



Lingvistická analýza automatizovanej anotácie datívu adjektíva v heterogénnych menných skupinách v Slovenskom národnom korpuse

KATARÍNA GAJDOŠOVÁ

Cieľom nášho príspevku je analyzovať anotáciu bigramu adjektívum v datíve a substantívum v inom páde v prípadoch, keď stoja vedľa seba a nie sú v Slovenskom národnom korpuse (ďalej SNK) označované jednotne, t. j. ďalšiemu adjektívu alebo substantívu nie je programom pridelený rovnaký tag (morfológická značka) v kategórii rodu, čísla alebo pádu. Výskum sme realizovali na materiáli SNK, konkrétne v jeho hlavnom korpuse *prim-4.0-snk-all*. Uvedený korpus obsahuje všetky texty dostupné v databáze SNK. Na analýzu sme zvolili najrozsiahlejší korpus SNK preto, aby sme podľa možnosti zachytili všetky nesprávne anotácie a aby boli diagnostikované chyby perspektívne odstránené vo všetkých podkorpusoch, ktoré sa z tohto hlavného korpusu generujú. Tematickou inšpiráciou na opísanie tohto javu bola sčasti aj štúdia T. Jelínka (2006, s. 155 – 179), v ktorej porovnáva nezhodu adjektíva s nasledujúcim substantívom vo všetkých pádoch na materiáli Českého národného korpusu (ČNK).

Morfológická anotácia SNK sa realizuje dvoma spôsobmi. Časť textov SNK je ručne morfológicky anotovaná. Ide o súbor dát, ktorý je verejne dostupný vo verziách *r-mak-1.0* (322 600 tokenov), *r-mak-2.0* (511 534 tokenov) a *r-mak-3.0* (1 207 939 tokenov). Na textoch z ručne anotovaného korpusu sa natrénoval automatický nástroj, tager, ktorý potom automaticky priradil značky všetkým tokenom (textovým jednotkám) v databáze SNK. Každému tokenu sú najskôr priradené všetky možné značky podľa slovníka – morfológického analyzátora. Následne sa na základe štatistických pravidiel vyberie tá morfológická značka, ktorá z možných vyhovuje najlepšie aj vzhľadom na najbližšie okolie – susedné slovo v bigrame alebo dve susedné slová v trigrame.

Štruktúra morfológickej značky v SNK je kombinovaná (na rozdiel napr. od pozičnej v ČNK), teda pozície v značke sú síce pevné a každá kategória má svoje vlastné označenie, ale každému slovu sa priradujú iba jemu vlastné kategórie (počet znakov v tagu je v SNK od 1 do 7, v ČNK 15). Tak je v SNK substantívam priradený najskôr slovný druh (značka S), na druhom mieste



paradigma v rámci slovného druhu, v prípade substantív – substantívna (S), adjektívna (A), zmiešaná (F) a neúplná (U). Tretie miesto patrí kategórii rodu, kde sa v mužskom rode rozlišuje životnosť a neživotnosť – maskulínium životné (m), maskulínium neživotné (i), feminínium (f), neutrum (n). Na štvrtom mieste sa nachádza číslo – singulár (s) alebo plurál (p), a posledná pozícia v morfolologickej značke substantív je vyhradená pre pád – nominatív (1), genitív (2), datív (3), akuzatív (4), vokatív (5), lokál (6), inštrumentál (7).

Zloženie morfolologickej značky adjektív je podobné: po slovnom druhu adjektívum (A) nasleduje informácia, o akú paradigmu v rámci slovného druhu ide – adjektívna (A), zmiešaná (F), neúplná (U). V rode, čísle a páde majú kongruujúce adjektíva rovnaké hodnoty ako príslušné substantíva. Posledná, šiesta pozícia v morfolologickej značke adjektív patrí stupňu – pozitív (x), komparatív (y), superlatív (z). Podrobnejšie informácie o morfolologickej anotácii SNK a jednotlivých značkách sú dostupné na <http://korpus.sk/usage/morpho/>.

Materiál sme selektovali z korpusu *prim-4.0-snk-all* použitím príkazu [tag="A.*3.*"] [tag="S.*[124567].*"]. V korpuse sa nachádza 7 714 výskytov slov, resp. dvojíc slov, ktoré spĺňajú kritériá zadaného príkazu. Predpokladáme, že časť materiálu bude predstavovať množina technických chýb, ktoré sú dôsledkom chybného priradenia značiek analyzátorom. Chceme však poukázať na to, že nejde len o chyby v značkovani, ale aj o gramaticky a stylisticky správne postavenie dvoch slov, ktoré môžu mať odlišnú morfolologicкую charakteristiku.

