

**Univerzita Karlova  
Filozofická fakulta**

*Ústav českého jazyka a teorie komunikace  
Filologie – Český jazyk*

**Autoreferát (teze) disertační práce**

Pavλίna Synková

**Popis staročeské apelativní deklinace  
(se zřetelem k automatické morfologické analýze textů Staročeské textové banky)**

**Description of Old Czech Common Nouns Declension  
(with regard to Automatic Morphological Analysis of Texts in Old Czech Text Bank)**

vedoucí práce: Doc. RNDr. Karel Oliva, Dr.

2017

## 1. Obecné představení práce

Práce přináší explicitní popis staročeské apelativní deklinace, který může být využit jako základ pro automatické generování tvarů. Tyto tvary mohou být poté využity pro přiřazování morfologických kategorií (rodu, čísla a pádu) a lemmatu k tvarům ve staročeských textech. Práce tak vytváří podklady pro první krok k přeměně textových bank, které v současnosti pro staročeské období existují, v širě využitelný nástroj lingvistického výzkumu. Substantiva byla pro první fázi práce vybrána z toho důvodu, že v současné češtině pokrývají zhruba 30 % textu (Bartoň et al. 2009, 130), tedy nejvíc ze všech slovních druhů (a není důvod předpokládat jejich výrazně nižší zastoupení v textech staročeských).

V celé práci se zohledňují staročeské texty pouze v transkribované podobě, vychází se z podoby transkripce využívané v textech Staročeské textové banky (Černá – Lehečka 2015, Daňhelka 1957, 1963, 1985, Staročeský slovník 1968).

Ze dvou možných přístupů k morfologickému značkování historických textů si práce vybírá ten, ve kterém se nástroj buduje bez využití nástrojů pro moderní fáze jazyka (toto základní rozlišení používá ve své práci Piotrowski (2012)). Výhodou takového přístupu je vznik systematického popisu formální morfologie daného období a s tím související možnost využít v automatické morfologické analýze i detailní lingvistickou informaci o jazyce daného období (k jakému deklinačnímu typu dané apelativum patří, jaké hláskové změny se v daném tvaru vyskytují apod.). Nevyhnutelnou cenou za tento přístup je časová náročnost přípravy takových podkladů a přímá závislost popisu na zdrojích, pomocí kterých je budován. Předkládaný popis tedy nutně představuje pouze základ, který bude s rozvojem použitých zdrojů (především staročeských slovníků, viz níže) třeba aktualizovat a dotvářet.

Na obecnější rovině práce testuje zvolený přístup jako celek – pokud na základě práce vznikne úspěšný nástroj pro automatickou morfologickou analýzu staročeských apelativ, bude možné stejný/podobný postup použít i pro ostatní slovní druhy.

### 1.1 Základní pojmy

**Staročeskými apelativy** jsou míněna obecná jména vyskytující se v textech od počátku česky psaných souvislých textů zhruba do roku 1500. Apelativa byla dále v práci omezena na jména s domácí substantivní deklinací. Stranou byly ponechány všechny výrazy, které by byly v současnosti považovány za číslovky.

Tvar apelativa je pro účely práce vhodné dělit na **tvarotvorný základ** a **tvarotvorný formant**. Jako tvarotvorný základ byla definována ta část tvaru, kterou mají všechny tvary v paradigmatu společnou, a jako tvarotvorný formant potom část, kterou se tvary od sebe liší. Nejčastějším tvarotvorným formantem je **koncovka**, ale vzhledem k tomu, že ve staročeském období existují paradigmata, v nichž se jednotlivé tvary liší přítomností a nepřítomností kmenotvorné přípony (srov. *břiem-ě – břem-en-e*), byl do popisu zaveden i pojem **zakončení**. Zakončení je tvarotvorný formant tvořený kmenotvornou příponou a koncovkou (např. tedy *-ene*). Takto definovaný tvarotvorný formant umožňuje popsat tvary přehledněji, než by to umožňovalo zavedení koncovek a různých podob základu apelativa v různých tvarech.

Z hlediska lemmatizace korpusu je třeba pro dané období určit, která formální podoba bude považována za **lemma**, protože tvary prochází mnoha hláskovými změnami. Stejně jako ve staročeských slovnících (viz níže) práce vychází z formální podoby lemmatu k roku 1300. V další fázi lemmatizace se plánuje s touto podobou pracovat jako s hyperlemmatem a umožnit uživateli najít nejen všechny podoby sdružené pod hyperlemma (tedy např. *viera* i *víra* po zadání hyperlemmatu *viera*), ale hledat také jednotlivé hláskové podoby hyperlemmatu jako samostatná lemmata (např. tedy po zadání lemmatu *viera* najít pouze tvary jako *vieře*, *vierami* apod., ale ne již *víra*, *vírami* apod.). V předkládané práci se pracuje pouze s podobami k roku 1300, tedy s hyperlemmaty, pro jednoduchost však o nich mluvím jako o lemmatech.

Jako lemma je chápána jedinečná podoba NOM.SG (nebo NOM.PL u pluralií tantum) uvedená v některém ze staročeských slovníků. Dvojice výrazů jako *koš* a *koš* nebo *sěkerá* a *sěkyra* tedy představují celkem čtyři lemmata. Je-li potřeba mluvit o formálně velmi podobných lemmatech se stejným významem, používám pojem apelativum a jeho podoby (např. *koš/koš*, *sěkerá/sěkyra*).

Ve staročeském období je třeba rozlišit několik typů variací apelativních tvarů.

Prvním typem jsou varianty, které zachycuje už podoba heslových slov ve staročeských slovnících. Staročeský slovník (1968, 21) je nazývá **hláskoslovné varianty**. Jedná se o velmi nepravidelné variace vyskytující se u jednotlivých apelativ (např. *sěkerá/sěkyra*, *chřtán/chrtan/hrtan*, *anjel/anděl*, *povraz/provaz/provraz*).

Další variaci vnáší do staročeských tvarů změny, které jsou buď pozůstatkem historického vývoje před staročeským obdobím (např. střídání *k* a *c* na konci tvarotvorného základu jako ve tvarech *ruka* – *rucě*, nebo zánik hlásky *e* ve tvarech jako *zed* – *dí*), nebo jsou spojeny přímo se sledovaným obdobím (např. přehláska *u* v *i* ve tvarech jako *dušu* – *duši* nebo monoftongizace ve tvarech jako *viera* – *víra*). Aby bylo možné s touto heterogenní skupinou variací pracovat, rozdělila jsem je pro účely této práce na dvě velké skupiny podle toho, zda je proměna snadno odvoditelná z formální stavby daného apelativa (např. každé *ie* se mění na *í*, každé *u* se po měkké souhlásce mění na *i*), nebo zda závisí na konkrétním lemmatu (srov. např. variaci *pes* – *psa* v kontrastu s tvary bez variace *les* – *lesa* nebo alternaci *kráva* – *kravám* ve srovnání s tvary bez alternace *krása* – *krásám*). První skupinu variací je možné do popisu zavést pouze jako pravidlo přepisu sekvence určitých písmen v určitém kontextu na jinou sekvenci, pro druhou skupinu variací je třeba jiný přístup. Pro první skupinu variací, které jsou popsateľné obecným pravidlem, používám označení **hláskové změny**, i když se v některých případech o hláskové změny ve smyslu užívaném obvykle v historické mluvnici nejedná (např. případy jako *ruka* – *rucě*, proteze a elize hlásek apod.). Při pojednání o hláskových změnách rozlišuji tvary původní a tvary nástupnické (vzniklé hláskovou změnou) u těch hláskových změn, kde má takové rozlišení smysl (např. *viera* – *víra*, *kóň* – *kuoň*). O variacích, které jsou vázány na určitá lemmata, mluvím jako o **alternacích**. Do této skupiny variací jsou zařazeny i variace, které sice mohou být odvozeny na základě formální podoby tvaru, ale pravidla pro jejich uplatnění by musela být formulována velmi složitě (např. změna v GEN.PL závisí zřejmě na složení skupiny konsonantů na konci tvarotvorného základu: viz např. tvary *hvězda* – *hvězd*, *válka* – *válek*, *jitro* – *jiter* i *jitr*). Mezi alternacemi se vyskytují jednak variace kvantitativní (jako *sól* – *soli*, *kráva* – *krav*, *ryba* – *rýb*), jednak kvalitativní (jako *pes* – *psa*, *vajce* – *vajec*, *křest* – *křtu*).

## 1.2 Zdroje práce

Jako východisko pro popis apelativní deklinace práce využívá tři hlavní zdroje. Prvním z nich je **Gebauerova mluvnice** (díly Hláskosloví a Tvarosloví – skloňování, vydání 1963 a 1960). Tato mluvnice je považována za základní, protože jako jediná z českých historických mluvnic uvádí množství autentického transliterovaného materiálu s podrobnou interpretací. Ostatní historické mluvnice (Gebauer 2007, Trávníček 1935, Vážný 1965, Komárek 1969, Lamprecht – Šlosar – Bauer 1986, Komárek 2012) byly používány pro kontrolu a případnou korekci či doplnění Gebauerových (1960, 1963) náhledů.

