

Klára Osolsobě (Brno) – Hana Žižková (Brno)

HOMONYMIE MEZI OIKONYMY A ANTROPONYMY ZAKONČENÝMI NA *-SLAV/-SLAVA* JAKO PROBLÉM AUTOMATICKÉ MORFOLOGICKÉ ANALÝZY

HOMONYMY BETWEEN OIKONYMS AND ANTHROPONYMS ENDING IN *-SLAV/-SLAVA* AS A PROBLEM OF AUTOMATIC MORPHOLOGICAL ANALYSIS

Homonymy at all levels, which is a distinct feature of all natural languages, is also one of the most significant obstacles to automatic natural language processing. In this paper, we will point out the morphosyntactic differences of Czech anthroponyms ending in *-slav* (*Miroslav*-type, masculine) and Czech oikonyms with the same ending (*Miroslav*-type, feminine) and Czech anthroponyms ending in *-slava* (*Miroslava*-type feminine) because its forms are homonymous with both: masculine anthroponyms and feminine oikonyms). The analysis of data from the Syn v8 corpus shows that word form homonymy significantly influences the results of automatic morphological analysis. We will document errors in the coverage of the automatic analyzer dictionary and, above all, errors in morphological tagging, and we will propose a solution to partially improve the automatic disambiguation of the given type of proper nouns.

Key words

anthroponym; oikonym; ending in *-slav/-slava*; corpus; tagging; disambiguation

Klíčová slova

antroponymum; oikonymum; zakončení na *-slav/-slava*; korpus; morfologické značkování; desambiguace

Úvod

Homonymie na všech úrovních, která je nepřehlédnutelnou vlastností každého přirozeného jazyka, je zároveň jednou z nejvýraznějších překážek automatického zpracování přirozeného jazyka. V tomto příspěvku poukážeme na morfosyntaktické odlišnosti antroponym zakončených na *-slav* (maskulina typu *Miroslav*), oikonym se stejným zakončením (feminina typu *Miroslav*) a antroponym zakončených na *-slava* (feminina typu *Miroslava*, jejichž tvary jsou homonymní s antroponymy maskuliny i s oikonymy femininy). Na analýze dat z korpusu Syn v8 ukážeme, že tvarová homonymie značně ovlivňuje výsledky automatické morfologické analýzy. Doložíme chyby v pokrytí slovníku automatického analyzátoru a především chyby v morfologickém značkování a navrhneme řešení na částečné zlepšení automatické desambiguace uvedeného typu vlastních jmen.

Historický exkurz

Významným pojmenovacím motivem, na jehož základě se tvořila místní jména od samého počátku češtiny, byla posesivita. Vlastní jméno posesivní vyjadřuje vlastnictví objektu konkrétní osobou nebo nositelem určité funkce. Je utvořeno z antroponyma, popř. apelativa, a to nejčastěji sufify *-ov-* nebo *-in-*. Skupina vlastních jmen posesivních zahrnuje typologicky pestrou škálu oikonym tvořených z antroponym, popř. apelativních označení osob reprezentujících určité funkce, např. *Vítkov*, *Opatov*, tj. „Vítkův/opatův majetek, hrad n. dvůr“ (David, 2017).

Vedle převažujícího sufify *-ov* se ale též uplatňoval praslovanský sufify **-jb*. V praslovanštině se spojoval jen s názvy živých bytostí (stsl. *člověčb*), jisté omezení bylo možné pozorovat u základů jo-kmenů, jejichž poslední konsonant nepodléhal alternaci (*-lb* apod.) Tato derivace se totiž už od praslovanských dob (od splynutí *-j-* s posledním konsonantem základu) manifestuje jen konsonantickou alternací, a proto by neměla smysl, srov. *roditelb* (subst.) > **roditel'b* (adjektivum, ale rovněž jmenná deklinace). Z tohoto důvodu sufify **-jb* v rané staré češtině svou produktivitu ztrácí (Lamprecht – Šlosar – Bauer, 1986). Po určitou dobu se však sufify **-jb* uplatňuje při tvoření posesivních místních jmen jako například např. *Boleslav*, *Břeclav*, *Chotěboř*, *Jaroměř*, *Litomyšl*, *Mladeč*, *Spytihněv*, tj. „Boleslavův“ (majetek, hrad n. dvůr) (David, 2017).

Co se týče přechodu od původně mužského rodu k rodu ženskému, Šrámek a Lutterer (2004) uvádějí, že právě tato starobylá přivlastňovací přípona *-jb* způsobila změkčení koncového konsonantu a podmínila přechod od původně rodu mužského (ten *Břecislav*, hrad Břetislavův) k rodu ženskému (ta *Břeclav* – srov. *Stará* a *Mladá Boleslav*). David s Rousem (2006) se zase domnívají, že přechod jmen od mužského rodu k ženskému byl způsoben patrně vlivem velkých skupin místních jmen ženského rodu, jako jsou jména typu *Lomnice*, *Roudná* a *Litoměřice*.

V dnešní češtině jsou naprosto bez výjimek antroponyma zakončená na *-slav* rodu mužského a místní jména zakončená na *-slav* rodu ženského.

Substantiva zakončená na *-slav/-slava* v synchronních korpusech a ve slovníku *MorfFlex*

V následující analýze se opíráme o data získaná z korpusu Syn v8 – spojení korpusů řady SYN, verze 8 z 12. 12. 2019, velikost: 5 391 362 082 pozic (Křen a kol., 2019).¹ Data jsou morfologicky značkována, přesné údaje k jednotlivým anotačním nástrojům viz Hnátková a kol., 2014.

¹ Viz <https://wiki.korpus.cz/doku.php/cnk:syn:verze8>.

Substantiva *Bohuslav*, *Boleslav*, *Borislav*, *Bořislav*, *Břetislav*, *Budislav*, *Domaslav*, *Jaroslav*, *Miroslav*, *Přibyslav*, *Soběslav*, *Sulislav*, *Svatoslav*, *Vladislav*, *Vlastislav*, *Vratislav*, *Zbraslav*, *Zbyslav*, *Zdeslav*, *Zdislav* mají ve slovníku *MorffFlex*² jak interpretaci antroponym (maskulin životných), tak oikonym (feminin). Většina mužských antroponym má protějšky mezi antroponymy ženského rodu zakončenými na *-slava*.

Ve slovníku *MorffFlex* jsou pouze jako oikonyma ženského rodu označena substantiva *Budislav G*,³ *Čáslav G*,⁴ *Nadslav G*, *Nosislav G*, *Perejaslav G*, *Předslav G*, *Preslav G*, *Sudislav G*,⁵ *Sulislav G*,⁶ *Vrchoslav G*, *Wodislav G*, *Zbraslav G*,⁷ nicméně v korpusech se okrajově některá vyskytují i jako antroponyma (viz doklady v poznámkách vč. interpretací přiřazených automatickou morfologickou analýzou, které nejsou s ohledem na nedostatečné pokrytí slovníku výsledkem chybné desambiguace⁸).

Ve slovníku *MorffFlex* figurují pouze jako antroponyma mužského rodu jména: *Blahoslav Y*, *Boguslav Y*, *Bojislav Y*, *Božislav Y*, *Břetislav Y*, *Bronislav Y*, *Čechoslav Y*, *Česlav Y*, *Dobroslav Y*,⁹ *Drahoslav Y*, *Hostislav Y*, *Kostislav Y*, *Krunoslav Y*, *Květoslav Y*, *Ladislav Y*, *Levoslav Y*, *Liboslav Y*, *Luboslav Y*, *Miloslav Y*, *Mnislav Y*, *Pobraslav Y*, *Pravoslav Y*, *Radoslav Y*, *Radslav Y*, *Rastislav Y*, *Rostislav Y*, *Stani-*

² Přestože morfologický slovník *MorffFlex* (Hajič – Hlaváčová, 2016) zahrnuje kromě morfologické značky (tagu) také informace o vlastních jménech a rozlišuje křestní jména (_Y), příjmení (_S) a geografické názvy (_G), není podle autorů slovníku v těchto interpretacích ani důsledný, ani zcela spolehlivý. Z tohoto důvodu jsme se pokusili formulovat co nejvíce pravidel pro desambiguaci, aniž bychom se o příslušné informace opírali.

³ Doklad z korpusu Syn v8 na chybu v tagu: *Příjmení Jozefa Zieľeňce (či Zieleniece?) má kořeny v polštině, Petry Buzkové vzniklo z osobního jména Budislav/Budislav/NNFSI----A-----*. Z kontextu plyne, že jde o antroponymum mužského rodu.

⁴ Doklad z korpusu Syn v8 na chybu v tagu: *Jméno vsi je odvozeno od středověkého jména Půta; již r. 1183 je uveden v listinách jakýsi Čáslav/Čáslav/NNFSI----A----- z Potína*.

⁵ Doklad z korpusu Syn v8 na chybu v tagu: *První zmínka o zdejším sídle je z roku 1285, kdy zde měl majetek Sudislav/Sudislav/NNFSI----A----- ze Zásmuk, královský kuchmistr*.

⁶ Doklad z korpusu Syn v8 na chybu v tagu: *První zmínka o obci Zásmuky se datuje k roku 1285, kdy byl majitelem zdejšího panství Sulislav/Sulislav/NNFSI----A----- ze Zásmuk*.

⁷ Doklad z korpusu Syn v8 na chybu v tagu: *Po sto letech se však jako držitel vsi připomíná okolo roku 1238 Zbraslav/Zbraslav/NNFSI----A----- z Miletína, významný šlechtic na dvoře krále Václava I*.

⁸ Viz <https://wiki.korpus.cz/doku.php/pojmy:desambiguace>.

