Abstrakt

Práce přináší explicitní popis staročeské apelativní deklinace, který může sloužit jako základ pro automatické vygenerování tvarů spojených s morfologickými charakteristikami a lemmatem. Tyto tvary mohou být poté využity pro přiřazování morfologických kategorií (rodu, čísla a pádu) a lemmatu k tvarům vyskytujícím se v elektronizovaných staročeských textech. Práce tak vytváří podklady pro první krok k přeměně textových bank, které v současnosti pro staročeské období existují, v šíře využitelný nástroj lingvistického výzkumu. Staročeským obdobím se přitom ve shodě s obecně přijatou periodizací myslí období od vzniku souvislých českých textů zhruba do roku 1500. Substantiva byla vybrána proto, že v současné češtině pokrývají zhruba 30 % textu, tedy nejvíce ze všech slovních druhů. V celé práci se zohledňují staročeské texty pouze v transkripci užívané v textech Staročeské textové banky budované v Ústavu pro jazyk český AV ČR, v. v. i. Pro automatickou morfologickou analýzu představuje transkripce velké usnadnění, protože standardizuje písmo i pravopis, zároveň je však třeba mít na zřeteli, že každá transkripce je interpretací a je do jisté míry závislá na rozhodnutí editora textu.

V práci se pro popis staročeské apelativní deklinace využívají historické mluvnice, staročeské texty a slovníky staré češtiny. Historické mluvnice slouží jako východisko práce, jejich tvrzení byla systematicky ověřována a doplňována pomocí textů interní verze Staročeské textové banky. Verze použitá pro většinu témat obsahovala 7,6 milionu tokenů, k jejímu prohledávání byl využit nástroj *Analýza tokenů*, který umožňuje a) po zadání tvarotvorných základů (tj. části slova, kterou mají společnou všechny tvary paradigmatu) a koncovek generovat tvary a hromadně je hledat v textech, b) prohledávat tvary (ve smyslu typů) vyfiltrované na základě hlásek, jimiž tvar končí. Část použitých textů (3,2 mil. tokenů) prozatím neprošla finální redakční kontrolou. Pokud bylo třeba použít materiál v nich obsažený, byly doklady kontrolovány přímo v kopiích rukopisů nebo edic, ze kterých texty pocházejí. Slovníky pro starou češtinu zpřístupněné elektronicky ve *Vokabuláři webovém* sloužily jako základ pro přehled o slovní zásobě staročeského období. Žádný z nich však nepokrývá staročeské období celé a slovníky se metodologicky liší, proto je třeba do budoucna počítat s rozšiřováním a zpřesňováním údajů z nich získaných.

Popis staročeské apelativní deklinace se skládá ze čtyř základních částí.

První část představuje popis koncovek jednotlivých deklinačních typů (odpovídajících kmenům). Koncovky v ní jsou popsány jednak v textové formě, jednak ve formě tabulek. V tabulkách se zohledňuje i původ a doložení koncovek. V rámci jednotlivých deklinačních typů je popsán různý počet vzorů. Vzor byl definován jako jedinečný soubor koncovek, kterými se tvoří tvary určité skupiny slov (např. pojmenování pro osoby nebo apelativ s tvarotvorným základem zakončeným na veláru). Celkem bylo popsáno 96 vzorů ve 22 deklinačních typech (nejvíce zástupců mají mužské o-kmeny, střední ьjo-kmeny a ženské a-kmeny).

V druhé části jsou popsány alternace, tedy změny tvarotvorného základu, které není možné nebo výhodné zavádět ve formě obecného pravidla, protože se nevyskytují u všech lemmat s danou formální stavbou (srov. *pes-Ø – ps-a*, ale *les-Ø – les-a*; *kráv-a – krav-ám*, ale *krás-a – krás-ám*), nebo by jejich zavádění formou pravidla bylo příliš složité (srov. *hvězd-a – hvězd-Ø*, *otázk-a – otázek-Ø*, *šacht-a – šacht-Ø*/*šachet-Ø*). Alternace jsou popsány v textové formě a pro jednotlivé typy alternací jsou zavedeny

značky, jež jsou použity ve čtvrté části práce – seznamu pro generování tvarů – jako signál, jaká alternace tvarotvorného základu se u daného lemmatu objevuje. Celkem bylo nalezeno asi 120 typů alternací, nejvíce lemmat zasahují alternace působené jerovým nebo vkladným *e*.

Ve třetí části jsou popsány hláskové změny, jež jsou zde pojímány jako formální proměny psaných tvarů, které lze zavést pomocí pravidla. Jedná se jednak o hláskové změny spojené s vývojem tvarů v daném období (např. *viera* – *víra*, *bóh* – *buoh*), vychází se přitom z hláskové podoby předpokládané k roku 1300, jednak o změny vznikající při spojování tvarotvorných základů a koncovek, z nichž některé jsou jen otázkou ortografie (např. *vlk+i = vlci*, *líň+em = líněm*). Hláskové změny jsou popsány ve formě textu a zároveň ve schematické formě jako pravidla, jaká písmena ubývají, přibývají, či se mění na jaká (případně i v jakém kontextu). Celkem bylo popsáno asi 100 takových pravidel.

Čtvrtou částí podkladů je seznam apelativních lemmat, která jsou přiřazena ke vzoru a případně i k typu alternace, pokud se daného apelativa alternace týká. Základ seznamu vznikl automatickou extrakcí apelativních lemmat ze slovníků staré češtiny a byl rozsáhle manuálně tříděn a obohacován. Obsahuje asi 29 000 lemmat. Ve spojení s ostatními částmi bude seznam apelativních lemmat použit jako základ pro generování tvarů: ze seznamu lemmat budou získány tvarotvorné základy, které budou na základě informace o vzoru kombinovány s koncovkami (při zohlednění případných alternací). Pravidla pro hláskové změny zajistí formování tvarů podle fonotaktických a pravopisných pravidel i vytvoření všech pravidelných nástupnických podob.

Kromě těchto částí obsahuje práce seznam výjimečných tvarů, jejichž systematické zavedení by podklady zbytečně zatěžovalo.

Výhodou zvoleného postupu při budování nástroje pro značkování (tagování) a lemmatizaci je vznik systematického popisu formální morfologie daného období a s tím související možnost využít v automatické morfologické analýze i detailní lingvistickou informaci (deklinační typ, hláskové změny). Nezbytnou cenou za tento přístup je časová náročnost a přímá závislost popisu na zdrojích, s jejichž pomocí je budován. Předkládaný popis tedy nutně představuje pouze základ, který bude s rozvojem použitých zdrojů 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.