Druhým zdrojem práce jsou **texty interní textové banky** (dále ITB), která představuje nejširší elektronicky zpřístupněnou a korpusově prohledávatelnou databázi textů ze staročeského období. Všechny texty jsou v ní uvedeny v transkribované podobě. Z hlediska periodizace staročeského období i z hlediska žánru se jedná o nevyvážený soubor textů. Tato textová banka vzniká v oddělení vývoje jazyka Ústavu pro jazyk český. Texty diachronní složky Českého národního korpusu jsou pro staročeské období podmnožinou textů obsažených v interní textové bance, proto jsem k nim nepřihlížela. Interní textová banka obsahuje všechny texty Staročeské a Středněčeské textové banky (tyto dvě banky jsou veřejně přístupné přes Vokabulář webový) a dále texty, které prozatím neprošly finální kontrolou editora. Jedná se o texty buď naskenované z edic, nebo transkribované z rukopisů. Tyto texty jsou označovány jako nespolehlivé. V práci byly využity jak spolehlivé, tak nespolehlivé texty ze staročeského období, tedy období zhruba od počátku souvislých českých písemných památek do roku 1500. Na základě jazykových charakteristik přiřazují pracovníci oddělení vývoje jazyka ke staročeským textům i některé texty ze začátku 16. století. I tyto texty byly považovány za zdroj materiálu pro popis, protože tyto texty jsou, nebo budou zařazeny do Staročeské textové banky a zde představované podklady chtějí pokrýt staročeské období v jeho celku.

ITB představuje proměnlivý zdroj informací – nepravidelně do ní přibývají nové texty, nebo se nespolehlivé texty po kontrole editora stávají spolehlivými. Představovaná práce vznikala mezi lety 2012 až 2016, a jednotlivé části tak měly k dispozici jinou verzi banky. Aby byl tento problém minimalizován, proběhla na závěr práce revize naprosté většiny témat na verzi 12. 02. 2015. U dvou témat nebyly kvůli velké časové náročnosti výsledky aktualizovány (analýza vznikové alternace proběhla na verzi 17. 05. 2013, analýza vzájemného přejímání koncovek ja-kmenové a i-kmenové deklinace proběhla na verzích 10. 12. 2013, 21. 02. 2014 a 02. 05. 2014). Verze 12. 02. 2015 obsahovala 7,6 milionu tokenů, z toho 4,4 milionu tokenů pocházelo ze spolehlivých textů, 3,2 milionu tokenů z textů nespolehlivých. Byl-li daný jev doložen v textech spolehlivých i nespolehlivých, byla vždy dána přednost dokladům z textů spolehlivých. Bylo-li třeba v práci citovat doklad z nespolehlivého textu, bylo dané místo vždy ověřeno ve fotokopiích rukopisů nebo v edicích v knihovně oddělení vývoje jazyka Ústavu pro jazyk český.

Třetím zdrojem práce jsou **novodobé slovníky pro staročeské období**, které zpřístupňuje Vokabulář webový. Tyto slovníky označuji souhrnně jako slovníky staročeské. V současnosti neexistuje staročeský slovník, který by slovní zásobu sledovaného období pokrýval celou. Gebauerův (a Smetánkův) Slovník staročeský (Gebauer, 1970) dospěl k písmeni *n* a jeho materiál je v mnoha ohledech nesystematicky vybrán (Homolková 2007). Nejkomplexněji metodologicky pojatý Staročeský slovník (Němec – Pečirková a kol., 1968–2008) pokrývá hesla s náslovím *na-* až *při-*. Malý staročeský slovník (Bělič – Kamiš – Kučera, 1978) obsahuje sice hesla z celé abecedy, ale je diferenční – jsou v něm uvedeny pouze lexikální jednotky obtížné pro uživatele současné češtiny. Konečně Elektronický slovník staré češtiny Vajdlová a kol., od

roku 2005 dosud) pokrývá prozatím všechna hesla v části abecedy od písmene *a* po *l* a v části *při* až *ž* doplňuje hesla, které neuvádí MSS. ESSČ plánuje postupné pokrytí celé abecedy. V práci vycházím z materiálu uvedeného ve všech těchto slovnících. I když jsou metodologicky nejednotné a pokrývají různé části abecedy, představují v současnosti neobsáhlejší zpracovaný přehled o staročeské slovní zásobě.

### 1.3 Použité nástroje

Pro práci s představenými zdroji bylo využito několika nástrojů. Pro prohledávání slovníků sloužilo uživatelské prostředí Vokabuláře webového, které umožňuje vyhledávání i pomocí regulárních výrazů. Pro práci s databází lemmat, která vznikla na základě staročeských slovníků, byl použit nástroj Open Refine, který umožňuje pracovat s komplikovanými daty (má pokročilé možnosti vyhledávání, filtrování i editace). Největším problémem bylo vyhledávání ve staročeských textech, protože hledání ve Staročeské textové bance i v ITB je v současnosti poměrně omezené. Zatímco Staročeská textová banka umožňuje vyhledávání omezit kontextem (umožňuje např. zadat, jak má končit výraz předcházející, a tak omezit homonymii tvarů), v textech ITB je možné vyhledávat pouze jeden slovní tvar. Dotazovací jazyk umožňuje zadání regulárních výrazů. Protože jsem potřebovala pracovat s ITB jako nejrozsáhlejší databází a postupné zadávání jednotlivých tvarů jednotlivých lemmat je velice zdlouhavé, a navíc nevyhnutelně poznamenané chybami a nedůslednostmi, vytvořil pro účely této práce Boris Lehečka nástroj Analýza tokenů v programu MS Excel, který umožňuje dva typy komplexnějšího vyhledávání. Umožňuje za prvé zadat všechny podoby tvarotvorných základů a všechny podoby koncovek, spojit tvarotvorné základy a koncovky a výsledné tvary vyhledat v ITB (která je přítomna jako databáze v samotném nástroji). Díky této funkci bylo možné systematicky vyhledávat naráz velké množství tvarů. Druhou funkcí nástroje Analýza tokenů je zobrazování seznamů typů tvarů vyfiltrovaných na základě koncové sekvence hlásek (typicky koncovek). Nalezené tvary vždy vyžadovaly ruční kontrolu, protože nástroj nepracuje s kontextem dokladů, zobrazuje pouze nalezený tvar.

### 1.4 Metodologické principy

Metodologicky práce bere za svoje východisko tvrzení v historických mluvnicích, která však ověřuje v textech ITB. Rozsah ověřování se lišil v závislosti na konkrétní problematice (více viz níže v úvodech ke shrnutí jednotlivých částí práce). Při ověřování konkrétního jevu (např. tvaru s určitou koncovkou) byly vždy ručně probrány všechny nalezené kontexty, pokud jich bylo méně než 500. Pokud bylo nalezených dokladů více než 500, bylo ručně probráno prvních 500, a pokud nebyl daný jev ověřen/nalezen, byl dále vyhledáván v textech Staročeské textové banky, která umožňuje omezit kontext dotazu. Pokud ani zde nebyl nalezen a neuvádí ho Gebauer (1960, 1963), je daný jev v práci označován jako nedoložený.

Obecným problémem morfologického popisu staročeského období je kontrast mezi systémovostí a doložením. Mnoho apelativ nemá doloženo celé paradigma, duál je obecně hůře doložen než singulár a plurál apod. Vzhledem k tomu, že cílem práce je popsat apelativní deklinaci tak, aby popis mohl sloužit automatické morfologické analýze, dávám vždy přednost systémovosti před doložením. Je totiž možné, že nově zpřístupněné texty budou obsahovat tvary, které současné verze použitých zdrojů neobsahují a systémové hledisko by je mělo pomoci pokrýt. Na druhou stranu je třeba mít na mysli, že se pracuje s jazykem bez rodilých mluvčích, s omezeným počtem dochovaných textů, a každý doklad tak má jinou

váhu než u jazyka současného. K doložení se proto důsledně přihlíží a popis není systematizován nebo zpravidelňován násilně nebo pro dané období nenáležitě.

Důraz na systémovost se projevuje zejména v tom, že je u všech apelativ předpokládáno plné paradigma, a to ve smyslu koncovek a zakončení (včetně málo doložených nebo vůbec nedoložených tvarů pro duál a vokativ plurálu) i ve smyslu podob tvarotvorného základu pro všechny tvary.

V celé práci se respektuje podoba lemmat uváděná ve staročeských slovnících. Nevýhodou tohoto přístupu je fakt, že tak byla do podkladů zahrnuta i apelativa, u kterých nejsou uvedeny doklady ze staročeského období (v GbSlov). Ověřovat, zda dané substantivum má některé tvary doložené ze staročeského období, je však mimo možnosti této práce. Nová lemmata jsou přidávána jen v odůvodněných případech a takové případy jsou vždy komentovány.

Slovníky jsou respektovány i v otázce řazení apelativ mezi pluralia nebo singularia tantum. U pluralií tantum je charakteristika evidentní z plurálové formy heslového slova, u singularií tantum existují mezi staročeskými slovníky takové rozdíly, že jsem se rozhodla jako singularia tantum pojímat pouze ta apelativa, která jako singularia tantum uvádí všechny staročeské slovníky, kde se apellativum vyskytuje.

Na staročeské slovníky spoléhám i v zachycení hláskoslovných variant, protože zkoumání nebo ověřování dalších hláskoslovných variant všech lemmat je mimo možnosti této práce.

Stejně tak je respektována podoba tvarů uváděných ve spolehlivých textech interní textové banky. Uvědomuji si, že se tak do podkladů dostává množství nejednotně řešených jevů, ale jejich sjednocení není v současnosti možné a popis tuto situaci do jisté míry odráží. Při nejistotě o správnosti nalezených tvarů byly dané kontexty konzultovány s editory textů. V případě, že se jednalo o chybu v transkripci, nejsou jevy v práci vůbec zmiňovány.