⁹ Doklad z korpusu Syn v8 na chybu v tagu: (...) *Bájo představení Cirkus Dobroslav/Dobroslava/NNFP2-----A-----* (...). Další kolokace označené jako NNFP2.* (femininum v genitivu plurálu) jsou např. *V časopisu Dobroslav; akademický spolek Dobroslav*. Domníváme se, že v těchto případech jde spíše o maskulina neživotná. Je ovšem otázka, zda je třeba víceroz interpretací ve slovníku zatěžovat desambiguaci.

slav Y,¹⁰ *Stranislav Y, Svjatoslav Y, Vácslav¹¹ Y, Věnceslav Y, Věroslav Y, Věslav Y, Věstislav Y, Vítězslav Y, Vitoslav Y, Vjačeslav Y, Vjekoslav Y, Vrastislav Y, Vseslav Y, Zbislav Y,¹² Zděslav Y, Zlatoslav Y,¹³* nicméně v korpusech se okrajově některá vyskytují i jako oikonyma (viz doklady v poznámkách vč. interpretací přiřazených automatickou morfologickou analýzou, které nejsou s ohledem na nedostatečné pokrytí slovníku výsledkem chybné desambiguace).

Jako antroponyma ženského rodu jsou ve slovníku *MorfFlex* interpretována substantiva: *Miroslava, Stanislava, Miloslava, Květoslava, Ladislava, Zdislava, Čáslava,¹⁴ Vladislava, Bohuslava, Vítězslava, Bronislava, Drahoslava, Věnceslava, Boleslava, Dobroslava, Vlastislava, Svatoslava, Přibyslava, Střezislava,¹⁵ Liboslava, Bořislava, Hostislava*. Ke všem ovšem existují i antroponyma maskulina životná, k řadě z nich i oikonyma ženského rodu zakončená na *-slav*.

Substantiva *Bratislava, Oslava, Ůslava, Sudslava* jsou ve slovníku *MorfFlex* interpretována pouze jako oikonyma ženského rodu/hydronyma. V korpusech se ovšem vyskytují i antroponyma mužského rodu *Bratislav¹⁶* a *Sudslav¹⁷*.

Slovník *MorfFlex* vykazuje některé nedostatky v pokrytí proprií zakončených na *-slav/-slava*. Navrhujeme, aby do slovníku *MorfFlex* byla doplněna: a) antroponyma mužského rodu životného (*Budislav, Bratislav, Čáslav, Sudislav, Sudslav, Sulislav, Střezislav, Zbraslav*), b) oikonyma ženského rodu (*Stanislav, Zbislav*), c) antroponyma ženského rodu (*Zlatoslava*) a přimlouváme se za d) vyřazení antroponyma ženského rodu *Čáslava*. Návrh se opírá o doklady v textech v korpusu Syn v8.

¹⁰ Doklad z korpusu Syn v8 na chybu v tagu: *Vě stejné době se u města **Stanislav/Stanislav/NNMS1**----**A**---- nacházela německá 1. a 8. tanková divize.*

¹¹ Všechny doklady z korpusu Syn v8 jsou ovšem překlepy.

¹² Doklad z korpusu Syn v8 na chybu v tagu: *Kapli v osadě **Zbislav/Zbislav/NNMS1**----**A**---- opraví obec Zhoř.*

¹³ Doklad z korpusu Syn v8 na chybu v tagu: (...) *řekla vedoucí laborantka **Zlatoslava/Zlatoslav/NNMS2**----**A**---- Přecechtělová.*

¹⁴ V korpusech Syn v8 je ovšem doloženo hojně přijetí *Čáslava*. Možnost femininní interpretace mají kolokace *společnost/firma/hospoda/penzion/polosamota/chata Čáslava* (přestože je tvar značený jako femininum, může jít stejně dobře o maskulinum, podle kontextu jde o nominativ jmenovací) a z dokladů na internetu vyplývá, že jde o přijetí majitele firmy. Na relevantní doklady křestního jména *Čáslava* jsme v korpusu Syn v8 nenarazili.

¹⁵ Slovník *MorfFlex* ovšem má jen antroponymum ženského rodu *Střezislava*, navzdory v korpusu Syn v8 okrajově doloženému maskulinu *Střezislav*: *Na zámek přijel **Střezislav/Střezislava/NNFP2**----**A**---- Megawat, proslavený šermíř (...).*

¹⁶ Doklad z korpusu Syn v8 na chybu v tagu: (...) *přiletěli si pro něj srbský ministr obrany **Bratislav/Bratislava/NNFP2**----**A**---- Gašić; (...) říká v rozhovoru pro MF DNES archeolog **Bratislav/Bratislava/NNFP2**----**A**---- Resutík.*

¹⁷ Doklad z korpusu Syn v8 na chybu v tagu: *Poprvé je o ní zmínka v **Zemských deskách z r. 1263, kdy o ní psal rytíř **Sudslav/Sudslava/NNFP2**----**A**---- z Dubného.***

Nedostatečné pokrytí slovníku se ovšem týká i dalších proprií zakončených na *-slav/-slava*. Lze je získat z dat korpusu Syn v8 dotazem [word="[[[:upper:]].*(slav|slava)" & tag="X.*"]. Množství z takto vyhledaných dat jsou překlepy, spojená slova atd. Objevují se ovšem i relevantní propria. S frekvencí 5 a více výskytů jsou v korpusu Syn v8 doložena antroponyma – křestní jména v mužské i ženské podobě, většinou pocházejí z různých slovanských jazyků, často jde o grafické varianty.¹⁸ Kromě skutečných antroponym jsou doložena žertovná jména (*Gumislav, Pepislav, Bečislav, Šlaposlav, Knihoslav, Dluhoslav*),¹⁹ *Pravdoslav, Hromoslav*), příjmení (*Latislav, Ratislav, Záslav, Pěvoslav, Vrchoslav*), oikonyma (*Sukoslav, Stejslav*),²⁰ *Donuslav*,²¹ *Tvrdoslav, Trojslava*,²² *Předslava*),²³ hydronymum *Reslava*. O produktivitě modelu svědčí hravé názvy jako jméno souboru *Pěslav*, kapela *Smrtislav*, divadelní spolek *Milislav*, pivo *Kvasslav*, sportovní tým ragbistek *Pepřeslav*, zkratkové slovo *Polyslav* označující skupinu slavistů potažmo slavistických konferencí.²⁴ V užití tato substantiva kolísají v rodě (mají tvary jak podle maskulin životných, tak neživotných). Lemmata nerozpoznaná automatickou morfologickou analýzou s frekvencí vyšší než 5 navrhuje zařadit do slovníku *MorfFlex*.

¹⁸ Seznam podle frekvence v korpusu Syn v8: *Vladyslav(a), Vítazoslav, Ludoslav(a), Czeslav, Vojslav(a), Hviezdoslav, Hvězdoslav(a), Mojslav(a), Krasoslav(a), Ljuboslav, Svetoslav, Mislav, Milislav, Luboslav(a), Myroslav(a), Kvetoslav(a), Ventzislav, Váslav, Světoslav(a), Přemyslav, Yvacheslav, Yaroslav, Věkoslav(a), Wladyslav(a), Šlechtislav, Viacheslav, Svetislav, Wieslav(a), Vítislav, Lubislav(a), Ladoslav, Vencislav, Ninoslav, Rostlislav, Vieroslav(a), Berislav, Držislav, Veaceslav, Uladzislav, Wenceslav, Wladislav(a), Venceslav, Vitoslav, Przemyslav, Mnohoslav, Zorislav, Věřislav, Mečislav, Rostyslav, Něhoslav(a), Žiarislav, Světlav, Leslav, Svěslav, Vněslav, Kristoslav, Bogislav(a), Čistoslav, Vlaslav, Chotěslav, Novislav, Siloslav, Ruslav, Ludislav(a), Zveneslav, Prvoslav, Sviatoslav, Byslav, Pivoslav, Lavoslav, Beloslav, Lutoslav, Unislav, Tsvetoslav, Svatislav, Doběslav, Veleslav(a), Mieczyslav, Vidoslav, Předislav, Zlatislava, Desislava, Kvitoslava, Cvetislava, Wislava, Preslava.*

¹⁹ Substantivum je hravou narážkou na ekonomickou politiku *Bohuslava Sobotky*.

²⁰ V korpusu Syn v8 je v kontextu (...) na hřišti **Stejslav** Hýskov. Jde o název fotbalového klubu. Na internetu je doloženo i antroponymum: (...) *1185 vlastnil Stejslav z Nitrianskej Blatnice* (...) (Koplotovce, 2020).

²¹ Jedná se o ukrajinskou námořní základnu na Krymu.

²² Jde to název vinice, tedy anoikonymum.

²³ V korpusu Syn v8 jsou doložena nejen antroponyma mužského rodu *Předslav*, ale i oikonymum ženského rodu: (...) *mladík v malé obci Předslava na Klatovsku* (...).

²⁴ Viz <http://www.polyslav-as.org/en/about> a také <http://nase-rec.ujc.cas.cz/archiv.php?art=7897>. Na internetu lze najít i další užití: (...) *že císař rakouský jest jediným pravdivým vládcem Mnohoslovanem (Polyslavem), an nejmíce národův a národkův, jazykův a nářečí žezlem svým chrání* (...).

Homonymní tvary proprií zakončených na *-slav/-slava*

Pro automatickou morfologickou analýzu není ovšem důležité jen pokrytí slovníku. Kvalitní anotace je neméně závislá na přesnosti desambiguace. Homonymní tvary antroponym zakončených na *-slav/-slava* a oikonym zakončených na *-slav* představují velký problém pro spolehlivou desambiguaci.

Jestliže slovní tvar začíná velkým písmenem a končí na *-slav* (formulováno v jazyce CQL `word="[:upper:].*slav"`), pak mohou nastat tyto případy lemmatizace a morfologické interpretace vyjádřené v podobě značek užívaných při značkování korpusů ČNK:

- a) lemma=".*slav" & tag="NNFS[14].*"
- b) lemma=".*slav" & tag="NNMS1.*"
- c) lemma=".*slava" & tag="NNFP2.*"

Jestliže slovní tvar začíná velkým písmenem a končí na *-slava* (`word="[:upper:].*slava"`), pak mohou nastat tyto případy:

- d) lemma=".*slav" & tag="NNMS[24].*"
- e) lemma=".*slava" & tag="NNFS1.*"

Jestliže slovní tvar začíná velkým písmenem a končí na *-slavy* (`word="[:upper:].*slavy"`), pak mohou nastat tyto případy:

- f) lemma=".*slav" & tag="NNMP[47].*"
- g) lemma=".*slava" & tag="NNFS2|NNFP[145].*"

Jestliže slovní tvar začíná velkým písmenem a končí na *-slavu* (`word="[:upper:].*slavu"`), pak mohou nastat tyto případy:

- h) lemma=".*slav" & tag="NNMS[36].*"
- j) lemma=".*slava" & tag="NNFS4.*"

Konkrétní případ homonymie uvnitř skupiny proprií zakončených na *-slav/-slava* je přehledně představen v následující tabulce 1.

