská univerzita.
DVOŘÁKOVÁ, Ž (2017): *Literární onomastika. Antroponyma*. Praha: FF UK.
DVOŘÁKOVÁ, Ž. (2019): Propria v moderních výkladových slovnících češtiny. *Naše řeč*, 102, 1–2, s. 119–132.
GEBAUER, J. (1960): *Historická mluvnice jazyka českého* 3/1. Praha: Nakladatelství ČSAV.
GEERAERTS, D. (2019): *Teorie lexikální sémantiky*. Praha: Karolinum.
GIGER, M. – ŠTĚPÁN, P. (2007): Pojmenované entity v počítačové lingvistice a vlastní jména. *Acta onomastica*, 48, 1, s. 44–53.
GRIES, S. Th. (2013): *Statistics for Linguistics with R. A Practical Introduction*. Berlin – Boston: De Gruyter Mouton.
HERNÁNDEZ-LORENZO, L. (2022): Stylistic Change in Early Modern Spanish Poetry Through Network Analysis (with an Especial Focus on Fernando de Herrera's Role). *Neophilologus*, 106, s. 397–417, https://doi.org/10.1007/s11061-021-09717-2.
HORÁLEK, J. (2002): *…nejde jen o slova*. Chomutov: Milenium Publishing.
HŘEBÍČEK, L. (2002): *Vyprávění o lingvistických experimentech s textem*. Praha: Academia.
JELÍNEK, J. – BEČKA, J. V. – TĚŠITELOVÁ, M. (1961): *Frekvence slov, slovních druhů a tvarů v českém jazyce*. Praha: SPN.
JORDAN, P. – MÁCHA, P. a kol. (2021): *Place-Names Politics in Multilingual Areas. A Comparative Study of Southern Carinthia (Austria) and the Těšín/Cieszyn Region (Czechia)*. London: Palgrave Macmillan Cham.
KNAPPOVÁ, M. (1989): *Rodné jméno v jazyce a společnosti*. Praha: Academia.
KOLÁŘOVÁ, I. (2019): Pod hradem Holštejn/Holštejnem, pod horou Milešovka/Milešovkou: konkurence shodného přívlastku a nominativu jmenovacího v postpozici vybraných jmen. In: P. Karlík – P. Malčík (eds.), *Svět podle Grepla*. Brno: Host, s. 111–121.
KOLÁŘOVÁ, I. (2020): V ulici Kobližná/Kobližné, do ulice Nerudova/Nerudovy: gramatické tvary adjektivních jmen v postpozici jména obecného. In: I. Kolářová – H. Svobodová (eds.), *Variantnost v onymii a dialektech*. Brno: Masarykova univerzita, s. 139–154.
KOVÁŘÍKOVÁ, D. – ŠKRABAL, M. – CVRČEK, V. – LUKEŠOVÁ, L. – MILIČKA, J. (2020): Lexicographer's lacunas or how to deal with missing representative dictionary