Z hlediska účelu předkládané práce je třeba zabývat se i otázkou propojování tvarů lemmat, tedy otázkou, jaké tvary má uživatel najít po zadání určitého lemmatu. Pro některé výzkumné otázky je vhodné najít širší spektrum tvarů, pro jiné zase spektrum užší. Protože propojení tvarů všech lemmat, kde by takový krok dával z jazykovědného hlediska smysl, je mimo možnosti této práce, bylo propojování lemmat zavedeno pouze v případě, kde jsou různá lemmata výsledkem morfologického vývoje malých deklinací (např. *bukva*, *bukev*, *bukvě*) nebo alternace působené jerovým *e* a jejího analogického napodobení (např. *bázen*, *bázn*) či výjimečné alternace (např. *dska*, *dcka*, *dcska* apod.). Propojení lemmat v těchto případech umožňuje neurčovat, k jakému lemmatu má být daný tvar řazen, a u malých deklinací zároveň umožňuje nepředpokládat plné paradigma podle nepůvodní deklinace.

Důležitou metodologickou připomínkou je i fakt, že celá práce využívá texty transkribované, protože právě pro ně jsou její výsledky určeny. Uvědomuji si, že každá transkripce představuje interpretaci a že se jednotliví editoři i při dodržování týchž zásad v pojímání mnoha jevů liší. Toto omezení je třeba mít na mysli u všech závěrů v práci konstatovaných. Pro hlubší poznání jednotlivých subtémat by bylo třeba zkoumat je přímo v rukopisech nebo alespoň v textech transliterovaných. Stejně tak je třeba mít na zřeteli i další omezení použitých zdrojů (staročeské slovníky se metodologicky odlišují a celou abecedu nepokrývá žádný, Staročeská textová banka je nevyvážený soubor textů a tato banka i ESSČ jsou zdroje

proměnlivé) i použitých metod (v popisu se nevychází přímo z textů, ale z historických mluvnic, jejichž tvrzení se ověřují v textech). I přes tato omezení však použité zdroje představují nejširší dostupnou základnu pro zvolený cíl a zvolená metodologie udržuje práci v realizovatelných mezích. Celý předkládaný popis vnímám jako základ určený k dalšímu rozvíjení.

## 1.5 Části práce

Popis apelativní deklinace, který má sloužit jako podklad pro tvoření tvarů využitelných následně pro automatickou morfologickou analýzu, musí zahrnovat seznam lemmat, ze kterého budou získány tvarotvorné základy, seznam koncovek a zakončení všech deklinačních typů, jimiž je možné od tvarotvorných základů získat tvary pro všechny pády a čísla, a přehled o proměnách tvarotvorných základů i koncovek, který zohledňuje jak pravidla kombinování tvarotvorných základů a koncovek, tak nástupnické podoby tvarů ve sledovaném období.

Proto má popis staročeské apelativní deklinace v této práci čtyři části. První část představuje popis koncovek a zakončení jednotlivých deklinačních typů. Koncovky a zakončení v ní jsou popsány jednak v textové formě, jednak ve formě tabulek, které koncovky a zakončení jednotlivých deklinačních typů shrnují. V druhé části jsou popsány všechny alternace, tedy změny tvarotvorného základu, které není možné nebo vhodné zavádět ve formě obecného pravidla, protože se nevyskytují u všech lemmat s danou formální stavbou (srov. *pes-∅ – ps-a*, ale *les-∅ – les-a*; *kráv-a – krav-ám*, ale *krás-a – krás-ám*), nebo by jejich zavádění formou pravidla bylo příliš složité (srov. *hvězd-a – hvězd-∅*, *otázk-a – otázek-∅*, *šacht-a – šacht-∅/šachet-∅*). Alternace jsou popsány v textové formě a pro jednotlivé typy alternací jsou zavedeny značky, které jsou použity ve čtvrté části práce – seznamu pro generování tvarů, jako signál, jaká alternace tvarotvorného základu se u daného apelativa objevuje. Ve třetí části jsou popsány hláskové změny, tedy takové formální proměny tvarů, které lze zavést pomocí pravidla. Hláskové změny jsou popsány ve formě textu a zároveň ve schematické formě jako pravidla, jaká písmena se mění na jaká (případně i v jakém kontextu). Čtvrtou částí podkladů je seznam apelativních lemmat, která jsou přiřazena ke vzoru a případně i k typu alternace, pokud se daného apelativa alternace týká. Práce obsahuje popis vzniku tohoto seznamu a představení údajů, které jsou v něm uvedeny. Seznam sám je kvůli svému rozsahu (obsahuje asi 29 000 apelativ) přiložen na CD. Na CD je přiložen také soubor se všemi tabulkami popisujícími vzory.

Kromě těchto částí práce obsahuje jako přílohu seznam výjimek, tedy seznam tvarů, ve kterých se vyskytuje nějaká výjimečná koncovka nebo výjimečná alternace či hlásková změna, jejíž systematické zavedení by podklady zbytečně zatěžovalo. V další příloze je uveden seznam všech textů ITB (ve verzi 12. 02. 2015) s uvedením, zda se jedná o text spolehlivý nebo nespolehlivý, a s počtem tokenů.

Následující části tezí přibližují základní metodologické postupy použité v jednotlivých částech práce a stručně opakují výsledky, ke kterým práce v jednotlivých částech dospěla.

## 2 Popis deklinačních typů

### 2.1 Principy popisu

Pro popis deklinačních typů je centrálním pojmem vzor. Vzor byl pro účely této práce definován jako jedinečný soubor koncovek/zakončení, jejichž prostřednictvím se tvoří tvary určité skupiny apelativ. Takto definovaný vzor je nejspecifičtější jednotkou popisu. Kvůli soudržnosti a přehlednosti výkladu je zavedena také úroveň obecnější – deklinační typ. Ten vychází z podoby kmene a je doplněn informací o rodu.

Deklinační typ představuje zobecnění. Ve většině případů sdružuje více vzorů. Protože v celém popisu usilují o vhodnou minimalizaci počtu vzorů, nebyly samostatné vzory zavedeny pro apelativa kolísající mezi dvěma vzory. Tato jména jsou řazena zároveň ke dvěma vzorům (např. apelativum *moč* je řazeno mezi ženské i-kmeny i mužské jo-kmeny). Samostatné vzory nejsou zaváděny ani pro pluralia a singularia tantum. Tato apelativa jsou zařazena ke vzorům, kam pravidelně patří (např. apelativum *húslí* k ženským i-kmenům). Na základě informace o omezení v čísle, která je zavedena do seznamu pro generování tvarů, budou od těchto apelativ tvořeny pouze tvary plurálu nebo singuláru podle příslušného vzoru.

Samostatné vzory jsou zavedeny pro apelativa nesklonná (pro každý rod jeden) a pro apelativa, u nichž se vyskytuje výjimečná kombinace podoby lemmatu a opěrného tvaru (o těchto vzorech se mluví jako výjimkách).

Vzor je popsán jednak v textu, kde je vysvětleno, proč je s určitými koncovkami/zakončeními ve vzoru třeba počítat, jednak ve formě tabulky. V tabulce se uvádí pád, číslo a rod, protože se u některých vzorů v paradigmatu vyskytuje změna rodu. Koncovky i zakončení se do vzoru zavádějí v podobě k roku 1300. U každé koncovky je v tabulce uveden její původ. Základem pro přiřazení původu je Gebauerova mluvnice (Gebauer 1960). Pro doplnění v místech, kde se o původu Gebauer nezmiňuje, jsou užívány i ostatní české historické mluvnice.

Vzorová slova jsou volena s ohledem na tradici, pokud nějaká existuje.

Pro pády a čísla se v celé práci užívají zkratky podle Leipzig Glossing Rules (2015).

Koncovky a zakončení jsou do vzorů zaváděny na základě popisu v historických mluvnících ověřeného v textech ITB. Systémové hledisko, které považují pro účely této práce za nutné, se projevuje v tom, že je pro všechna apelativa (kromě výjimek a singularií a pluralií tantum) předpokládáno celé paradigma (tvary pro všechny pády i čísla) a dále v tom, že jsou v popisu rozlišovány koncovky pravidelné a zvláštní. Koncovky pravidelné jsou zařazovány do vzoru, koncovky zvláštní pouze do seznamu výjimek. O zařazení koncovky mezi koncovky pravidelné rozhoduje její systémovost u dané deklinace a frekvence doložení. U zvláštních koncovek je nutné rozlišovat koncovky ojediněle doložené a pravděpodobně omyly (mezi nimi jsou nejčastější případy tzv. zvrtné analogie). Koncovky vyskytující se u daného tvaru pravděpodobně omylem nejsou řazeny ani do seznamu výjimek.

Stanovování toho, jaké koncovky patří do vzoru, se poněkud lišilo u malých a velkých deklinací.



Malé deklinace jsou deklinace s maximálně 80 zástupci a často představují zbytky deklinačních typů, které se ve staročeském období rozkládají, přejímají koncovky jiných deklinačních typů, a jejich paradigma proto obsahuje velké množství variace. Z těchto důvodů byly tvary zástupců těchto deklinačních typů zkoumány pomocí nástroje Analýza tokenů. Všechny možné (v mluvnicích předpokládané a jinak systémové) tvary všech zástupců daných deklinačních typů byly pomocí tohoto nástroje vyhledány v textech ITB. Popis vzorů pak vznikl na základě těchto tvarů.