Zajímavý doklad tvarové homonymie nacházíme v následujícím textu: *Den s Mírou ve Veltrusech sobota 10. března Veltruští zámečtí zvou všechny **Mirolav**y, Míry i Mirky, aby oslavili svůj svátek poněkud netradičně. Pokud jste **nositelem/nositelkou** jména **Mirolav/Mirolava**, přijďte do Veltrus a vydejte se na komentovanou prohlídku s názvem *Váhy, míry, závaží na zámku vám ukáží* (...). Tvar **Mirolav**y je co do gramatického rodu nedesambiguovatelný, neboť označuje osoby obou pohlaví. V desambiguační praxi užívané při anotování korpusů se v analogických případech dává přednost hierarchicky neutrálnější interpretaci (maskulinum životné).*

S ohledem na rozsah článku se budeme nadále zabývat pouze problematikou homonymních tvarů zakončených na *-slav/-slava*, tedy případy a) až e).

Tabulka 1

lemma=	"[:upper:].*slav" &	"[:upper:].*slav" &	"[:upper:].*slav" &	"[:upper:].*slava" &	"[:upper:].*slava" &	"[:upper:].*slava" &
tag=	„NNMS1.*“	„NNFS1.*“	„NNFS4.*“	„NNFP2.*“		
word="[:upper:].*slav"	Jak informoval Miroslav Chlan (...)	(...) patří do katastru města Miroslav	V opačném směru bude provoz veden přes Miroslav do Hostěradic.	To už musí posoudit manželé a přátelé všech Miroslav . ²⁵		
tag=	„NNMS2.*“	„NNMS4.*“		„NNFS1.*“		
word="[:upper:].*slava"	Černobílá kolekce fotografií Miroslava Kolbasy.	Jmenovitě jde o Miroslava Tarnóczyho.		(...) vedoucí muzea Miroslava Štýbrová.		
tag=	„NNMP4.*“	„NNMP7.*“		„NNFS2.*“	„NNFP1.*“	!“NNFP4.*“
word="[:upper:].*slavy"	(...) budou muset schovat své Miroslavy , Kalouska a Šloufa (...)	(...) hraje spolu s Miroslavy Šimandly – otcem a synem (...)			(...) bylo nás tam pět rodiček a z toho dvě Miroslavy ²⁶ úplně stejného věku (...)	Kdybychom měli vyjmenovat všechny Miroslavy , ²⁷ které se nějak proslavily, (...)
tag=	„NNMS3.*“	„NNMS6.*“			„NNFS4.*“	
word="[:upper:].*slavu"	(...) lidé podobní Miroslavu Skalickému (...)	(...) o Miroslavu Šloufovi od něj neuslyšíme (...)			(...) šefku komory Miroslavu Němcovou (...)	

Zběžnou rešerší v korpusu Syn v8 (dotazem na slovní tvar začínající velkým písmenem a končící na *-slav* nebo *-slava* s vyloučením apelativního substantiva *oslava*²⁸ [word="[:upper:].*(slav|slava)" & lemma!="oslava"]) získáme celkem

²⁵ Uvedený doklad nepochází z korpusu Syn v8, je upravený pro názornou představu. Dokladů v korpusu Syn v8 je poměrně málo, ale vyskytují se: (...) *manželé a přátelé všech **Drahoslav**, **Drahan** a **Drahušek**; (...) **Patronkou všech **Zdišlav**.***

²⁶ Uvedený doklad nepochází z korpusu Syn v8, je upravený pro názornou představu. Ve skutečnosti se v korpusu Syn v8 vyskytuje takto: (...) *bylo nás tam pět rodiček a z toho dvě **Jaroslavy** úplně stejného věku.*

²⁷ Doklad jsme jednoduše v celém korpusu Syn v8 nedokázali najít, což nevylučuje jeho výskyt. V korpusu Araneum jsme našli jeden relevantní doklad (*Kdybychom měli vyjmenovat všechny **Jaroslavy**, které se nějak proslavily, trvalo by to do **Silvestra**.*), který jsme pro názornost upravili.

²⁸ Apelativum *oslava* je homonymní s hydronymem *Oslava*. Tuto homonymii automatická morfologická analýza zahrnuje, a sice prostřednictvím dvou lemmat (s velkým a malým počáteč-

3 716 726 výskytů tvarů sledovaných proprií. V desambiguaci není na první pohled mnoho chyb, nicméně zapátráme-li u jednotlivých slov, chyby se začnou objevovat.²⁹ Z tohoto důvodu jsme se zaměřili na případy, které by bylo možné správně desambiguovat pomocí pravidel. Pravidla lze opřít o grafické, formální, morfosyntaktické a lexikální vlastnosti kontextu, v němž se slovní tvary končící na *-slav/-slava* vyskytují v korpusech.

Kolokace a víceslovná pojmenování, která lze zařadit do databáze LEMUR³⁰

Poměrně dobře by bylo možné vylepšit desambiguaci v případech, kdy proprium (antroponymum) zakončené na *-slav* je v kontextu příjmení významné osoby, respektive případy, kdy se slovo zakončené na *-slav/-slava* vyskytuje bezprostředně před slovním tvarem s velkým počátečním písmenem, které lze interpretovat jako příjmení.

Typickým případem je *Bohuslav Martinů*. V korpusu Syn v8 je celkem 2 200 výskytů spojení **Bohuslav Martinů**. V 1 903 případech je tvar *Bohuslav* interpretován jako životné maskulinum v nominativu singuláru. 297 má interpretaci jako femininum (jak oikonymum s lemmatem *Bohuslav*, tak antroponymum *Bohuslava*). V témže korpusu je celkem 16 976 výskytů spojení **Bohuslava Martinů**. V 608 případech je tvar *Bohuslava* interpretován jako nominativ singuláru feminina *Bohuslava*.

V případě spojení **Jaroslav Seifert** jsou mylně jako tvary oikonyma interpretovány pouze dva doklady. V případě spojení **Jaroslava Seiferta** je mylně jako tvar feminina *Jaroslava* interpretován jeden doklad. V případě spojení **Miroslav Horníček** jsou všechny tvary interpretovány správně. V případě spojení **Miroslava Horníčka** je mylně jako tvar feminina *Miroslava* interpretován jeden doklad.

ním písmenem), tvarová homonymie není problém, neboť tvarosloví apelativa a propria se neliší. Problémem nepříliš spolehlivé desambiguace lemmatu *oslava/Oslava* se na tomto místě nebudeme zabývat.

²⁹ Například vyfiltrujeme-li případy, kdy je tvar *Jaroslav* interpretován jako – oikonymum ženského rodu, narazíme dosti často na kolokace typu *manželé Dana a Jaroslav/Jaroslav/NNFS1 -----A----- Stodolovi*, které jsou se zvolenou interpretací naprosto v rozporu, podobně je tomu i v kolokacích propria *Bohuslav Martinů*, *Bohuslav Sobotka* atd.

³⁰ Databáze víceslovných lexikálních jednotek LEMUR (Multiword Expression Lexical Database) vzniká péčí Ústavu teoretické a počítačnické lingvistiky FF UK a Ústavu Českého národního korpusu FF UK a využívá se při desambiguaci korpusů Českého národního korpusu, viz Jelínek a kol., 2018. Doplněním uvedeného typu víceslovných jednotek by se vyřešil i problém homonymie a chyb v desambiguaci tvarů zakončených na *-slavu*, které jsme v tomto článku ponechali stranou.

Dotazy:

[word="[:upper:].*slava" & word!="(Bratislava|Oslava)" & tag!="..[MI].*"]
[word="[:upper:].*([ay]ého)"]

[word="[:upper:].*slava" & word!="(Bratislava|Oslava)" & tag!="..[MI].*"]
[word="[:upper:].*([ay]ého)"]

Lze získat další kandidáty.

Dále jsme testovali dotaz na tvar **slava* v kontextu proprií příjmení na **ová*: [word="[:upper:].*slava" & tag!="..FS1.*"][word="[:upper:].*ová"]. Lze jím získat 13 086 dokladů, z toho 13 085 chybně desambiguovaných proprií zakončených na *-slava*. Jediný správně označovaný doklad je příjmení *Čáslava* (např.: *Snímky: Karel Čáslava Nová dětská hřiště jsou již v provozu*).³¹ Jak jsme navrhli výše, ve slovníku by nemělo být uvedeno antroponymum ženského rodu *Čáslava*, ale spolu s příjmením *Čáslava* pouze maskulinum *Čáslav* a femininum ženského rodu *Čáslav*. Při desambiguaci by tedy stačilo zohlednit jednoduché kontextové pravidlo a odstranit tak další tisícovku chyb.

Frekvenční distribuce takto získaných dvojic slovních tvarů je zdrojem dalších adeptů na spojení rodné jméno + příjmení, které lze uložit do databáze *LEMUR*.³²

Podobně lze postupovat i u oikonym jako *Mladá Boleslav*, *Stará Boleslav*.

Dotazem [word="(Mlad(á|ou))(Star(á|ou))"][word="Boleslav" & tag!="..FS[14].*"] lze dohledat 46 chybně desambiguovaných tvarů.³³

Návrh pravidel desambiguace proprií zakončených na *-slav/-slava* v koordinovaných skupinách

Všimli jsme si, že se antroponyma zakončená na *-slav/-slava* velice často vyskytují v kontextu výrazu *manžel[ée]/rodiče*, který je následován souřadně spojený-

³¹ Případy „falešných syntaktických vztahů“ typu: *Miroslava Kalousková vypráskala, že si nevzal roušku (...)* se v korpusu Syn v8 nevyskytly.