forms on the example of Czech. *International Journal of Lexicography*, 33, 1, s. 90–103, https://doi.org/10.1093/ijl/ecz027.
KÖHLER, R. (1986): *Zur linguistischen Synergetik. Struktur und Dynamik der Lexik*. Bochum: Brockmeyer.
LÁBUS, V. – VRBÍK, D. (2018): *Toponyma v krajině a možnosti jejich výzkumu*. Liberec: Technická univerzita v Liberci.
LAMPRECHT, A. – ŠLOSAR, D. – BAUER, J. (1986): *Historická mluvnice češtiny*. Praha: SPN.
LEVÝ, J. (1971): *Bude literární věda exaktní vědou?* Praha: Československý spisovatel.
MARKERT, K. – NISSIM, M. (2007): Metonymic proper names. A corpus-based account. In: A. Stefanowitsch – S. Th. Gries (eds.), *Corpus-Based Approaches to Metaphor and Metonymy*. Berlin: De Gruyter-Mouton, s. 152–174.
MÍSTECKÝ, M. – KUBÁT, M. (2021): Přepište dějiny české stylistiky: Registr je nový styl. *Naše řeč*, 104, 3, s. 184–189.
MOTSCHENBACHER, Heiko (2020): Corpus Linguistic Onomastics. A Plea for a Corpus Based Investigation of Names. *Names*, 68, 2, s. 88–103, https://doi.org/10.1080/00277738.2020.1731240.
NameTag 2. Praha: ÚFAL [on-line]. <https://ufal.mff.cuni.cz/nametag/2> cit. 2022-10-10.
NEUWIRTHOVÁ, K. (2020): *Toponymie Vratimova* [bakalářská práce]. Ostrava: Ostravská univerzita.
NICK, I. M. (2017): Squaw Teats*, Harney Peak, and Negrohead Creek*: A Corpus-Linguistic Investigation of Proposals to Change Official US Toponymy to (Dis)honor Indigenous US Americans. *Names*, 65, 4, s. 223–234, https://doi.org/10.1080/00277738.2017.1370067.
OSOLSOBĚ, K. – ŽIŽKOVÁ, H. (2020): Homonymie mezi apelativy a proprii jako problém automatické morfologické analýzy češtiny. *Acta onomastica*, 61, 1, s. 161–174.
OSOLSOBĚ, K. – ŽIŽKOVÁ, H. (2021): Homonymie mezi oikonymy a antroponymy zakončenými na *-slav*/*-slava* jako problém automatické morfologické analýzy. *Acta onomastica*, 62, 2, s. 377–403.
PDT – Vlastní jména a názvy. Praha: ÚFAL [on-line]. <https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/t-layer/html/ch07s08s02.html> cit. 2022-10-10.
PETKEVIČ, V. (2017): Disambiguace / Desambiguace. In: P. Karlík – M. Nekula – J. Pleskalová (eds.), *CzechEncy – Nový encyklopedický slovník češtiny*. Brno: Ústav českého jazyka FF MU [on-line]. <https://www.czechency.org/slovnik/DISAMBIGUACE%20/%20DESAMBIGUACE> cit. 2022-10-10.
PETRŮ, E. (1968): *Exaktní metody v literárněvědné práci*. Olomouc: Univerzita Palackého.
PLECHÁČ, P. – KOLÁR, R. (2017): *Kapitoly z korpusové versologie*. Praha: Akropolis.
PLESKALOVÁ, J. (1992): *Tvoření pomístních jmen na Moravě a ve Slezsku*. Jinočany: H+H.
PLUSKAL, M. (1989): Hodnocení kvality textu učebnic zeměpisu. *Sborník ČSGS*, 94, 4, s. 285–292.
PTÁČNÍKOVÁ, M. (2021): *Mezi Stalinovou třídou a Husákovým tichem. Kapitoly z pražské toponymie*. Praha: Academia.
RAMBOUSEK, A. – HORÁK, A. – PARKIN, H. (2018): Software Tools for Big Data Resources in Family Names Dictionaries. *Names*, 66, 4, s. 246–255, https://doi.org/10.1080/00277738.2018.1453276.
RICHTA, R. a kol. (1966): *Civilizace na rozcestí*. Praha: Svoboda.

SEDLAČÍKOVÁ, B. (2013): *Historie matematické lingvistiky*. Praha: CERM.
SGALL, P. – HAJIČOVÁ, E. – PIŤHA, P. (1982): *Učíme stroje česky*. Praha: Panorama.
SCHREIBMAN, S. – SIEMENS, R. (2008, eds.): *A Companion to Digital Literary Studies*. Oxford: Blackwell.
SMEJKALOVÁ, K. (2013): Odlišnosti gramatických kategorií apelativ a proprií (antroponym a toponym). In: O. Uličný – O. Bláha (eds.), *Studie k moderní mluvnici češtiny 6. Úvahy o české morfologii*. Olomouc: Univerzita Palackého, s. 65–84.
SSJČ: *Slovník spisovného jazyka českého*. In: *Internetová jazyková příručka*. Praha: Ústav pro jazyk český AV ČR, v. v. i. [on-line]. <https://prirucka.ujc.cas.cz/?slovo=Mnichov> cit. 2022-10-10.
STEFANOWITSCH, A. (2020): *Corpus linguistics: A guide to the methodology*. Berlin: Language Science Press.
STRAKOVÁ, J. – STRAKA, M. – ŠEVČÍKOVÁ, M. – ŽABOKRTSKÝ, Z. (2017): Czech Named Entity Corpus. In: N. Ide – J. Pustejovsky (eds.), *Handbook of Linguistic Annotation*. Dordrecht: Springer Netherlands, s. 855–873.
ŠEVČÍKOVÁ, M. – ŽABOKRTSKÝ, Z. – KRŮZA, O. (2007): *Zpracování pojmenovaných entit v českých textech*. Praha: ÚFAL [on-line]. <https://ufal.mff.cuni.cz/~zabokrtsky/reports/techrep-ne-2007.pdf> cit. 2022-10-10.
ŠIMANDL, J. (2000): Morfologická problematika v jazykové poradně 1. *Naše řeč*, 83, 2, s. 57–76.
ŠMEJKALOVÁ, M. (2015): *Praporu věren i ve ztraceném boji: Vladimír Šmilauer – život a dílo filologa (1895–1983)*. Praha: Academia.
ŠMILAUER, V. (1966): *Novočeská skladba*. Praha: SPN.
ŠMILAUER, V. (1972): *Nauka o českém jazyku*. Praha: SPN.
ŠMILAUER, V. (1974): Úvodem. In: F. Kopečný, *Průvodce našimi jmény*. Praha: Academia, s. 9–29.
ŠRÁMEK, R. (1976): Slovotvorný model v české toponymii. *Slovo a slovesnost*, 37, 2, s. 112–120.
ŠRÁMEK, R. (1999): *Úvod do obecné onomastiky*. Brno: Masarykova univerzita.
ŠRÁMEK, R. (2007): Onomastika. In: J. Pleskalová – P. Karlík – M. Krčmová – R. Večerka (eds.), *Kapitoly z dějin české jazykovědné bohemistiky*. Praha: Academia, s. 377–425.
ŠTĚPÁN, P. (2020): Psaní velkých písmen u toponym v češtině: polemický pohled. *Acta onomastica*, 61, 2, s. 417–439.
ŠTÍCHA, F. a kol. (2013): *Akademická gramatika spisovné češtiny*. Praha: Academia.
ŠTÍCHA, F. a kol. (2021): *Velká akademická gramatika spisovné češtiny* 2. *Morfologie: Morfologické kategorie / Flexe. Část* 2. Praha: Academia.
TĚŠITELOVÁ, M. (1965): Profesor Vladimír Šmilauer sedmdesátníkem. *Naše řeč*, 48, 5, s. 257–263.
THIELEN, Ch. (1995): *An Approach to Proper Name Tagging for German* [on-line]. <https://arxiv.org/pdf/cmp-lg/9506024.pdf> cit. 2022-10-10.
TSE, G. Y. W. (2005): *A Corpus-based Study of Proper Names in Present-day English. Aspects of Gradience and Article Usage*. Berlin – Oxford: Peter Lang.
TUŠKOVÁ, J. M. (2011): *Deklinační systém femininních oikonym v češtině. Synchronní pohled na základě Českého národního korpusu*. Praha: Nakladatelství Lidové noviny.