Popis vzorů u velkých deklinací více spoléhá na Gebauerův (1960) popis. Jednotlivé koncovky byly ověřovány a zkoumány pouze v případě, že byly přejaty z jiných deklinací, byly málo doložené, nebo doložené u Gebauera (1960) až z pozdějších období vývoje češtiny. Zkoumání těchto koncovek a zakončení probíhalo pomocí seznamů tvarů vyfiltrovaných na základě koncovkových hlásek v nástroji Analýza tokenů.

Koncovky pro daný tvar náležité nebo jediné doložené byly do vzoru zařazeny vždy. U malých deklinací jsou do vzoru řazeny i doklady ojedinělých koncovek/zakončení tehdy, pokud se jedná o systémový vztah k jiné deklinaci. Výjimky (tedy koncovky deklinací, které se zkoumanou deklinací pravidelný analogický vztah nemají) jsou řazeny do seznamu výjimek. U velkých deklinací jsou do seznamu výjimek řazeny koncovky/zakončení s několika málo doklady (zhruba od jednoho do tří), vždy se ale zároveň přihlíží k celkovému počtu lemmat v deklinaci (tisíce versus nízké stovky lemmat) a systémovým vztahům k jiným deklinacím.

Do seznamu výjimek jsou tvary řazeny v hláskové podobě, která byla doložena v textech (nepřirazuje se k nim podoba základní k roku 1300).

## 2.2 Základní přehled vzorů

Základní přehled o vzorech, ke kterým práce dospěla, je uveden v Tabulce 1. U každého vzoru je uveden počet lemmat, která k němu byla přiřazena. Název vzoru se skládá ze zkratky *subst* pro označení substantivního vzoru, zkratky pro rod (u vzorů, v nichž se vyskytuje více rodů, je jako první vždy uveden rod mužský, po něm ženský a po něm střední, aby měly kombinace ustálené pořadí), dále je v názvu vzoru uvedeno označení kmene a zvolené vzorové slovo.

vzor	počet lemmat	vzor	počet lemmat
subst.n.ɔjo-kmen.znamenie	6520	subst.n.n-kmen.jmě	11
subst.f.a-kmen.žena	4295	subst.f.ɔja-kmen.zmie	9
subst.m.o-kmen.dub	2516	subst.f.ja-kmen.kršně	7
subst.m.o-kmen.zvuk	2147	subst.m.u-kmen.syn	7
subst.f.i-kmen.kost	2117	subst.f.ja-kmen.hospodyni	6
subst.m.o-kmen.hoch	1592	subst.f.výjimka.vieca	6
subst.f.ja-kmen.duše	1530	subst.m.o-kmen.bóh	6
subst.m.jo-kmen.muž	1085	subst.f.i-kmen.dřvi	5
subst.n.o-kmen.město	844	subst.f+n.nt+i-kmen.dietě	5
subst.m.jo-kmen.meč	739	subst.f+n.ɔjo-kmen.polúsestřie	5
subst.m.jo-kmen.otec	568	subst.m+n.ɔja-kmen.hrabie	5
subst.m.jo-kmen.učitel	531	subst.f.r-kmen.dci	4

subst.m.jo-kmen.hrniec	496	subst.m.ja+ɔjo-kmen.rukojmě	4
subst.m.o-kmen.had	353	subst.m.o+i-kmen.lud	4
subst.m.ja-kmen.panošě	349	subst.f.ɔv-kmen.tykev	3
subst.m.o-kmen.pták	345	subst.m.jo-kmen.přietel	3
subst.m.o-kmen.pán	339	subst.m.výjimka.duspivo	3
subst.m.o-kmen.chlap	301	subst.m+f.ɔja-kmen.bratřie	3
subst.n.jo-kmen.moře	297	subst.n.jo-kmen.poledne	3
subst.n.o-kmen.klubko	267	subst.n.o-kmen.slovo	3
subst.m+f.a-kmen.vladyka	153	subst.f.ja-kmen.rez	2
subst.n.nt-kmen.kuřě	152	subst.f.r-kmen.máti	2
subst.f.ja-kmen.dýně	144	subst.f+n.a+o-kmen.kamna	2
subst.f.ɔja+ja-kmen.historie	142	subst.f+n.o+i-kmen.oko	2
subst.f.ja-kmen.tvrz	106	subst.m.jo-kmen.turnaj	2
subst.f.ja-kmen.prorokyni	80	subst.m.n-kmen.den	2
subst.m.jo-kmen.peniez	69	subst.m.n-kmen.týden	2
subst.f.ɔv-kmen.svekrev	66	subst.m.ɔja-kmen.pradlé	2
subst.m.jo-kmen.vítěz	63	subst.m+n.nt-kmen.kniežě	2
subst.m.ɔja-kmen.sudí	52	subst.n.o-kmen.koleno	2
subst.m.o-kmen.zeměnin	43	subst.n.s+jo-kmen.nebe	2
subst.m+f.a-kmen.evangelista	37	subst.f.i-kmen.hosti	1
subst.f.ja-kmen.přě	30	subst.f.indecl.apokalypsis	1
subst.f.ja-kmen.obnož	29	subst.f.ja-kmen.vánocě	1
subst.m.indecl.alef	25	subst.f.ɔja-kmen.laní	1
subst.m.n-kmen.kámen	25	subst.f.ɔja-kmen.lodí	1
subst.m.t-kmen.loket	21	subst.f.ɔja-kmen.rolí	1
subst.f.ja-kmen.výheň	20	subst.f+n.a+o-kmen.játry	1
subst.m+n.ɔjo-kmen.podkonie	20	subst.m.i-kmen.zvěř	1
subst.f.ja-kmen.mzhyni	19	subst.m.n-kmen.jelen	1
subst.m+n.ɔja-kmen.markrabie	15	subst.m.u-kmen.vól	1
subst.n.indecl.tau	15	subst.m.u-kmen.vrch	1
subst.f.ɔja-kmen.paní	14	subst.m.výjimka.báťa	1
subst.n.n-kmen.rámě	14	subst.m.výjimka.popele	1
subst.m.ɔjo-kmen.řebří	13	subst.m.výjimka.řanca	1
subst.m.i-kmen.host	12	subst.m.výjimka.střiža	1
subst.m.i-kmen.náv	12	subst.m+f.o+i-kmen.prs	1
subst.m.u-kmen.čin	12	subst.n.o-kmen.dietko	1

Tabulka 1: Vzory staročeské apelativní deklinace a počet lemmat k nim přiřazených

Pokud jsou počty lemmat rozděleny do vzorů, ne primárně podle deklinačních typů, tvoří o-kmenová maskulina nejpočetnější vzor, protože se vlivem vývoje osobovosti a dalších faktorů rozpadají na více vzorů. Na prvních místech tabulky tak jsou střední ɔjo-kmeny a ženské a-kmeny, u kterých nebyly nalezeny žádné faktory rozrůzňující deklinaci do více vzorů.

Z tabulky je dále zřejmé, že téměř polovina vzorů má jen jedno až pět lemmat. Nejvíce vzorů s jedním nebo dvěma zástupci mají maskulina.

Pozorujeme-li vzory podle deklinačních typů, ke kterým patří, lze z tabulky vyčíst i rozrůzněnost deklinačních typů z hlediska vzorů. Více než 5 vzorů zahrnují tyto deklinační typy: nejvíce rozrůzněnou deklinací jsou ženské ja-kmeny s 12 vzory, poté následují mužské o-kmeny s 10 vzory, mužské jo-kmeny s 9 vzory, ženské bja-kmeny se 7 vzory a střední o-kmeny se 6 vzory. U každého z těchto typů se na počtu vzorů podílejí jiné faktory – u ženských ja-kmenů souvisí vzory s uchováním původních tvarů na jedné straně a s přejímáním novotvarů podle ženských i-kmenů na straně druhé. U deklinací mužských zase s vývojem osobovosti a se zakončením tvarotvorného základu. U ženských bja-kmenů souvisí rozrůznění s rozpadem deklinace, u středních o-kmenů se zachováním starých tvarů u některých zástupců a zakončením tvarotvorného základu.

### 3 Alternace

#### 3.1 Principy popisu

Druhou částí práce je popis alternací tvarotvorného základu. Pojmenování alternace bylo zvoleno pro takové změny v tvarotvorném základu, které není vhodné nebo možné zachycovat formou obecně platných pravidel jako hláskové změny, protože jsou buď vázány na konkrétní lemma (*pes – psa*, ale *les – lesa*, *kráva – krav*, ale *krása – krás*), nebo jsou podmíněny stavbou tvarotvorného základu (např. počtem slabik nebo typem souhlásek na jeho konci), a jejich zavedení formou pravidel by tak bylo velmi složité.

Byly popsány celkem tři obecné typy takových alternací: a) alternace v kvantitě, tedy prodloužení a krácení vokálů způsobující rozdíly v tvarotvorném základu jména např. ve tvarech *diera – děr*, *práh – prahu*, *sól – soli*, *břiemě – břěmeno*, *ryba – ryb*, b) alternace působené jerovým nebo vkladným *e*, tedy změny v tvarotvorném základu, v nichž v některých tvarech oproti tvarům jiným přibývá, přesunuje se nebo se ztrácí *e* – např. *ovcě – ovec*, *hra – her*, *orel – orla*, *rtek – retku*, c) alternace výjimečné, neboli takové změny tvarotvorného základu, které se vyskytují u jednotlivých lemmat a nelze je dobře zobecnit. Takové změny se vyskytují např. v tvarech jako *čest – ctí*, *dřevo – drva* apod.