³² Většina uvedených vyhledaných příjmení, v jejichž kontextu se vyskytují chyby v desambiguaci proprií zakončených na *-slav*, mohou být i apelativa (adjektivum *hořejší*, apelativní substantiva *krček*, *doubrava*, *hlavnička*, *sobotka* atd.). Z toho důvodu se nelze spoléhat na značkování proprií ve slovníku *MorfFlex*, neboť i když jsou příjmení uložena jako propria, jejich značkování je předmětem desambiguace.

³³ Testovali jsme rovněž dotaz na další víceslovná oikonyma [word="(Velk(á|ou))(Mal(á|ou)|Horní|Dolní)"][word="*slav" & tag!="..FS[14].*"], ale relevantní doklady jsme nenašli. Rovněž jsme testovali dotaz na případy potenciálních víceslovných oikonym s prvním slovem *Mladá* a *Stará*, s vyloučením druhého slova *Boleslav* (dotaz: [word="(Mlad(á|ou))(Star(á|ou))"][word="*slav" & word!="Boleslav" & tag!="..FS[14].*"]). Kromě 83 překlepů ve slově *Boleslav* jsme našli pouze 4 nerelevantní doklady (*ředitel Domova Mladá Miloslav/Miloslav/NNMSI-----A----- Müller*).

mi křestními jmény a za nimi stojí vlastní jméno zakončené na *-ov[iy]*: *manželé Marie a Jaroslav Stejskalovi*.

Konstatovali jsme, že ne vždy je desambiguace antroponym zakončených na *-slav/-slava* správně. Velmi často je tvar maskulina antroponyma zakončeného na *-slav* interpretován jako tvar feminina ženského rodu zakončeného na *-slav*, méně často pak je tvar antroponyma feminina zakončeného na *-slava* interpretován jako tvar antroponyma mužského rodu zakončeného na *-slav* a naopak.

Navrhli jsme kontextová pravidla pro desambiguaci. Některá pravidla počítají s pravopisně správnými tvary v okolí analyzovaného tvaru. Korpusy ovšem zahrnují i texty s pravopisnými chybami. Těmi se v následující analýze systematicky nezabýváme, pouze na ně okrajově upozorňujeme.

Poznamenáváme, že pravidla týkající se dvojice s antroponymy zakončenými na *-slav/-slava* se týkají i dvojic s antroponymy *Václav/Václava* (ale i dalších podobných, jako například **mil/ *mila*, **mír/ *míra*, tedy *Bohumil/Bohumila*, *Jaromír/Jaromíra* atd.). Vzhledem k rozsahu tohoto článku je ponecháme stranou.

Pravidla zahrnující v podmínce tvary *rodiče/manžel[ée]* je třeba nazírat s následující výhradou. Tvar *rodiče* má homonymní tvar pro nominativ a akuzativ, navíc může být rozvinut postponovaným (i koordinovaným) přívlastkem v genitivu. Máme-li tvar *rodiče* následovaný dvojicí potenciálních proprií, mezi nimiž je slučovací spojka *a*, pak můžeme předpokládat tyto interpretace: *rodiče Stanislava a Miroslava* ← (*Stanislavov[iy] a Miroslavov[iy]*) *rodiče* | (*Stanislavovi rodiče a Miroslava/Fem.*) | (*Stanislavovy rodiče a Miroslava/Mask.*) | (*rodiče/Nom. Stanislava/Fem. a Miroslava/Fem.*) | (*rodiče/Ak. Stanislava/Mask. a Miroslava/Mask.*). Poslední dvě varianty je možné vykládat a) jako stejnopohlavní rodičovský pár a b) jako nepár. Na tyto případy upozorňujeme, přestože některé z interpretací jsou velmi málo pravděpodobné a pravidla, která navrhuje, se takovými případy nezabývají.

Pravidla jsou formulována tak, aby se neopírala o automatickou morfologickou analýzu tvarů v kontextu slova zakončeného na *-slav/-slava*. V pravidlech počítáme s variantami kombinací antroponym žen a mužů, přičemž vždy jeden člen dvojice má explicitně tvar zakončený na *-slav/-slava*. V pravidlech 1, 1a předpokládáme, že před *a* stojí femininum, nicméně se může vyskytovat i maskulinum a za *a* maskulinum končící na *-slav*. V pravidlech 2, 2a předpokládáme, že před *a* stojí maskulinum končící na *-slav* a po *a* stojí femininum, nicméně se může vyskytovat i maskulinum. V pravidlech 3, 3a, předpokládáme, že před *a* stojí femininum zakončené na *-slava* a za *a* maskulinum, kdežto v pravidlech 4, 4a je pořadí opačné. V pravidlech 5, 5a předpokládáme, že před *a* stojí maskulinum (tvar na *-slava*) v akuzativu a za *a* femininum, kdežto v pravidlech 6, 6a je pořadí opačné. Je patrné, že preferované slovosledné varianty (v koordinované skupině je na prvním místě jméno ženy a na druhém jméno muže – odraz zdvořilostní strategie) jsou početnější, zachytíme v nich tudíž většinou i více chyb v desambiguaci.

Pravidlo 1. Jestliže v levém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slav*

[word="[:upper:].*slav"]

je slovní tvar s počátečním velkým písmenem následovaný spojkou *a*

<-2,-1> **[word="[:upper:].*"] [lc="a"]**

a v pravém kontextu následuje slovní tvar s velkým počátečním písmenem zakončeným na *-ovi*

<1,1> je **[word="[:upper:].*ovi"]**,

pak platí, že slovní tvar končící na *-slav* je maskulinum životné v nominativu singuláru a jeho lemma začíná velkým písmenem

[word="[:upper:].*slav" & tag="NNMS1.*" & lemma="[:upper:].*slav"].

Tomuto pravidlu odporují výsledky desambiguace tvaru propria zakončeného na *-slav* v korpusu Syn v8 v 1417 případech, viz CQL dotaz 1 v tabulce 2 níže. 2 251 dokladů je označováno správně.

Tabulka 2

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
1.	[word="[:upper:].*"] [lc="a"] [word="[:upper:].*slav" & tag!=".MS1.*"] [word="[:upper:].*ovi"]	1417 chyb (chybovost 39 %)	<i>Profesionální pěstouni Jana a Jaroslav/Jaroslav/NNFSI</i> ----A---- Toušovi.

Pravidlo 1a. Jestliže v levém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slav*

[word="[:upper:].*slav"]

je slovní tvar *rodiče* nebo *manželé*, za ním tvar s počátečním velkým písmenem následovaný spojkou *a*

<-3,-1> **[lc=".*rodiče.*manželé"] [word="[:upper:].*"] [lc="a"]**

a v pravém kontextu nenásleduje slovní tvar s velkým počátečním písmenem zakončeným na *-ovi*

<1,1> **[word!="[:upper:].*ovi"]**,

pak velmi pravděpodobně³⁴ platí, že slovní tvar končící na *-slav* je maskulinum životné v nominativu singuláru a jeho lemma začíná velkým písmenem

[word="[:upper:].*slav" & tag="NNMS1.*" & lemma="[:upper:].*slav"].

³⁴ S velmi málo pravděpodobnou interpretací, jako např. *rodiče Martina a Jaroslav* ← *Martinovi rodiče, kteří jsou zároveň rodiči několika Jaroslav*, nepočítáme. Bereme v úvahu pouze pravděpodobnější interpretace (*rodiče Martina a Jaroslav* ← *rodičovský pár Martina/žena a Jaroslav/muž* | *Martinovi rodiče a Jaroslav/muž*).

Tomuto pravidlu odporují výsledky desambiguace tvaru propria zakončeného na *-slav* v korpusu Syn v8 v 212 případech, viz CQL dotaz 1a v tabulce 2a níže. 419 dokladů je označováno správně.

Tabulka 2a

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
1a.	[lc=".*rodičel.*manželé"] [word="[:upper:].*"] [lc="a"] [word="[:upper:].*slav" & tag!=".MS1.*"] [word!="[:upper:].*ovi"]	212 chyb (chybovost 34 %)	<i>Manželé Veronika</i> <i>a Jaroslav/Jaroslav/NNFS1-----A-----</i> <i>si synka odvezou domů do Poličky.</i>

Pravidlo 2. Jestliže v pravém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slav*

[word="[:upper:].*slav"]

je slovní tvar spojka *a*, za ní následuje slovní tvar s velkým počátečním písmenem následovaný slovním tvarem s velkým počátečním písmenem zakončeným na *-ovi* <1,3> [lc="a"] [word="[:upper:].*"] [word="[:upper:].*ovi"],

pak platí, že slovní tvar končící na *-slav* je maskulinum živočné v nominativu singuláru a jeho lemma začíná velkým písmenem

[word="[:upper:].*slav" & tag="NNMS1.*" & lemma="[:upper:].*slav"].

Tomuto pravidlu odporují výsledky desambiguace tvaru propria zakončeného na *-slav* v korpusu Syn v8 v 137 případech, viz CQL dotaz 2 v tabulce 3 níže. 2 049 dokladů je označováno správně.

Tabulka 3

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
2.	[word="[:upper:].*slav" & tag!=".MS1.*"] [lc="a"] [word="[:upper:].*"] [word="[:upper:].*ovi"]	137 chyb (chybovost 6 %)	(...) žili <i>Jaroslav/Jaroslav/NNFS1</i> <i>-----A-----</i> a Dana Stodolovi necečých <i>pět let v pronájmu</i>

Pravidlo 2a. Jestliže v levém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slav*

[word="[:upper:].*slav"]

je slovní tvar *rodiče* nebo *manželé*

<-1,-1> [lc=".*rodičel.*manželé"]

a v pravém kontextu následuje slovní tvar spojka *a*, za ní tvar s velkým počátečním písmenem, za nímž nenásleduje slovní tvar s velkým počátečním písmenem zakončený na *-ovi*

<1,3> **je** [lc="a"] [word="[:upper:].*"] [word!="[:upper:].*ovi"],

pak velmi pravděpodobně³⁵ platí, že slovní tvar končící na *-slav* je maskulinum životné v nominativu singuláru a jeho lemma začíná velkým písmenem

[word="[:upper:].*slav" & tag="NNMS1.*" & lemma="[:upper:].*slav"].