Priradenie rozdielných značiek môžeme rozdeliť do niekoľkých skupín a v rámci nich do príslušných podskupín. Dokladový materiál uvádzame v takej podobe, ako sa nachádza v databáze SNK, pri hľadaných (kľúčových) slovách je zobrazená aj ich morfolologicкая značka. V zátvorke za každým dokladom je uvedená základná, skrátená informácia o konkrétnom zdroji (zoznam úplných bibliografií všetkých textov spracovaných v SNK je dostupný na <http://korpus.juls.savba.sk/stats/index.sk.html>).

ADJEKTÍVUM V DATÍVE A SUBSTANTÍVUM V AKOMKOLVEK PÁDE

Formálne chyby v texte

Nástroj, ktorý automaticky označuje tokeny príslušnou morfolologicкую značkou, počíta so správnym tvarom slova bez formálnych či gramatických chýb. Texty do korpusu však prichádzajú aj s chybami, ktoré sú často formálne – ide o nesprávne rozdelenie alebo odseknutie slova, či o absenciu koncovej interpunkcie. Nástroj s takýmto textovým materiálom nedokáže pracovať korektne a chyby sa odrazia práve v morfolologickej anotácii.

Do tejto skupiny radíme bigramy, ktorým k ich celistvosti buď niečo chýba, alebo majú niečo navyše.



1) Formálna absencia bodky a chybné spojené slová

Uplatňovanie spisovných tvarov oproti < nárečovým/nárečový/AAfp3x Používanie/používanie/SSns1> spisovných tvarov je nasledovné : v rádiu 31 % , sviečka 11 % , sviatok 10 % , staršiu ženu 18 % , nie je 9 % , niečo 12 % , piaty mesiac 18 % , prsteň 11 % . (Michal Jozefovič: Z nárečia Radošoviec a okolia) Liv Ullmannová (z autobiografie Premena) Hokejový turnaj vyvrcholil strhujúcim finálovým zápasom < cielavedomejšiemu/cielavedomý/AAns3y Švédsko/švédsko/SSns1> rozšírilo menoslov krajín , ktoré vyhrali olympijský hokejový turnaj , na päť (okrem neho Kanada , V . Británia , USA a ZSSR , resp . SNŠ) . (Pravda 1994-mar)

Prvý bigram nepredstavuje mennú skupinu. K formálnemu spojeniu prišlo preto, lebo veta je zrejme podnadpisom v texte a opodstatnene nemá na konci bodku. Vzhľadom na to, že texty v databáze SNK sú radené za sebou, stretli sa tu v jednom bigrame slová z rôznych kontextov, ktoré nemajú okrem náhodnej polohy nič spoločné. Automatická anotácia však prebehla v oboch zložkách bigramu správne.

Druhý príklad ukazuje častú chybu - výpadok bodky a slabiku navyše (správne *cielavedomejšie Švédsko*). Tak sa do formálneho bigramu dostávajú slová z dvoch odlišných syntagiem. Adjektívum je však v tomto prípade anotované správne, keďže v analyzátore sú len tri možnosti anotácie tvaru adjektíva *cielavedomejšiemu* (AAms3y, AAis3y, AAns3y) - všetky v datíve. Substantívum je tiež anotované správne.

2) Absencia slova v syntagme

1 . septembra bola škola priradená k < Ludovej/ludový/AAfs3x umenie/umenie/SSns4> (LŠU) v Humennom a stala sa jej pobočkou ... (Sninské noviny 2004/44)

V príklade absentuje podstatné meno *škola*, a tak sa vedľa seba dostali slová *k ludovej* a *umenie*, ktoré je navyše aj v nesprávnom tvare. Adjektívum je anotované správne, hoci vzniknutý bigram vykazuje nezhodu.

3) Rozdelené slovo v texte

Ide o textové chyby, ktoré vznikli pri zalamovaní textu, rozdeľovaní slov na konci riadku alebo boli pri rekonštrukcii textu v procese OCR nedôsledne opravené. V uvedených prípadoch ide najmä o takú zámenu, ktorá je analogická s jestvujúcim slovom. „Slovo“ vzniklo omylom odstránenými alebo pridanými hláskami a jeho výskyt môže bez vysvetlenia pôsobiť náhodne, nezmyselne.

... mali by sa sústreďiť na to , aby pomohli < rozhlasovým/rozhlasový/AAm-p3x pracovník/pracovník/SSms1> om ako profesionálnym používateľom jazyka dosiahnuť pohotovosť pri používaní hovoreného slova ... (Kultúra slova 1994/5)



S < **vojim/AFmp3x nepriateľom/SSms7** > a ich hriechu , či už sú zlí , blázniví alebo hlúpi musíme čeliť dôkladne . (Dan B. Allender, Tremper Longman - Pravá tvár lásky)

Keďže Božia moc je v protiklade k < **prirodzenej/AAfs3x sla/SSfs1** > - bosti stvorenia, keďže Jeho múdrosť je v úplnom proti - klade k najmenšej chybe v porozumení alebo hlúposti ... (Arthur W. Pink - Božia dokonalosť)