Pro zkoumání alternací jako východisko sloužil zejména Gebauerův popis deklinace (Gebauer 1960), v některých případech i jeho hláskosloví (Gebauer 1963). Významným zdrojem informací byly staročeské slovníky. Pro lepší přehled o staročeské kvantitě jsem vycházela také z nepublikované práce Milady Nedvědové (rukopis).

Vzhledem k tomu, že tyto popisy jsou pro účely vytváření podkladů pro ge nerování tvarů všech apelativ sledovaného období kusé, byla k apelativům zmiňovaným ve všech zdrojích přidávána jména s podobnou stavbou nalezená ve staročeských slovnících, pokud bylo oprávněné se domnívat, že sledovaný jev má širší uplatnění (např. potenciálně se týká dvojslabičných a-kmenových apelativ s první slabikou dlouhou). Tvary těchto apelativ byly vyhledávány pomocí nástroje Analýza tokenů a nalezené doklady byly ověřovány v textech.

Pro alternaci v kvantitě byly takto vyhledány a ověřeny tvary asi 300 lemmat.

U alternace působené jerovým a vkladným *e* bylo třeba rozlišovat alternaci, při které dochází ve tvarech s koncovkou oproti bezkoncovkovému tvaru v NOM.SG k zániku *e* na určitém místě tvarotvorného

základu (tato alternace bývá nazývána (Mluvnice češtiny 2 1986, 257) zániková, viz např. *orel – orla*), a alternaci, kde se naopak oproti NOM.SG s koncovkou v bezkoncovkovém pádě (typicky GEN.PL) *e* objevuje (alternace bývá nazývána vzniková, viz např. *ovcě – ovec*). I když se jedná o tutéž situaci – na určitém místě tvarotvorného základu je nutné počítat s vložením, nebo zánikem *e* – byl ve výkladu rozdíl mezi vznikovou a zánikovou alternací využit, protože případy prvního typu bylo možné zkoumat jinak než případy typu druhého. Zatímco pro vznikovou alternaci představují spolehlivý zdroj informací staročeské slovníky, které uvádějí opěrný tvar včetně alternace (*orel, orla*), pro alternaci vznikovou žádný takový zdroj neexistuje.

Alternace zániková byla tedy zkoumána především na základě dat ze staročeských slovníků – ze všech apelativ, která mají lemma bez koncovky, byla na základě opěrného tvaru, který se neskládal pouze z koncovky, ale také z alternující části tvarotvorného základu, vybrána apelativa bez zřetelného sufixu. Na apelativa neodvozená zřetelným sufixem jsem se zaměřila proto, že právě u nich byly u Gebauera (1963) nalezeny některé nepravidelné tvary. Spolu s těmito jmény byla ze staročeských slovníků vybrána všechna apelativa, která nevykazovala pravidelnou alternaci typu bezkoncovkové tvary s *e*, tvary s koncovkou bez něho (schematicky KeK\_KK, např. *orel – orla*). U všech takto vybraných apelativ proběhlo zkoumání tvarů pomocí nástroje Analýza tokenů (jednalo se celkem asi o 600 lemmat), u deminutiv typu *domček* (jež pravidelně vykazují alternaci typu KeK\_eKK, tedy *domček – domečku*) bylo možné díky jejich specifické a formálně výrazné stavbě zkoumat tvary přímo v korpusovém manažeru ITB pomocí regulárních výrazů.

Pro alternaci vznikovou na rozdíl od alternace zánikové staročeské slovníky žádné jasné vodítko neposkytují a podoba tvaru u a-kmenů a o-kmenů není odvoditelná ani na základě Gebauerova (1960) popisu, který pro daný jev představuje nejobsáhlejší zdroj. Podoba tvarů tak musela být odvozována na základě analýzy dokladů z textů ITB. Ze staročeských slovníků byly za tímto účelem vybrány všechny kombinace posledních dvou souhlásek v tvarotvorném základu o-kmenů a a-kmenů (např. *zk, zd, kn, tr, rch*). V seznamu tokenů v nástroji Analýza tokenů potom bylo na základě tokenů zakončených např. *zk a zek, zd a zed, kn a ken, tr a ter, rch a rech* zjišťováno, zda se v textech ITB vyskytují tvary s alternací (*otázek, oken*), nebo bez ní (*hvězd, patriarch*), nebo obojí (*jitr, jiter*). Tvary nalezené v seznamech nástroje Analýza tokenů byly vždy ověřovány v textech ITB kvůli homonymiím, nespolehlivé doklady byly dohledávány v kopiích rukopisů nebo edicích. Tímto způsobem byly získány skupiny sekvencí souhlásek, které se vyskytují pouze bez alternace, pouze s alternací a s alternací i bez ní.

Apelativa, v jejichž tvarotvorném základu nastává výjimečná alternace, byla nalezena na základě Gebauerových mluvnic (1960, 1963) a staročeských slovníků, které nepravidelnosti ve skloňování signalizují v opěrném tvaru.

Zjištěné alternace tvarotvorného základu jsou u lemmat, v jejichž tvarech se vyskytují, označeny v seznamu pro generování tvarů. Pro označení typu alternace, tvaru, ve kterém k ní dochází, a její konkrétní podoby je užívána značka čitelná hlavně pro lidského vnímatele.

Lemmata s alternací vzniklou vývojem *jerů* nebo vkladným *e* mají zavedenu značku začínající E0. Podoba tvarotvorného základu v pádech s koncovkou a v pádech bez ní je ve značce popsána schematicky pomocí sekvence značky K pro konsonant a e pro hlásku *e*, kterou se signalizuje změna na konci

tvartvorného základu. Nejčastější situaci představuje střídání sekvence KeK v tvarech bez koncovky (např. sekvence *rel* nebo *zek* ve tvarech jako *orel*, *otázek*) a KK v tvarech s koncovkou (např. sekvenci *rl* nebo *zk* ve tvarech jako *orlovi*, *otázka*).

Lemmata s výjimečnou alternací mají zavedenu alternaci typu základ, v jejíž značce se vypisuje celá podoba tvarotvorného základu v určitých tvarech (např. pro lemma *čest* to jsou ve tvarech s koncovkou podoby *čest*, *čs*, *čc*, *ct*).

### 3.2 Základní přehled alternací

Tabulka 2 přináší přehled základních typů alternací a počtů lemmat k nim přiřazených.

typ alternace	počet lemmat
E0	6412
kvantita	500
základ	336
prehlaska	12
souhlasky	8

Tabulka 2: Základní typy alternací a počet lemmat k nim přiřazených

Z tabulky je zřejmé, že nejhojněji zastoupeným typem alternace ve sledovaném období (z hlediska počtu lemmat) je alternace vzniková a zániková (E0). Alternace není omezena pouze na deklinační typy s bezkoncovkovým NOM.SG, vyskytuje se i u deklinací, které mají NOM.SG s koncovkou, ale bezkoncovkově je GEN.PL (a-kmeny, ja-kmeny, střední jo-kmeny a o-kmeny). Druhou nejčastější alternací je alternace v kvantitě. Velkou část těchto lemmat tvoří dvojslabičná a-kmenová apelativa, u ostatních deklinací není alternace tak rozšířená. Alternace typu základ je velmi heterogenní, co se týče jednotlivých podtypů alternace, protože tato skupina zahrnuje všechny alternace, které nelze dobře zobecnit. Alternace typu přehlaska a souhlasky se ve sledovaném období týkají pouze jednotlivých apelativ.

Tabulka 3 uvádí přehled konkrétních značek alternací (tj. značek s uvedením, jakých tvarů se alternace týká a jak se v tvaru projevuje), které se vyskytují u více než 10 lemmat. Z celkových 118 značek alternací se jich u více než 10 lemmat vyskytuje pouze 15.

alternace	počet lemmat
E0-bezkoncovkove_tvary-KeK/tvary_s_koncovkou-KK	5026
E0-bezkoncovkove_tvary-KeK+KK/tvary_s_koncovkou-KK	1172
kvantita-GEN.PL-dlouha+kratka	344
kvantita-cele_paradigma-dlouha+kratka	137
E0-bezkoncovkove_tvary-KeK/tvary_s_koncovkou-eKK	127
zaklad-bezkoncovkove_tvary-řec+rec/tvary_s_koncovkou-rc+řc	63
zaklad-cele_paradigma-ěnin+an+ěn	46
E0-bezkoncovkove_tvary+koncovka_mi+ma-KeK/tvary_s_koncovkou-KK	27
E0-bezkoncovkove_tvary-KeK/tvary_s_koncovkou-KK+KeK	24
zaklad-bezkoncovkove_tvary-něK/tvary_s_koncovkou-ňK	23
zaklad-bezkoncovkove_tvary-něK+ňeK/tvary_s_koncovkou-ňK+nK	18
zaklad-bezkoncovkove_tvary-ščeK+štěK/tvary_s_koncovkou-ščK	16

kvantita-NOM+ACC+VOC.SG_pred_koncovkou_ě-dlouha/zbytek_paradigmatu-kratka	14
zaklad-bezkoncovkove_tvary-ščeK+štěK+šteK/tvary_s_koncovkou-ščK+šťK+štK	13
zaklad-pokud_koncovka_zacina_ě- dšč+dcšč+dčt+dct+dc+ct+st+dsc+dcc+dsc+dcsc+dščc+dšč+cc+desc+destc+dešč+deščc+deš tc/GEN.PL- desk+desek+deck+decsk+desck+deščk+dešk+cek+destek+dešek+dešček+deštek+deštěk/zb ytek_paradigmatu-dsk+dck+dsck+dcsk+dščk+dšk+ck+desk+de	11

Tabulka 3: Značky alternací vyskytující se u více než 10 lemmat

## 4 Hláskové změny

### 4.1 Principy popisu

Kapitola hláskové změny se zaměřovala na popis takových formálních změn tvarotvorných základů, koncovek i celých tvarů, které jsou pro staročeské období dobře popsateľné pravidlem (a není je tedy třeba primárně vztahovat pouze k určitým lemmatům). Jako hláskové změny byly v této kapitole pro jednoduchost označovány všechny změny hlásek a jejich skupin projevující se v písmu, i když některé z těchto jevů jako hláskové změny obvykle označovány nejsou (např. střídání velár a sykavek na konci tvarotvorných základů jako v příkladu *vlk – vlci*) a některé jsou pouze záležitostí pravopisného systému (např. značení sekvence hlásek *ň* a *e* ve tvarech jako *koněm*).