Tomuto pravidlu odporují výsledky desambiguace tvaru propria zakončeného na *-slav* v korpusu Syn v8 v 1 případě, viz CQL dotaz 2a v tabulce 3a níže. 128 dokladů je označováno správně.

Tabulka 3a

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
2a.	[lc=".*rodiče . *manželé"] [word="[:upper:].*slav" & tag!=".MS1.*"][lc="a"] [word="[:upper:].*"] [word!="[:upper:].*ovi"]	1 chyba (chybovost 0,8 %)	(...) <i>se bez prostředků ocitli i manželé Dobroslav/Dobroslava/NNFP2</i> -----A----- a Kateřina Novotná.

Pravidlo 3. Jestliže v pravém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slava*

[word="[:upper:].*slava"]

je slovní tvar spojka *a*, za ní následuje slovní tvar s velkým počátečním písmenem, který nekončí na samohlásky *a, e, ě, i, o, u, y, á, é, ó, ů, ý*, a za ním následuje slovní tvar s velkým počátečním písmenem zakončený na *-ovi*

<1,3> [lc="a"] [word="[:upper:].*[^aečiouyáéóůý"] [word="[:upper:].*ovi"]

pak platí, že slovní tvar na *-slava* je femininum v nominativu singuláru a jeho lemma začíná velkým písmenem

[word="[:upper:].*slava" & tag="NNFS1.*" & lemma="[:upper:].*slava"].

Tomuto pravidlu odporují výsledky desambiguace tvaru propria zakončeného na *-slava* v korpusu Syn v8 v 139 případech, viz CQL dotaz 3 v tabulce 4 níže. 614 dokladů je označováno správně.

Tabulka 4

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
3.	[word="[:upper:].*slava" & tag!=".FS1.*"][lc="a"] [word="[:upper:].*[^aečiouyáéóůý"] [word="[:upper:].*ovi"]	139 chyb (chybovost 18 %)	Vystavují <i>Jaroslava/Jaroslav/NNMS4</i> -----A----- a Jan Solovjevovi (...)

³⁵ S velmi málo pravděpodobnými interpretacemi, jako např. *rodiče Jaroslav a Martina* ← (*rodiče několika Jaroslav, kteří jsou zároveň Martinovými rodiči* | (*rodiče/ak. několika Jaroslav a Martina/mask. ak. | rodiče/nom. několika Jaroslav a Martina/fem. nom.*) nepočítáme.

Pravidlo 3a. Jestliže v levém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slava*

[word="[:upper:].*slava"]

je slovní tvar *rodiče* nebo *manželé*

<-1,-1> **[lc=".*rodiče].*manželé"]**

a v pravém kontextu je slovní tvar spojka *a*, za ní následuje slovní tvar s velkým počátečním písmenem, který nekončí na samohlásky *a, e, ě, i, o, u, y, á, é, ó, ů, ý*, a za ním nenásleduje slovní tvar s velkým počátečním písmenem zakončený na *-ovi* <1,3> **[lc="a"] [word="[:upper:].*[^aeěiouyáéóůý"] [word!="[:upper:].*ovi"]**, pak velmi pravděpodobně³⁶ platí, že slovní tvar končící na *-slava* je femininum v nominativu singuláru a jeho lemma začíná velkým písmenem

[word="[:upper:].*slava" & tag="NNFS1.*" & lemma="[:upper:].*slava"].

Tomuto pravidlu odporují výsledky desambiguace tvaru propria zakončeného na *-slava* v korpusu Syn v8 v 65 případech, viz CQL dotaz 3a v tabulce 4a níže.³⁷ 99 dokladů je označováno správně.

Tabulka 4a

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
3a.	[lc=".*rodiče].*manželé" [word="[:upper:].*slava" & tag!=".FS1.*"] [lc="a"] [word="[:upper:].*[^aeěiouyáéóůý"] [word!="[:upper:].*ovi"]	65 chyb (chybovost 40 %)	<i>Domů do Chrustric si synka odvezou manželé Drahoslava/Drahoslav/NNMS2-----A----- a Miroslav.</i>

Pravidlo 4. Jestliže v levém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slava*

[word="[:upper:].*slava"]

je slovní tvar s velkým počátečním písmenem, který nekončí na samohlásky *a, e, ě, i, o, u, y, á, é, ó, ů, ý*, následovaný slovním tvarem spojkou *a*

<-2,-1> **[word="[:upper:].*[^aeěiouyáéóůý"] [lc="a"]**

³⁶ S velmi málo pravděpodobnými interpretacemi, jako např. *Rodiče Miroslava a Jindřich si syna odvezli domů (...)* čti: *Miroslavovi rodiče a nějaký Jindřich si syna odvezli domů (...)*, nepočítáme.

³⁷ Pravidla 3, 3a, 4, 4a jsou formulována tak, že zachycují pouze chyby v kontextu maskulin, která v nominativu singuláru končí na souhlásku, nebo na *-i* (tedy nikoli jména jako *Ivo, René, Arne, Ota, Harry, Joe, Otto, Henry, José...*, ale ani hypokoristika jako *Jirka, Kája, Standa, Pepa, Franta, Ondra...*, přičemž mnohá z nich jsou homonymní s femininy – *Vlasta, Stáňa* atd.). Například v dokladu: *Raduji se z něj manželé Zdislava/Zdislav/NNMS2-----A----- a Yussine z Ústí nad Orlicí (...)*, nebude chyba desambiguace tvaru *Zdislava* pravidlem zachycena.

a v pravém kontextu je slovní tvar s velkým počátečním písmenem zakončený na *-ovi*

<1,1> je [word="[:upper:].*ovi"]

pak platí, že slovní tvar zakončený na *-slava* je femininum v nominativu singuláru a jeho lemma začíná velkým písmenem

[word="[:upper:].*slava" & tag="NNFS1.*" & lemma="[:upper:].*slava"].

Tomuto pravidlu odporují výsledky desambiguace tvaru propria zakončeného na *-slava* v korpusu Syn v8 v 21 případech, viz CQL dotaz 4 v tabulce 5 níže. 308 dokladů je označováno správně.

Tabulka 5

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
4.	[word="[:upper:].*[^aeěiouyáéóůý]" lc="a"] [word="[:upper:].*slava" & tag!=".FS1.*"] [word="[:upper:].*ovi"]	21 chyb (chybovost 6%) 1 chyba ³⁸	<i>Zasloužili se o to manželé Vítězslav a Radoslava/Radoslav/NNMS2</i> <i>-----A----- Doubravovi (...) uviděla své rodiče, Karin a Rastislava/Rastislav/NNMS4-----A----- Jelínkovi</i>

Pravidlo 4a. Jestliže v levém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slava*

[word="[:upper:].*slava"]

je slovní tvar *rodiče* nebo *manželé*, za ním následuje slovní tvar s velkým počátečním písmenem, který nekončí na samohlásky *a, e, ě, i, o, u, y, á, é, ó, ů, ý, a* za ním následuje slovní tvar spojka *a*

<-3,-1> [lc=".*rodiče.*manželé"] [word="[:upper:].*[^aeěiouyáéóůý]" lc="a"]

a v pravém kontextu nenásleduje slovní tvar s velkým počátečním písmenem zakončený na *-ovi*

<1,1> je [word!="[:upper:].*ovi"],

pak velmi pravděpodobně platí, že slovní tvar zakončený na *-slava* je femininum v nominativu singuláru a jeho lemma začíná velkým písmenem

[word="[:upper:].*slava" & tag="NNFS1.*" & lemma="[:upper:].*slava"].

³⁸ Pravidlo generuje v jednom případě chybu. Selhání není důsledkem restriktivní formulace pravidel, která počítají pouze s antroponymy mužského rodu, jejichž nominativy singuláru končí na souhlásku nebo na *i*, a s antroponymy ženského rodu, jejichž akuzativ končí na *-u* nebo *-i*, tedy nikoliv s neskloňnými femininy končícími na souhlásku, či jiný vokál (*Ráchel, Karin, Rút, Ellen, Ester, Dagmar, Chloé* atd.). Příčinou selhání pravidla je chyba ve shodě v analyzovaném textu. Pokud by v dokladu nebyla gramatická chyba a zněl by: *uviděl své rodiče, Karin a Rastislava Jelínkovi* (...), pravidlo by jej ignorovalo.

Tomuto pravidlu neodporují žádné výsledky desambiguace tvaru propria zakončeného na *-slava* v korpusu Syn v8, viz pravidlo 4a v tabulce 5a níže. 36 dokladů je označováno správně.

Tabulka 5a

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
4a.	[lc=".*rodičel.*manželé"] [word="[:upper:]]. *[^aečiouyáéóůý"] [lc="a"] [word="[:upper:]]. *slava" & tag!=".FS1.*" [word!="[:upper:]].*ovi"]	0 chyb (chybovost 0 %)	----- 39

Pravidlo 5. Jestliže v levém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slava*

[word="[:upper:]].*slava"

je slovní tvar začínající velkým písmenem končící na *-u* nebo *-i*,⁴⁰ za ním následuje slovní tvar spojka *a*

<-2,-1> **[word="[:upper:]].*[ui]" [lc="a"]**

a v pravém kontextu následuje slovní tvar s velkým počátečním písmenem zakončený na *-ovy*

<1,1> **[word="[:upper:]].*ovy"**,

pak velmi pravděpodobně⁴¹ platí, že slovní tvar na *-slava* je maskulinum životné v akuzativu singuláru, jeho lemma začíná velkým písmenem a končí na *-slav*

[word="[:upper:]].*slava" & tag="NNMS4.*" & lemma="[:upper:]].*slav".

Tomuto pravidlu odporují výsledky desambiguace tvaru propria zakončeného na *-slava* v korpusu Syn v8 v 25 případech, viz CQL dotaz 5 v tabulce 6 níže. 59 dokladů je označováno správně.

³⁹ V korpusu Araneum Bohemicum Maximum jsme vyhledali jeden doklad, který ukazuje na chybu v desambiguaci (*Posledními majiteli byli manželé Bohumír a Bohuslava/Bohuslav/NNMS2-----A---- Škodovy*). Pravidlo by chybu opravilo, přestože je v kontextu gramatická chyba, neboť tvar, ve kterém je chyba, není zahrnut do podmínky.