TUŠKOVÁ, J. M. (2012): Transonymizace oikonym a její projevy v morfologické rovině. *Acta onomastica*, 53, 1, s. 324–334.
TUŠKOVÁ, J. M. (2019): Korpusy češtiny a výzkum (gramatických vlastností) proprií. In: A. Chomová – J. Krško – I. Valentová (eds.), *Konvergencie a divergencie v propriálnej sfére*. Bratislava: Veda, s. 126–137.
TUŠKOVÁ, J. M. – ŽIŽKOVÁ, H. (2019): Variantní a dubletní tvary českých maskulinních oikonym a jejich fungování v současné psané komunikaci. In: K. Klímová – H. Žižková (eds.), *Čeština v pohybu* 2. *Teoretické a komunikační aspekty proprií*. Brno: Masarykova univerzita, s. 7–29.
UHLÍŘOVÁ, L. (1975): O frekvenci příslovečného určení v souvislém textu. *Naše řeč*, 58, 3, s. 133–142.
UHLÍŘOVÁ, L. (2017): Kvantitativní lingvistika. In: P. Karlík – M. Nekula – J. Pleskalová (eds.), *CzechEncy – Nový encyklopedický slovník češtiny*. Brno: Ústav českého jazyka FF MU [on-line]. <https://www.czechency.org/slovnik/KVANTITATIVN%C3%8D%20LINGVISTIKA> cit. 2022-08-21.
VAJTROVÁ, V. (2021): *Toponymie Řepišť*. Bakalářská práce. Ostrava: Ostravská univerzita.
VÁŽNÝ, V. (1964): *Historická mluvnice* 2. Praha: SPN.
YODER, P. J. (2006): Predicting Lexical Density Growth Rate in Young Children With Autism Spectrum Disorders. *American Journal of Speech-Language Pathology*, 15, s. 378–388, https://doi.org/10.1044/1058-0360(2006/035).
ZIPF, G. K. (1949): *Human Behavior and the Principle of Least Effort*. Boston: Addison-Wesley.
ZMĚLÍK, R. (2017): Bude literární věda exaktní vědou ve století jednadvacátém? *Bohemica litteraria*, 20, 2, s. 35–45, https://doi.org/10.5817/BL2017-2-3.
ZMĚLÍK, R. (2018): Kvantitativní a korpusový výzkum v literární vědě: možnosti a přístupy. *Slovo a slovesnost*, 79, 1, s. 47–65.

Jaroslav David
ORCID: 0000-0003-2443-5351
jaroslav.david@osu.cz
Katedra českého jazyka
Filozofická fakulta
Ostravská univerzita
Reální 3
701 03 Ostrava

Michal Místecký
ORCID: 0000-0002-9183-4435
michal.místecky@osu.cz
Katedra českého jazyka
Filozofická fakulta
Ostravská univerzita
Reální 3
701 03 Ostrava

Tento článek je volně přístupný na základě licence *Creative Commons Uveďte původ 4.0 Mezinárodní veřejná licence*, která umožňuje používání, šíření a reprodukci článku za podmínky, že původní článek bude řádně citován. Licenční podmínky jsou dostupné zde: https://creativecommons.org/licenses/by/4.0/legalcode.cs.

This is an open access article under the terms of the *Creative Commons Attribution 4.0 International Public License*, which permits use, distribution and reproduction in any medium, provided the original article is properly cited. The license terms are available here: https://creativecommons.org/licenses/by/4.0/legalcode.

View publication stats