Uplatnění hláskových změn umožňuje z lemmat i koncovek, které jsou do podkladů zaneseny v podobě k roku 1300, získat všechny jejich podoby nástupnické (např. aplikací změn *ie > í* a *ě > e* vznikne od tvaru *vieř* podoba *víře*) a také získat náležité tvary při spojování tvarotvorných základů a koncovek (např. tvarotvorný základ *vlk* pro spojení s koncovkou *-i* potřebuje pravidlo *k+i=ci*, aby vznikl náležitý tvar *vlci*).

Při popisu hláskových změn sloužilo jako východisko Gebauerovo Hláskosloví (Gebauer 1963), které bylo soustavně konfrontováno se staročeskými slovníky a texty ITB. Hláskové změny byly popsány ve formě textu a formou pravidel.

### 4.2 Základní přehled hláskových změn

Následující přehled shrnuje změny, které pokládám za vhodné pro tvoření apelativních tvarů a jejich nástupnických podob zavést.

Změny jsou do tabulek rozděleny podle toho, zda a) je třeba je aplikovat na lemma před získáním tvarotvorného základu (změna *šč > šť* totiž značně rozmnoží tvarotvorné základy končící na *ť*, které představují pro automatické tvoření tvarů složitý jev) a je třeba s nimi počítat při získávání tvarotvorných základů u měkkých a smíšených deklinací, b) nastávají pouze v tvarotvorném základu, c) pouze v koncovce, d) pouze na konci tvarotvorného základu, e) v tvarotvorném základu i v koncovce a f) spolu s předchozími možnostmi také až ve výsledném tvaru (srov. např. sekvenci *ti* ve tvaru *milosti*, která podléhá změně na *ci*). To, zda je toto dělení změn pro generování tvarů vhodné, může ukázat až praktická aplikace. Stejně tak až praktické využití ukáže, v jakém pořadí je vhodné aplikovat jednotlivé hláskové změny.

změna	kontext	příklad
šč > šť	na konci slova	plášč, plášť
šč > št	před ě, ie, i, í	púščě, púště
šče > ště		hradišče, hradiště

Tabulka 4: Změny, které je třeba aplikovat na lemma před získáním tvarotvorného základu

změna na konci TZ (po odtržení koncovky)	výskyt	příklad TZ
b, f, l, m, p, s, v, z > b', f', l', m', p', s', v', z'	měkké deklinace, i- kmeny, ѡv-kmeny	zem, zem', král, král', otep, otep'
d, t, n = d', t', ň	měkké deklinace	hád' (hádě), bář (bářě), znameň (znamenie), sršeň (sršen)
d, t, n > d', t', ň d', t', ň > d, t, n	i-kmeny	dan, daň, kád, kád', kost, kost'

Tabulka 5: Změny při získávání tvarotvorného základu u měkkých a smíšených deklinací

změna	kontext	příklad
ú > ou	na začátku slova	ústa, ousta
o > ó > uo > ů u > ú > ou a > á e > é i > í y > ý	na začátku slova	ucho, úcho, oucho
é > ý	po h, ch, k, r, d, t, n	déka, dýka
ý > ej		pýcha, pejcha
o > vo	na začátku slova	ovcě, vovcě
ře, ře, ře > tře, dě, ně		hňetenie, hnětenie
rstv > řstv řstv > rstv tstv > ctv, tctv čstv > ctv, cstv žstv > stv šťstv > stv		kněžství, kněství
tedln > teln, tedn		smrtedlnost, smrtelnost, smrtednost
dln > dn, ln, nn, n		mučedlník, mučedník, mučelník, mučenník, mučeník
nn > n		panna, pana
kk > k		měkkost, měkost
vz > z	na začátku slova před K	vzdviž, zdviž

Tabulka 6: Hláskové změny, jež se vyskytují pouze v tvarotvorném základu

změna	kontext	příklad
o > ě ó > ie	po ž, š, č, ř, c, j, d', t', ň; po b, f, m, p, s, v, z na konci TZ u měkkých deklinací, po b', f', m', p', s', v', z' na konci TZ	mužovi, mužěvi

o > e ó > é	po l na konci TZ u měkkých deklinací, po l' na konci TZ	královi, králevi
ě = e ie = é	po l	tele, zelé
óv > ó		mužov, mužó

Tabulka 7: Hláskové změny, jež se vyskytují pouze v koncovkách

změna	kontext	příklad
u > i ú > í	po ž, š, č, ř, c, j, d', t', ň; po b, f, l, m, p, s, v, z na konci TZ u měkkých deklinací, po b', f', l', m', p', s', v', z' uvnitř TZ a na konci TZ	klúč, klíč
ě > e	po ž, š, č, ř, j, s, z, c	třeví, třeví
ie > é	po ž, š, č, ř, j, s, z, c	mužie, mužé
ú > ou	po h, ch, k, r, d, t, n, b, f, l, m, p, s, v, z, c	ženú, ženou
ó > uo > ú		bóh, buoh, bůh
o > uo	ne na začátku slova	právo, právuo
ie > í		viera, víra
é > í	po b, f, l, m, p, s, v, z, ž, š, č, ř, c, j	řemének, řemínek; jmelé, jmelí

Tabulka 8: Hláskové změny, jež se vyskytují v tvarotvorném základu i v koncovkách

změna	kontext	omezení výskytu	příklad
zh = žd	i, ě, ie		rózh: róždě
k = c g = z h = z ch = š r = ř	i, ě, ie		vlk: vlci
k = č g = ž h = ž ch = š r = ř	e	o-kmeny	člověk: člověče
c = č z = ž	e	jo-kmeny	otc: otče
d', t', ň > d, t, n d, t, n > d', t', ň	nulová koncovka	měkké deklinace a i-kmeny	oheň: oheň, ohen povodeň: povoden, povodeň jeskyň: jeskyn, jeskyň
d', t', ň = d, t, n	před i, í, ě, ie		koň: koně
t', d', ň + e = tě, dě, ně			koň: koněm, zeť: zetěm
d', t', ň + e = d'ě, t'ě, ně		jo-kmeny	koň: koňem
d', t', ň > d, t, n	před mi, ma		koň: koňmi, konmi
b', f', l', m', p', s', v', z' = b, f, l, m, p, s, v, z	ě, ie, i, í, Ø, K		otep' + i: otepi
šč > št'	nulová koncovka		púšč: púšč, púšť
Kr > Ker	nulová koncovka		bratr: bratr, brater, bratir



Kr > Kir Kl > Kil			
st > sc	nulová koncovka		zlost: zlost, zlosc
b > p d > t h > ch z > s ž > š	nulová koncovka		chléb: chléb, chlép

Tabulka 9: Hláskové změny, jež se vyskytují pouze na konci tvarotvorného základu

změna	kontext	příklad
šč > št	před ě, ie, i, í	púščě, púště
šče > ště		púšče, púště
šč > št, št	před konsonantem	kléščky, kléštky, kléštky
šč > št	před a, á, o, ó, é, u, ú, Ø	ščáva, štáva púšču, púšťu
r > er r > ir, ri l > il, li	mezi konsonanty	srdce, serdce, sirdce, sridce
t > c	před ě, ie, i, í	mostě, moscě
tc > c dc > c		rádcě, ráčě
tc > ts dc > ds		rádcě, rádsě
b > p d > t h > ch z > s ž > š	před p, t, ch, k, s, š	nehty, nechty
p > b t > d ch > h s > z š > ž	před b, d, h, z, ž	prosba, prozba
b', f', l', m', p', s', v', z' > b, f, l, m, p, s, v, z	e, é, a, á, u, ú, o, ó	omám' enec, omámenec, otep'ú, otepú

Tabulka 10: Hláskové změny, jež je třeba aplikovat na výsledný tvar

## 5 Seznam pro generování tvarů

### 5.1 Představení seznamu a principy jeho vytváření

Poslední částí předkládané práce je seznam pro generování tvarů. Tento seznam představuje syntézu poznatků získaných během analýzy tvarů staročeské apelativní deklinace a zároveň základ pro generování tvarů. Jedná se o seznam lemmat získaných ze staročeských slovníků, která jsou přiřazena ke vzorům a zároveň k alternacím tvarotvorného základu, pokud se u daného lemmatu nějaká alternace

vyskytuje. Kromě toho je v seznamu uvedeno i případné omezení v čísle, pokud je dané lemma plurale nebo singulare tantum (v takovém případě se od daného lemmatu budou podle přiřazeného vzoru tvořit pouze tvary plurálu, nebo singuláru). Posledním údajem je poznámka, ve které se uvádí, zda bylo lemma do seznamu doplněno (převážně se jednalo o důsledek limitů automatické extrakce) a zda v některém ze staročeských slovníků zavedeno je, nebo se jedná o lemma ve slovníku neuváděné. Na základě kombinace informací v tomto seznamu a informací v tabulkách koncovek a zakončení, které popisují vzory, lze získat všechny tvary staročeských apelativ v podobě k roku 1300. Aplikací pravidel popisujících hláskové změny se a) ve vzniklých tvarech uplatní fonotaktické a pravopisné zákonitosti, b) získají nástupnické podoby tvarů.