⁴⁰ Restriktivní formulace pravidel 5, 5a, 6, 6a, která počítají pouze se sklonnými antroponymy ženského rodu, jejichž akuzativ singuláru končí na *-u* nebo *-i*, a nikoliv s nesklonnými antroponymy ženského rodu (*Ráchel, Karin, Rút, Ellen, Ester, Dagmar, Noemi, Lulu, Veru, Kiki, Cony...*), je diskutabilní, viz následující poznámka.

⁴¹ Pravidla 5, 6 mohou generovat chyby v případech, že se v kontextu tvaru zakončeného na *-slava* bude vyskytovat koordinované femininum nesklonné končící na *-i/-u* (např. *Noemi, Lulu*, popř. hypokoristika *Veru, Maru, Kristi...*). V takovém případě nemusí být tvar zakončený na *-slava* [tag="NNMS4.*" & lemma="[:upper:]].*slav", ale musí (*Noemi a Jaroslava Královy*),

Tabulka 6

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
5.	[word="[:upper:].*ui"] [lc="a"][word="[:upper:].*slava" & tag!=".MS4.*"] [word="[:upper:].*ovy"]	25 chyb (chybovost 30 %)	<i>A jak nepochválit zkušené choreografy Jitku a Ladislava/Ladislava/NNFSI -----A----- Košíkovy</i>

Pravidlo 5a. Jestliže v levém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slava*

word="[:upper:].*slava"

je slovní tvar *rodiče* nebo *manžele*, za ním následuje slovní tvar s velkým počátečním písmenem, který končí na samohlásky *u* nebo *i*, a za ním následuje slovní tvar spojka *a*

<-3,-1> [lc=".*rodič|. *manžele"] [word="[:upper:].*ui"] [lc="a"]

a v pravém kontextu nenásleduje slovní tvar s velkým počátečním písmenem zakončený na *-ovy*

<1,1> [word!="[:upper:].*ovy"]

pak velmi pravděpodobně⁴² platí, že slovní tvar zakončený na *-slava* je maskulinum životné v akuzativu singuláru, jeho lemma začíná velkým písmenem a končí na *-slav*

[word="[:upper:].*slava" & tag="NNMS4.*" & lemma="[:upper:].*slav"].

Tomuto pravidlu odporují výsledky desambiguace tvaru *propria* zakončeného na *-slava* v korpusu Syn v8 v 27 případech, viz CQL dotaz 5a v tabulce 6a níže. 96 dokladů je označováno správně.

Tabulka 6a

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
5a.	[lc=".*rodič . *manžele"] [word="[:upper:].*ui"] [lc="a"][word="[:upper:].*slava" & tag!=".MS4.*"] [word!="[:upper:].*ovy"]	27 chyb (chybovost 22 %)	<i>Eliška Kaplanová je první radostí pro rodiče Vendulu a Jaroslava/Jaroslava/NNFSI -----A----- z Rájce u Chocně.</i>

může (*Miroslava a Lulu Baltazarovy*) být [tag="NNFS1.*" & lemma="[:upper:].*slava"]. Na takové případy jsme ovšem v korpusu Syn v8 nenarazili.

⁴² Pravidlo vylučuje nepravděpodobné leč možné čtení v případech nesklonných feminin: *rodiče Noemi a Jaroslava* ← *Noeminy rodiče a Jaroslava*, *rodiče Veru a Jaroslava* ← *Veruniny rodiče a Jaroslava*.

Pravidlo 6. Jestliže v pravém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slava*

[word="[:upper:].*slava"]

je slovní tvar spojka *a*, za ní následuje slovní tvar s velkým počátečním písmenem, který končí na samohlásky *u* nebo *i*, a za ním následuje slovní tvar s velkým počátečním písmenem zakončený na *-ovy*

<1,3> [lc="a"] [word="[:upper:].*[ui]"] [word="[:upper:].*ovy"]

pak velmi pravděpodobně⁴³ platí, že slovní tvar na *-slava* je maskulinum životné v akuzativu singuláru, jeho lemma začíná velkým písmenem a končí na *-slav*

[word="[:upper:].*slava" & tag="NNMS4.*" & lemma="[:upper:].*slav"]

Tomuto pravidlu odporují výsledky desambiguace tvaru propria zakončeného na *-slava* v korpusu Syn v8 ve 4 případech, viz CQL dotaz 6 v tabulce 7 níže. 40 dokladů je označováno správně.

Tabulka 7

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
6.	[word="[:upper:].*slava" & tag!=".MS4.*"] [lc="a"] [word="[:upper:].*[ui]"] [word="[:upper:].*ovy"]	4 chyby (chybovost 9 %)	(...) <i>dopravila eskorta Jaroslava/Jaroslava/NNFS1</i> ----- <i>A</i> ----- <i>a Danu Stodolovy</i> (...)

Pravidlo 6a. Jestliže v levém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slava*

[word="[:upper:].*slava"]

je slovní tvar *rodiče* nebo *manžele*

<-1,-1> [lc=".*rodiče|. *manžele"]

a v pravém kontextu následuje slovní tvar spojka *a*, za kterým následuje slovní tvar s velkým počátečním písmenem, který končí na samohlásky *u* nebo *i*, po kterém nenásleduje slovní tvar s velkým počátečním písmenem zakončený na *-ovy*

<1,3> [lc="a"] [word="[:upper:].*[ui]"] [word!="[:upper:].*ovy"],

pak velmi pravděpodobně⁴⁴ platí, že slovní tvar končící na *-slava* je maskulinum životné v akuzativu singuláru, jeho lemma začíná velkým písmenem a končí na *-slav*

[word="[:upper:].*slava" & tag="NNMS4.*" & lemma="[:upper:].*slav"].

⁴³ Viz poznámka výše, doklady typu *Miroslava a Lulu Baltazarovy přišly včas* (...).

⁴⁴ S přípustnou interpretací jako např.: *bylo pro rodiče Rostislava a Věru překvapením* ← *bylo pro Rostislavovy rodiče a Věru překvapením*, nepočítáme.

Tomuto pravidlu odporují výsledky desambiguace tvaru propria zakončeného na *-slava* v korpusu Syn v8 v 1 případě, viz CQL dotaz 6a v tabulce 7a níže. 12 dokladů je označováno správně.

Tabulka 7a

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
6a.	[lc=".*rodičel.*manžele"] [word="[:upper:].*slava" & tag!="..MS4.*"] [lc="a"] [word="[:upper:].*[ui]" & word!="[:upper:].*ovy"]	1 chyba (chybovost 8 %)	<i>Měli jsme u nás manžele Jaroslava/Jaroslav/NNMS2-----A----- a Aničku Lehkých z Hrobčic.</i> ⁴⁵

Ověření možnosti pravidel desambiguace proprií zakončených na *-slav* v kontextu vybraných substantiv označujících místo/sídlo

V rámci hledání způsobů zlepšení desambiguace jsme se snažili zmapovat situaci spojení lemmat označujících místo/sídlo (např. *obec, město, lokalita, vesnice, městys, oblast, zámek, ves, vesnička, městečko, víska, tvrz, hrad*) a tvarů zakončených na *-slav*.

V mnoha případech jsme našli správně desambiguovaná maskulina, neboť před lemmaty s významem místo předcházela substantiva typu *starosta* a po tvaru končícím na *-slav* následovalo příjmení (např. *starosta obce Miroslav Kovář*). Na základě pozorování jsme přistoupili k formulaci následujícího pravidla.

Pravidlo 7. Jestliže v levém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slav*

[word="[:upper:].*slav"]

je slovní tvar lemmatu *obec, město, lokalita, vesnice, městys, oblast, zámek, ves, vesnička, městečko, víska, tvrz* nebo *hrad*

<-1,-1>

[lemma="obec|město|lokalita|vesnice|městys|oblast|zámek|ves|vesnička|městečko|víska|tvrz|hrad"]

a v pravém kontextu nenásleduje slovní tvar s velkým počátečním písmenem <1,1> **[word!="[:upper:].*"]**,

⁴⁵ Zvolená interpretace předpokládá dva *Jaroslavovy* manžele (tedy dva muže, kteří byli/jsou manželé dotyčného *Jaroslava*) a *Aničku*. Jde o interpretaci vysoce nepravděpodobnou, přestože je gramaticky (a někde třeba i právně) přípustná.

a zároveň pravý kontext neobsahuje lemma z následované slovním tvarem s velkým počátečním písmenem

<1,2> [lemma!="z"][word="[:upper:].*"]⁴⁶,

pak platí, že slovní tvar zakončený na *-slav* je femininum v nominativu singuláru, jeho lemma začíná velkým písmenem a končí na *-slav*

[word="[:upper:].*slav" & tag="NNFS1.*" & lemma="[:upper:].*slav"].

Tomuto pravidlu odporují výsledky desambiguace tvaru propria oikonyma zakončeného na *-slav* v korpusu Syn v8 ve 2370 případech, viz pravidlo 7 v tabulce 8 níže. 2 549 dokladů je označováno správně.

Tabulka 8

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
7.	[lemma="obec město lokalita vesnice městys oblast zámek ves vesnička městečko vískal tvrz hrad"] [word="[:upper:].*slav" & tag!=".F.*"] [word!="[:upper:].*"] N-filtr <0,1> [lemma="z"] [word="[:upper:].*"]	2370 chyb (chybovost 48 %)	(...) který se v minulých dnech v obci <i>Jaroslav/Jaroslav/NNMSI----A----</i> vloupal do plechové haly v areálu bývalého JZD (...)

Ověření možnosti pravidel desambiguace proprií zakončených na *-slav* v kontextu substantiv označujících funkce (*předseda, trenér, mluvčí...*)

Sledovali jsme desambiguaci tvarů zakončených na *-slav/-slava* ve spojení s názvy funkcí. Na základě pozorování dat jsme navrhli následující pravidlo.

Pravidlo 8. Jestliže v levém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slav*

[word="[:upper:].*slav"]

je slovní tvar lemmatu končící na řetězec *starosta, předseda*, nebo lemmatu *trenér, mluvčí*

<-1,-1> [lemma=".*starosta|.*předseda|trenér|mluvčí"]

a v pravém kontextu následuje slovní tvar s velkým počátečním písmenem

<1,1> [word="[:upper:].*"]

⁴⁶ Podmínka vyloučí správně desambiguované případy dokladů jako: *Domnělý zakladatel hradu Jaroslav/Jaroslav/NNMSI----A----* z Turnova měl sídlo věnovat (...); (...) přijede na zámek *Vratislav/Vratislav/NNMSI----A----* z Pernštejna (...); *Statui* věnoval obci *Jaroslav/Jaroslav/NNMSI----A----* ze Šternberka (...).

pak platí, že slovní tvar na *-slav* je maskulinum životné v nominativu singuláru, jeho lemma začíná velkým písmenem a končí na *-slav*

[word="[:upper:].*slav"&tag="NNMS1.*"&lemma="[:upper:].*slav"].

Tomuto pravidlu odporují výsledky desambiguace tvaru propria zakončeného na *-slav* v korpusu Syn v8 ve 166 případech, viz CQL dotaz 8 v tabulce 9 níže.⁴⁷ 142 249 dokladů je označováno správně.

Tabulka 9

CQL dotaz	dotaz na doklady odporující pravidlu	počet chyb (Syn v8)	příklad chyby v desambiguaci (Syn v8), který je porušením pravidla
8.	[lemma=".*starosta. *předseda trenér mluvčí"] [word="[:upper:].*slav" & tag!="..M.*"] [word="[:upper:].*"]	166 chyb (chybovost 0,1 %)	(...) řekl jejich mluvčí <i>Jaroslav/Jaroslav/NNFS4-----A-----</i> <i>Haid.</i>

Řada antroponym a řada oikonym jako další případy koordinace

Testovali jsme také pravidla na vylepšení desambiguace v případech, kdy proprium (antroponymum nebo oikonymum) zakončené na *-slav/-slava* stojí v kontextu dalších proprií. Takové řady lze definovat jako slovní tvary s počátečním velkým písmenem oddělené čárkami.

Pravidlo 9a. Jestliže v levém i pravém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slav*

[word="[:upper:].*slav"]

stojí tvar s počátečním velkým písmenem označovaný jako maskulinum neživotné, femininum, či neutrum oddělený od tvaru zakončeného na *-slav* čárkami

<-2,-2> [word="[:upper:].*" & tag="..[IFN].*"] [word="\","]

<2,2> [word="\","] [word="[:upper:].*" & tag="..[IFN].*"],

pak s velkou mírou pravděpodobnosti⁴⁸ platí, že slovní tvar končící na *-slav* je

⁴⁷ O nedostatečném pokrytí slovníku svědčí např. interpretace tvarů *Ludoslav/X.**, *Ljuboslav/X.**, *Vladyslav/X.** atd. V ostatních případech jsou tvary nerozpoznané morfologickou analýzou (značované jako *X.**) překlepy a spojená slova. Přehled o nerozpoznaných tvarech a překlpech je možné získat tak, že pomocí p-filtru vyhledáme [tag="X.*"] v intervalu <0,0>.

⁴⁸ Toto pravidlo neplatí 100%. To, že podmínka je formulována příliš široce, způsobí, že pravidlo opraví i tvar zakončený na *-slav*, správně desambiguovaný jako maskulinum životné. První příčinou je homonymie mezi apelativy a proprii: (...) 155 s. *Zahrádka, Miroslav/Miroslav/NNMS1-----A-----, Dogmata a živý literární proces* (...); Čs. obec sokolská, 1926 *ŠLAJER, Jaroslav/Jaroslav/NNMS1-----A-----, Husitské revoluční hnutí a husitská tradice* (...).

femininum v nominativu nebo akuzativu singuláru, jeho lemma začíná velkým písmenem a končí na *-slav*

[word="[:upper:].*slav" & tag="NNFS[14].*" & lemma="[:upper:].*slav"].

Stav desambiguace slovních tvarů s velkým počátečním písmenem, které končí na *-slav*, získaný dotazem

[word="\,"][word="[:upper:].*slav"][word="\,"]

p-filtr<-1,-1> [word="[:upper:].*" & tag="..[IFN].*"]

p-filtr <1,1> [word="[:upper:].*" & tag="..[IFN].*"]

z korpusu Syn v8⁴⁹ a výsledky ruční kontroly desambiguovaných výsledků i počty chyb opravených i generovaných navrženým pravidlem ukazuje následující tabulka 10a.

Tabulka 10a

počet tvarů ([word="[:upper:].*slav"]) a tag	je správně tag	má být tag	tag by navržené pravidlo opravilo	chyby neopravené navrženým pravidlem	chyby vnesené navrženým pravidlem
983/NNFS1.*	970/NNFS1.*	13/NNMS1.*	-	13	-
59/NNFS4.*	59/NNFS4.*	0	-	0	-
3/NNFP2.*	0	3/NNFS1.*	3	0	0
201/NNMS1.*	15	186/NNFS[14].*	201	0	15
6/X.*	0	3/NNFS[14].* 3/NNMS1.*	6	0	3
1/NNIS1.*	0	1/NNFS1.*	1	0	0
1/NNIP2.*	0	1/NNFS1.*	1	0	0

V případě tří chyb vnesených navrženým pravidlem by pomohlo doplnění lemmatu *Sukoslav* do slovníku *MorfFlex*, viz výše. Většinu (12 z 15) dalších případů chyb vnesených navrženým pravidlem a 13 chyb neopravených navrženým pravidlem

V obou případech je ve slovníku morfologického analyzátoru substantivum *Zahrádka* a *Šlajer* interpretováno jednak jako maskulinum životné – apelativum, jednak jako femininum – *zahrádka* nebo maskulinum neživotné – *šlajer*. Tyto interpretace byly chybně desambiguovány. Druhou příčinou je fakt, že antroponyma ženského rodu mohou stát v koordinované řadě s antroponymy mužského rodu: (...) *narodili se Justýna, Jakub, Anna, Miroslav/Miroslav/NNMS1-----A-----, Tereza, Adam a Kristýna* (...). Ve sledovaných datech jsou tyto případy v menšině, nicméně se objevují.

⁴⁹ (https://context.korpus.cz/freqs?ctxattrs=word%2Clemma%2Ctag&attr_vmode=mixed&pa gesize=40&refs=%3Ddoc.title&q=~1uWaZfjnzRD&viewmode=kwic&attrs=word%2C lemma%2Ctag&corpname=syn_v8&attr_allpos=all&fcrit=tag%2Fe%201%3C0&flimit=1 &freq_sort=fpage=1&ftt_include_empty=).

lem, by bylo možné odstranit, pokud bychom při formulaci pravidla využili informace o příslušnosti k různým typům proprií, které zachycuje slovník *MorfFlex*.⁵⁰ Pokud v kontextu <-2,-2> a <2,2> od tvaru zakončeného na *-slav* stojí tvary s počátečním velkým písmenem označované jako feminina_ geografické názvy oddělené od tvaru zakončeného na *-slav* čárkami, jde o oikonymum femininum končící na *-slav*.⁵¹ Pokud v kontextu <-2,-2> a <2,2> od tvaru zakončeného na *-slav* stojí tvary s počátečním velkým písmenem označované jako feminina_ jména osob oddělené od tvaru končícího na *-slav* čárkami, jde o antroponymum maskulinum životné zakončené na *-slav*.⁵²

Pravidlo 9b. Jestliže v levém i pravém kontextu slovního tvaru s velkým počátečním písmenem zakončeného na *-slav*

[word="[:upper:].*slav"]

stojí tvar s počátečním velkým písmenem označovaný jako maskulinum životné oddělený od tvaru zakončeného na *-slav* čárkami

<-2,-2> **[word="[:upper:].*" & tag="..M.*"]**[word="\",",]

<2,2> **[word="\",","]**[word="[:upper:].*" & tag="..M.*"],

pak s velkou mírou pravděpodobnosti platí, že slovní tvar končící na *-slav* je maskulinum v nominativu singuláru, jeho lemma začíná velkým písmenem a končí na *-slav*

[word="[:upper:].*slav" & tag="NNMS1.*" & lemma="[:upper:].*slav"].

Stav desambiguace slovních tvarů s velkým počátečním písmenem, které končí na *-slav*, získaný dotazem

[word="\",","][word="[:upper:].*slav"]**[word="\",","]**

p-filtr <-1,-1> **[word="[:upper:].*" & tag="..M.*"]**

p-filtr <1,1> **[word="[:upper:].*" & tag="..M.*"]**

z korpusu Syn v8,⁵³ a výsledky ruční kontroly desambiguovaných výsledků i počty chyb opravených i generovaných navrženým pravidlem ukazuje následující tabulka 10b.

⁵⁰ Jak jsme uvedli výše, slovník *MorfFlex* (Hajič – Hlaváčová, 2016) zahrnuje informace o vlastních jménech. Tato rozlišení by bylo možné aplikovat v pravidlech pro desambiguaci antroponymu a oikonymu v koordinovaných řadách, a sice přidáním příslušné informace ve formě podmínky. Problém spočívá v tom, že výsledky užití pravidel, která využívají výsledky automatické morfologické analýzy, na ní jsou z principu závislé (např. v případě řady homonymních proprií jako *Vladislav*, *Vratislav*, *Soběslav* by uvažované pravidlo nefungovalo).

⁵¹ Např.: (...) *Luka nad Jihlavou*, *Předboř*, *Svatoslav*, *Příseka* či *Bitovčice* (...).

⁵² Např.: (...) *Lucie*, *Andrea*, *Jaroslav*, *Tereza* a *Marek* (...).

⁵³ (https://kontext.korpus.cz/freqs?ctxattrs=word%2Clemma%2Ctag&attr_vmode=mixed&pagewidth=40&refs=%3Ddoc.title&q=~7oCe8STFL2kS&viewmode=kwic&attrs=word%2Clemma%2Ctag&corpname=syn_v8&attr_allpos=all&fcrit=tag%2Fe%201%3C0&flimit=1&freq_sort=&fpage=1&ft_include_empty=).