Seznam vznikl v několika krocích. Prvním z nich byla automatická extrakce seznamu apelativních lemmat ze staročeských slovníků s použitím rodu a malého písmena lemmatu jako indikátoru. Pro ESSČ byla použita verze z prosince 2014 a počítá se s aktualizací seznamu pro novější verze. Lemmata byla automaticky předběžně rozřazena ke vzorům podle koncovky v reprezentativním tvaru a koncovky tvaru opěrného. Automatickou extrakci provedl s použitím tabulky kombinací koncovek v NOM.SG a GEN.SG, kterou jsem sestavila, Boris Lehečka. Další práce se seznamem už byla pouze manuální a prováděla jsem ji sama.

Nejprve byla vyhledána všechna lemmata, která se do seznamu dostala na základě jmenného rodu, ale jedná se o číslovky a zájmena. Stejně tak byla vyhledána apelativa cizího původu skloňovaná podle cizích deklinací a apelativa skloňovaná podle deklinace adjektivní. Všechna tato lemmata byla ze seznamu vyřazena. Dále byla vyhledána lemmata, u nichž automatická procedura nechala vzor nerozhodnutý, nebo jim přiřadila více vzorů zároveň. Tato lemmata byla roztríděna a přiřazena ke vzorům (celkem se ve všech těchto krocích jednalo asi o 6 000 lemmat). Do seznamu byla doplněna lemmata, která mají ve slovnících vícenásobnou rodovou charakteristiku, ale jen jedno heslové slovo, a automatická procedura pro ně proto vytvořila pouze jeden záznam.

Dále byla ke vzoru ručně přiřazována lemmata v těch deklinačních typech, které obsahují více vzorů. Např. u mužských o-kmenů to znamenalo roztrídění jmen osob, živočichů a neživých entit, u ženských ja-kmenů hledání apelativ s tvarotvorným základem zakončeným na *n/ň* i apelativ jednoslabičných. Při rozdělování lemmat ke vzorům podle významu byl jako rozhodující brán význam uvedený jako první v daném slovníku.

V následujícím kroku byly kódovány alternace a označena pluralia a singularia tantum.

Seznam pro generování tvarů obsahuje 28 775 lemmat. Lemmata jsou přiřazená k 96 vzorům, různých alternací bylo přiřazeno 118. Singularií tantum bylo nalezeno 22, pluralií tantum 260. Doplněno bylo 252 lemmat, které staročeské slovníky obsahují, a 48 lemmat, která neobsahují.

## 5.2 Ukázka části seznamu

Ukázka části seznamu je uvedena v Tabulce 11, celý seznam je vzhledem ke svému rozsahu k práci přiložen na CD.

lemma	vzor	omezení	alternace	poznámka
šle	subst.f.ja-kmen.přě	plurale_tantum		
šlechetenstvie	subst.n.bjo-kmen.znamenie			
šlechetenstvo	subst.n.o-kmen.město		E0-bezkoncovkove_tvary-KeK+KK/ tvary_s_koncovkou-KK	
šlechetnička	subst.f.a-kmen.žena		E0-bezkoncovkove_tvary-KeK/ tvary_s_koncovkou-KK	
šlechovník	subst.m.o-kmen.hoch			
šlechťnosť	subst.f.i-kmen.kost			
šlechťnosť	subst.f.ja-kmen.duše	plurale_tantum		
šlechť	subst.f.a-kmen.žena		E0-bezkoncovkove_tvary-KeK+KK/ tvary_s_koncovkou-KK	
šlechť	subst.m+f.a-kmen.vladyka		E0-bezkoncovkove_tvary-KeK+KK/ tvary_s_koncovkou-KK	pridane_lemma _ve_sl_je
šlechťenie	subst.n.bjo-kmen.znamenie			
šlechťic	subst.m.jo-kmen.vítěz			

Tabulka 11: Ukázka části seznamu pro generování tvarů

## 6 Citovaná literatura

Bartoň, Tomáš – Cvrček, Václav – Čermák, František – Jelínek, Tomáš – Petkevič, Vladimír (2009): *Statistiky češtiny*. Praha: Nakladatelství Lidové noviny.

Bělič, Jaromír – Kamiš, Adolf – Kučera, Karel (1978): *Malý staročeský slovník*. Praha: Státní pedagogické nakladatelství.

Černá, Alena M. – Lehečka, Boris (2015): *Metodika přípravy a zpracování elektronických edic starších českých textů* [online]. [cit. 20. 08. 2016]. Praha: oddělení vývoje jazyka Ústavu pro jazyk český AV ČR, v. v. i. Dostupné z: [http://vokabular.ujc.cas.cz/soubory/nastroje/Methodika/Methodika\\_pripravy\\_a\\_zpracovani\\_elektronickyh\\_edic\\_DF12P010VV028.pdf](http://vokabular.ujc.cas.cz/soubory/nastroje/Methodika/Methodika_pripravy_a_zpracovani_elektronickyh_edic_DF12P010VV028.pdf).

Daňhelka, Jiří (1957): Obecné zásady ediční a poučení o starém jazyce českém. In: B. Havránek, J. Hrabák a kol. (eds.), *Výbor z české literatury od počátků po dobu Husovu*. Praha: Nakladatelství Československé akademie věd, s. 25–35. Dostupné z: <http://vokabular.ujc.cas.cz/moduly/edicnipoznamka.aspx?id=DanhelkaVybAkadI>.

Daňhelka, Jiří (1963): Obecné zásady ediční a poučení o češtině 15. století. In: B. Havránek, J. Hrabák, J. Daňhelka a kol. (eds.), *Výbor z české literatury doby husitské*. Svazek 1. Praha: Nakladatelství Československé akademie věd, s. 31–41. Dostupné z: <http://vokabular.ujc.cas.cz/moduly/edicnipoznamka.aspx?id=DanhelkaVybAkadII>.

Daňhelka, Jiří (1985): Směrnice pro vydávání starších českých textů. In: *Husitský Tábor: sborník Muzea husitského revolučního hnutí Tábor*, 8, s. 285–301. Dostupné z: <http://vokabular.ujc.cas.cz/moduly/edicnipoznamka.aspx?id=DanhelkaSmernice>.

- Gebauer, Jan (1960): *Historická mluvnice jazyka českého. Díl III. Tvarosloví. I. Skloňování*. Praha: Nakladatelství Československé akademie věd.
- Gebauer, Jan (1963): *Historická mluvnice jazyka českého. Díl I. Hláskosloví*. Praha: Nakladatelství Československé akademie věd.
- Gebauer, Jan (2007): *Historická mluvnice jazyka českého. Díl IV. Skladba*. Praha: Academia.
- Gebauer, Jan (1970): *Slovník staročeský. Díl I [A–J], Díl II [K–N]*. Praha: Academia.
- Homolková, Milada (2007): Historická lexikografie. In: J. Pleskalová a kol., *Kapitoly z dějin české jazykovědné bohemistiky*. Praha: Nakladatelství Academia, s. 199–242.
- Komárek, Miroslav (1969): *Historická mluvnice česká 1. Hláskosloví*. Praha: Státní pedagogické nakladatelství.
- Komárek, Miroslav (2012): *Dějiny českého jazyka*. Brno: Host.
- Lamprecht, Arnošt (1986): Hláskosloví. In: A. Lamprecht, D. Šlosar, J. Bauer, *Historická mluvnice češtiny*. Praha: Státní pedagogické nakladatelství, s. 25–128.
- The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses* [online]. [cit. 12. 05. 2016]. Dostupné z: <https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>.
- Nedvědová, Milada (rukopis): *Kvantita samohlásek ve starší češtině*. Rukopis uložen v oddělení vývoje jazyka Ústavu pro jazyk český AV ČR, v. v. i.
- Němec, Igor – Pečírková, Jaroslava a kol. (1968–2008): *Staročeský slovník*. Praha: Academia.
- Piotrowski, Michael (2012): *Natural Language Processing for Historical Texts*. Morgan & Claypool Publishers.
- Staročeský slovník. Úvodní stati, soupis pramenů a zkratek* (1968). Praha: Academia.
- Trávníček, František (1935): *Historická mluvnice československá*. Praha: Melantrich.
- Vajdllová, Miloslava a kol. (2005–): *Elektronický slovník staré češtiny* [online]. Oddělení vývoje jazyka Ústavu pro jazyk český AV ČR, v. v. i. Dostupné z: <http://vokabular.ujc.cas.cz/hledani.aspx>.
- Vážný, Václav (1964): *Historická mluvnice česká II. Tvarosloví. 1. část Skloňování*. Praha: Státní pedagogické nakladatelství.