Tabulka 10b

počet tvarů ([word="[:upper:]]. *slav"]) a jejich tagování	je správně tag	má být tag	tag by navržené pravidlo opravilo	chyby neopravené navrženým pravidlem	chyby vnesené navrženým pravidlem
440/NNMS1.*	439/NNMS1.*	1/NNFS1.*	-	-	1 ⁵⁴
X.*/20	0	19/NNMS1.* ⁵⁵ 1/NN[MI]S1.* ⁵⁶	20	0	0
NNFS1.*/3	1/NNFS1.*	2/NNMS1.*	3	0	1 ⁵⁷
Vl.*/1 ⁵⁸	0	1/NNMS1.*	1	0	0

Pravidla 9a, 9b samozřejmě nelze aplikovat stejně dobře jako pravidla 1 až 8, viz výše, protože jednak vnášejí do správně desambiguovaných výsledků chyby, jednak se opírají o výsledky automatické morfologické analýzy (značkování i desambiguace tvarů v kontextu desambiguovaných tvarů končících na *-slav*⁵⁹). Z těchto důvodů je nezařazujeme do závěrečné frekvenční analýzy.

Frekvenční analýza chyb v desambiguaci

V tabulce 11 uvádíme souhrnné počty získané testováním chyb, které by bylo možné odstranit navrženými pravidly.

Celkem jsme na základě výše uvedených rešerší odhalili 19 945 chyb v desambiguaci slovních tvarů končících na *-slav/-slava* (word="*(slav|slava)"), které lze navrženými pravidly uvést na pravou míru.

⁵⁴ Doklad z korpusu Syn v8: (...) *kolem lyžařských vleků přes Malý Stožek, Česlav/Česlav/NNMS1-----A-----, Velký (886 m) a Malý Sešov* (...).

⁵⁵ Ve třech případech jde o chyby v originálním textu (spojená slova jako např. *Timotej/Bohuslav, Šujík Miroslav*), které použité nástroje tokenizace (Hnátková a kol., 2014) analyzují jako samostatné tokeny.

⁵⁶ Jde o název hudební skupiny *Smrtislav*. Z dokladů na internetu je patrné, že užitě proprium kolísá v rodě.

⁵⁷ Doklad z korpusu Syn v8: (...) *František Nejedlo, Čáslav/Čáslav/NNFS1-----A-----, Vladimír Myslivec, Semteš 2* (...). Substantivum *Nejedlo* je označováno jako maskulinum životné (příjmení), lemma je s velkým písmenem. Substantivum *Vladimír* je maskulinum životné (rodné jméno).

⁵⁸ Doklad z korpusu Syn v8: *Ve starých kalendářích Achác, Děpold, Hostivit, Veleoslav/veleslavit/Vl-S---2--A----I, Vilibald Církevní kalendář Vilibald* (...).

⁵⁹ Například v kontextu substantiva *Vratislav* se vyskytnou substantiva *Vladislav* a *Soběslav*. Vyloučení jedné/druhé interpretace u substantiva *Vratislav* pak bude nutně závislé na správné desambiguaci substantiv *Vladislav* a *Soběslav*. Ta ovšem přesahuje možnosti automatické morfologické analýzy, neboť jde spíše o desambiguaci sémantické roviny jazyka.

Tabulka 11

typ chyb	počet chyb v desambiguaci v Syn v8
chyby v antroponymech známých osobností a v kombinaci rodné jméno a příjmení	15 314
chyby ve víceslovných oikonymech	46
chyby v koordinovaných skupinách	2 049
chyby v kontextu pojmenování míst	2 370
chyby v kontextu pojmenování funkcí	166
Celkem	19 945

Závěr

Propria zakončená na *-slav/-slava* jsou častými jmény osob a míst. Jako taková se hojně vyskytují v českých korpusech.

Analýza dat zahrnutých ve slovníku *MorfFlex* ukázala velmi dobré pokrytí uvedeného typu proprií. Přesto při porovnání interpretací obsažených ve slovníku *MorfFlex* s daty v korpusu Syn v8 se ukázaly některé dílčí nedostatky. K jejich odstranění navrhuje doplnění slovníku *MorfFlex* a) o 11 interpretací lemmat, jejichž tvary jsou chybně interpretovány automatickou morfologickou analýzou a b) o frekventovaná lemmata nerozpoznaná automatickou morfologickou analýzou, jejichž seznam uvádíme výše. Přimlouváme se také za odstranění antroponyma ženského rodu *Čáslava* a navrhuje ponechání pouze příjmení maskulinum *Čáslava*.

Homonymie proprií, jejichž tvary končí na *-slav/-slava*, představuje ovšem daleko vážnější problém pro automatickou morfologickou analýzu. Při podrobnějším zkoumání korpusových dat lze zjistit poměrně značné množství chyb v desambiguaci.

Navrhlí jsme proto pravidla, která by pomohla odstranit existující chyby v desambiguaci sledovaných tvarů proprií. Tato pravidla lze rozdělit do několika skupin. První skupinu tvoří přehled frekventovaných víceslovných pojmenování rodné jméno + příjmení a víceslovné názvy oikonym (*Mladá/Stará Boleslav*). Tyto jednotky navrhuje zařadit do databáze *LEMUR*. Další skupinou jsou kombinace tvarů **slava* **ová* (v doložených dokladech prezentuje konstrukce rodné jméno femininum následované příjmením femininem, viz výše), které lze desambigovat na základě jednoduchého kontextového pravidla. Následují pravidla popisující tvary zakončené na *-slav/-slava* v různě definovaných koordinacích a pravidla, která mají zabránit chybné interpretaci tvarů zakončených na *-slav* v kontextu lexémů, které vylučují jednu interpretaci a podporují druhou. Poslední skupinu tvoří pravidla, jimiž jsme se snažili ověřit možnost využití souvškyt tvarů – kandidátů na kongruentní skupiny. U většiny pravidel je spolehlivost stoprocentní, pouze u poslední skupiny nefungují pravidla stoprocentně.

Aplikace navržených pravidel by vylepšila stávající desambiguaci v případě 19 945 chybně desambiguovaných tvarů. Tento výsledek není nikterak statisticky významný (0,54 % všech tvarů zakončených na *-slav/-slava*), přesto svou cenu má.

Chyby v desambiguaci jsou patrným důsledkem použití statistických metod desambiguace. Domníváme se, že pravidlová desambiguace opřená o výše uvedené lingvistické intuice (koordinované skupiny v širším kontextu, lexikální obsazení kolokací) je schůdnou cestou k vylepšení morfologického značkování uvedeného typu proprií. UVÁŽÍME-LI, že korpusy rostou především velkým přílivem publicistických textů, v nichž se uvedená propria vyskytují s vysokou frekvencí, pak lze předpokládat, že investice do pravidel podporujících jejich správné interpretace nemusí být bez významu. Navíc by některá pravidla (konkrétně pravidla 1 až 6) bylo možné aplikovat i na dvojice podobných typů proprií (např.: **mil// *mila, *mír// *míra, Petr//Petra, Pavel//Pavla, ...*).

Text vznikl za podpory projektu MUNI/A/1181/2020 Gramatika a lexikon češtiny.

LITERATURA

- BENKO, V. (2015): *Araneum Bohemicum Maximum, verze 15.04*. Praha: Ústav Českého národního korpusu FF UK. [online] <<http://www.korpus.cz>> cit. 2020-03-04.
- BENKO, V. (2014): Aranea: Yet Another Family of (Comparable) Web Corpora. In: P. Sojka – K. Horák – I. Kopeček – K. Pala (eds.), *TSD 2014, LNAI 8655*. Berlin: Springer International Publishing, s. 257–264.
- DAVID, J. – ROUS, P. (2006): *Neviditelní svědkové minulosti: místní a pomístní jména na Vysočině*. Praha: Academia.
- DAVID, J. (2017): Vlastní jméno posesivní. In: P. Karlík – M. Nekula – J. Pleskalová (eds.): *CzechEncy – Nový encyklopedický slovník češtiny* [online] <https://www.czechency.org/slovník/VLASTNÍ_JMÉNO_POSESIVNÍ> cit. 2020-03-14.
- HAJIČ, J. – HLAVÁČOVÁ, J. (2016): *MorfFlex CZ 160310*. Praha: LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. [online] <<http://hdl.handle.net/11234/1-1673>> cit. 2020-03-06.
- HNÁTKOVÁ, M. a kol. (2014): The SYN-series corpora of written Czech. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavík: ELRA, s. 160–164. [online] <http://www.lrec-conf.org/proceedings/lrec2014/pdf/294_Paper.pdf> cit. 2020-04-14.
- JELÍNEK, T. a kol. (2018): Variabilita českých frazémů v úzu. *Časopis pro moderní filologii*, 100, 2, s. 151–175.
- Koplotovce* (2020): [online] <https://www.koplotovce.sk/download_file_f.php?id=621919> cit. 2020-04-14.

- KŘEN, M. a kol. (2019): Korpus SYN, verze 8 z 12. 12. 2019. Praha: Ústav Českého národního korpusu FF UK. [online] <<https://www.korpus.cz>> cit. 2020-04-16.
- LAMPRECHT, A. – ŠLOSAR, D. – BAUER, J. (1986): *Historická mluvnice češtiny*. Praha: Státní pedagogické nakladatelství.
- LUTTERER, I. – ŠRÁMEK, R. (2004): *Zeměpisná jména v Čechách, na Moravě a ve Slezsku: slovník vybraných zeměpisných jmen s výkladem jejich původu a historického vývoje*. Havlíčkův Brod: Tobiáš.
- OSOLSOBĚ, K. – MACHALOVÁ, J. (2013): Hypokoristika z rodných jmen v Korpusu soukromé korespondence. In: Z. Hladká a kol., *Soukromá korespondence jako lingvistický pramen*. Brno: Masarykova univerzita, s. 33–59.

Klára Osolsobě
osolsobe@phil.muni.cz
Ústav českého jazyka
Filozofická fakulta
Masarykova univerzita
Arna Nováka 1
602 00 Brno

Hana Žížková
zizkova@phil.muni.cz
Ústav českého jazyka
Filozofická fakulta
Masarykova univerzita
Arna Nováka 1
602 00 Brno