## 7 Odborný životopis

### 7.1 Vzdělání

2003-2009: obory Český jazyk a literatura a Lingvistika – fonetika na FF UK, diplomová práce Morfologická a syntaktická charakteristika českých číslovek vyjadřujících počet entit, jejich souborů a druhů

### 7.2 Praxe od ukončení VŠ studia

od února 2016: pracovnice oddělení vývoje jazyka Ústavu pro jazyk český AV ČR

od října 2011: asistentka na ÚČJTK FF UK

od září 2009: asistentka na Ústavu formální a aplikované lingvistiky MFF UK

### 7.3 Publikační činnost

#### Monografie

*Discourse and Coherence. From the Sentence Structure to Relations in Text.* Praha: Univerzita Karlova v Praze, MFF, ÚFAL, 2015. 274 s. [s Š. Zikánovou, E. Hajičovou, B. Hladkou, J. Mírovským, A. Nedoluzhko, L. Polákovou, K. Rysovou, M. Rysovou a J. Václem]

#### Skripta a učebnice

*Čítanka 6. Učebnice pro základní školy a víceletá gymnázia.* Plzeň: Fraus, 2016. [s K. Šebestou, K. Šormovou, A. Králíkovou a A. Markem]

*Úvodní jazykový seminář: výklad a cvičení.* Praha: Karolinum, 2014. [elektronický zdroj; s R. Adamem, M. Benešem, I. Bozděchovou, F. Martínkem, H. Prokšovou, L. Saicovou Římalovou]

#### Studie a odborné články

Designing CzeDLex – A Lexicon of Czech Discourse Connectives. In: *Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation (PACLIC 2016)*. Seoul, Korea, s. 1–9 [s J. Mírovským, M. Rysovou a L. Polákovou]

Signals of Attribution in the Prague Dependency Treebank. In: *14th International Workshop on Treebanks and Linguistic Theories (TLT 2015)*. Warszawa, Poland, s. 292–299. [s L. Polákovou a J. Mírovským]

Usage of Electronic Sources of Historical Czech for Formal Description of Appellative-Noun Declension. In: *Sborník Semináře o digitálních zdrojích a službách ve společenských a humanitních vědách.* Univerzita Karlova v Praze, MFF, ÚFAL, Praha, 2015, s. 65–73. [s B. Lehečkou]

Zachycení výstavby textu v Pražském závislostním korpusu. *Slovo a slovesnost*, 76, 2015, s. 163–197. [s Š. Zikánovou, E. Hajičovou, J. Mírovským, A. Nedoluzhko, L. Polákovou a M. Rysovou]

Describing Old Czech Declension Patterns for Automatic Text Analysis. *Mundo Eslavo*, 13, 2014, s. 7–17. [s B. Lehečkou a K. Olivou]

Sentence Structure and Discourse Structure (Possible parallels). In: *Dependency Linguistics. Recent advances in linguistic theory using dependency structures*. John Benjamins Publishing Company, Amsterdam, The Netherlands, 2014, s. 53–74. [s L. Polákovou a J. Mírovským]

Discourse Relations in the Prague Dependency Treebank 3.0. In: *The 25th International Conference on Computational Linguistics (Coling 2014), Proceedings of the Conference System Demonstrations*. Dublin, Ireland, s. 34–38. [s L. Polákovou a J. Mírovským]

Genres in the Prague Discourse Treebank. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík, Iceland, s. 1320–1326. [s L. Polákovou a J. Mírovským]

Číslovky úhrnné, souborové a druhové. In: Petra Adámková et al., *Studie k moderní mluvnici češtiny 6. Úvahy o české morfologii*. Olomouc: Univerzita Palackého, 2013, s. 102–117.

Introducing the Prague Discourse Treebank 1.0. In: *Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013)*. Nagoya, Japan, s. 91–99. [s L. Polákovou, J. Mírovským, A. Nedoluzhko, Š. Zikánovou a E. Hajičovou]

Subordinators with Elaborative Meanings in Czech and English. In: *Proceedings of the Second International Conference on Dependency Linguistics (Depling 2013)*. Prague, Czech Republic, s. 128–136. [s L. Polákovou a J. Mírovským]

Nejčastější konektivní prostředky kauzálního vztahu v Pražském závislostním korpusu. *Studie z aplikované lingvistiky*, 3/2012, s. 35–52.

*Manual for Annotation of Discourse Relations in Prague Dependency Treebank*. Technical report no. 2012/47, Institute of Formal and Applied Linguistics, Charles University in Prague, Prague, 2012, 83 s. [s L. Polákovou, Š. Zikánovou, Z. Bedřichovou, J. Mírovským, M. Rysovou, J. Zdeňkovou, V. Pavlíkovou a E. Hajičovou]

Interplay of Coreference and Discourse Relations: Discourse Connectives with a Referential Component. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*. Istanbul, Turkey, s. 146–153. [s L. Polákovou a J. Mírovským]

Analyzing the Most Common Errors in the Discourse Annotation of the Prague Dependency Treebank. In: *Proceedings of the 11th International Workshop on Treebanks and Linguistic Theories (TLT 2012)*. Lisboa, Portugal, s. 127–132. [s L. Polákovou a J. Mírovským]

Does Tectogramatics Help the Annotation of Discourse? In: *Proceedings of the 24th International Conference on Computational Linguistics (Coling 2012)*. Mumbai, India, s. 853–862. [s L. Polákovou a J. Mírovským]

Semi-Automatic Annotation of Intra-Sentential Discourse Relations in PDT. In: *Proceedings of the Workshop on Advances in Discourse Analysis and its Computational Aspects (ADACA) at Coling 2012*. Mumbai, India, s. 43–58. [s L. Polákovou a J. Mírovským]

Vybrané problematické aspekty konektivních prostředků v rámci anotace mezivýpovědních významových vztahů v PDT. *Bohemica Olomucensia*, č. 2, 2011, s. 138–147.

Sentence Structure and Discourse Structure: Possible Parallels. In: *Proceedings of the International Conference on Dependency Linguistics (Depling 2011)*. Barcelona, Spain, s. 233–240. [s L. Mladovou a J. Mírovským]

Typical Cases of Annotators' Disagreement in Discourse Annotations in Prague Dependency Treebank. In: *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*. Valetta, Malta, 2010. s. 2002–2006. [s Š. Zikánovou, L. Mladovou a J. Mírovským]

## Data

*Prague Discourse Treebank 2.0*. Data/software, ÚFAL MFF UK, Prague, Czech Republic, <http://hdl.handle.net/11234/1-1905>, Dec 2016 [s M. Rysovou, J. Mírovským, E. Hajičovou, A. Nedoluzhko, R. Ocelákem, J. Perglerem, L. Polákovou, V. Pavlíkovou, J. Zdeňkovou a Š. Zikánovou]

*Prague Dependency Treebank 3.0*. ÚFAL MFF UK, Prague, Czech Republic, <http://ufal.mff.cuni.cz/pdt3.0/>, 2013. [s E. Bejčkem, E. Hajičovou, J. Hajičem, V. Kettnerovou, V. Kolářovou, M. Mikulovou, J. Mírovským, A. Nedoluzhko, J. Panevovou, L. Polákovou, M. Ševčíkovou, J. Štěpánkem a Š. Zikánovou]

*Prague Discourse Treebank 1.0*. ÚFAL MFF UK, Prague, Czech Republic, <http://ufal.mff.cuni.cz/discourse/>, 2012 [s L. Polákovou, Š. Zikánovou, E. Hajičovou, J. Mírovským, A. Nedoluzhko, M. Rysovou, J. Zdeňkovou, V. Pavlíkovou, J. Perglerem a R. Ocelákem]

## Recenze

Nové, korpusově založené mluvnice češtiny. *Český jazyk a literatura*, 66, 2015, s. 44-50, 96-99. [s R. Adamem, I. Bozděchovou a E. Lehečkovou]

Nová klasifikace souvětí v současné češtině. *Naše řeč*, 98, 2015, s. 147-151.

Ohlédnutí za prvními třemi ročníky Studií z aplikované lingvistiky (SALi). *Naše řeč*, 5, 2013, s. 265–269.

Zdenka Rusínová: Pokušení struktury. Brno: Host, 2011. 152 s. Studie osobností brněnské lingvistiky, 2. *Slovo a slovesnost*, 4, 2013, s. 298–302.

Novátorská studie o slovosledu ve starší češtině. *Naše řeč*, 2011, 94, s. 209–212. [Š. Zikánová: Postavení slovesného přísudku ve starší češtině (1500–1620), Karolinum, Praha 2009]

## 7.4 Zapojení v grantech

- od 2017 spoluřešitelka grantu Implicitní vztahy v textové koherenci (GAČR 17-03461S)
- od 2016 spolupráce na obohacování výzkumné infrastruktury RIDICS (Výzkumná infrastruktura pro diachronní bohemistiku, LM2015081)
- 2014–2016 spoluřešitelka grantu Vícejazyčná korpusová anotace jako podpora jazykových technologií (Kontakt LH14011)

- od 2015 spoluřešitelka grantu TextLink-cz (TextLink: Skladba diskurzu v evropských jazycích, LD15052)
- od 2014 spolupráce na grantu TextLink: Skladba diskurzu v evropských jazycích
- 2013–2014 spolupráce na grantu Diskurzní konektory v češtině (GA UK 36213)
- 2013 a 2014 řešitelka grantu Automatická morfologická analýza staročeských apelativ (VG2013)
- 2012–2015 spoluřešitelka grantu Koreference, diskurs a aktuální členění v kontrastivním pohledu (GAČR P406/12/0658)
- od 2012 zapojení v grantu LINDAT-Clarin: Institut pro analýzu, zpracování a distribuci lingvistických dat (LM2010013)
- 2010–2012 spoluřešitelka grantu Mezivětné vztahy (GAUK 4036/2009)
- 2010–2012 spoluřešitelka grantu Kontakt – K počítačové analýze struktury textu (ME 10018)
- 2010–2011 spoluřešitelka grantu Od struktury věty k textovým vztahům (GAČR GA405/09/0729)