No vo vyspelých krajinách sa už trend jednoznačne presunul k < **plochým/AAp3x tele/SSns6** > vizerom s plazmovou či LCD obrazovkou . (Brejk 07/10)

V prvom prípade je adjektívum aj substantívum automaticky anotované správne. V druhom prípade pomohla analyzátoru pri anotácii substantíva predložka s (chybou v texte vznikol falošný predložkový pád), avšak určit druh paradigmy na druhom mieste v morfolologickej značke sa mu pri vzniknutom tvare *vojim* nepodarilo. Preto sa na druhom mieste v tagu pri tomto neznámom „slove“ nachádza značka *F* (zmiešaná paradigma – skloňujú sa podľa nej adjektíva patriace k vzoru *matkin, otcov*). V treťom a štvrtom prípade sa analyzátor riadil pri priradovaní morfolologickej značky adjektívu datívnou predložkou. Vzniknuté substantívum v štvrtom prípade má takisto správny tag: analyzátor vybral jedinú správnu možnosť z paradigmy lemy *telo* – lokál.

4) Chybné spojené slová

Spojené slová, zabudnutá slabika, nedokončené slovo, ktoré sa v texte spojilo s nasledujúcim slovom, znemožňujú pri automatizovanej anotácii priradenie korektnej morfolologickej značky anotovanej textovej jednotke, najmä však nasledujúcemu substantívu.

Niektoré z nich sa zdali byť radikálnejšie , ale ja si myslím , že my sme boli bližšie k skutočnej < **partizánskej/AAfs3x armyslím/armyslie/SSns7** > , že my sme boli bližšie k skutočnej partizánskej ar § máde ako ktorékoľvek z nich . (Andrew M. Greeley - Panna a mučeníčka)

Vďaka tejto < **milosrdnej/AAfs3x nevedomostiMaľko/SSms1** > kráča v ústrety nepríjemnosti ešte nepríjemnejšej . (Gustáv Murín - Ako sa máš)

Adjektívam je v oboch prípadoch priradená správna morfologická značka, navigátorom štatistickému nástroju sú prítomné datívne predložky. Chybné zapísané substantíva sa analyzátor snažil rozpoznať v prvom prípade cez inštrumentálovú koncovku *im* – skloňovanie podľa vzoru *vysvedčenie* (porov. priradenie lemy *armyslie*), v druhom prípade sa nástroj zamerl na časť nesprávne spojeného slova *Maľko*, ktoré vyhodnotil takisto správne.

5) Slovo v nesprávnom tvare

Tento druh nesprávnej anotácie je dôsledkom absencie hlásky alebo dĺžky v existujúcom slove. Keďže tak niekedy vznikajú formálne homonymá s iný-



mi slovami, použije anotačný nástroj pri priradovaní morfolologickej značky tú, ktorá zodpovedá vzniknutému známemu tvaru. Používateľ však z kontextu vie, že ide len o náhradné priradenie.

Alebo bol celý prípad jasný z písomnej agendy epidemiológa , alebo tu bol čísi úmysel , aby sa epidemiológ k < celej/AAfs3x kaze/SSis6 > nevyjadril . (Pravda 1994/01)

Dcéry sú v zhode so svojou mladistvosťou zobrazené vo svetlých šatách , ako to zodpovedá < dobovej/AAfs3x mode/SSis6 > . (Beatrix Bastlová - Na prvý pohľad?!)

Adjektívum aj novovzniknuté substantívum je v oboch prípadoch označované dobre, teda morfológická značka je tvaru priradená správne. Chybné je slovo v texte. Miesto feminína *kauza* sa v texte ocitlo maskulínium *kaz* v prvom prípade a miesto feminína *móda* zas maskulínium *mod* v druhom prípade.

Osobitnú podskupinu výskytov tu tvoria prípady, v ktorých je substantívum zo skúmaného bigramu v základnom tvare a chýba mu správna datívna koncovka. Slovu je automaticky priradená kategória nominatívu. Aj tu môžeme do istej miery hovoriť o homonymii v širšom význame – o homonymii tvarov.

V hre bol milión pre hokejový klub a 250 tisíc pre popradské basketbalistky . Viceprimátor predstúpil pred poslancov s návrhom , aby bola < hokejovému/hokejový/AAis3x klub/klub/SSis1 > ŠKP Poprad poskytnutá dotácia vo výške jeden milión korún . (Východoslovenské noviny - Korzár 2002/09)

Začali sme využívať priestory bývalého ČSZTV – budovu , ktorá v rámci reštitúcie pripadla , podobne ako i celý školský komplex budov , nášmu zriaďovateľovi < Biskupskému/AAis3x úrad/SSis1 > v Nitre . (10 rokov Gymnázia sv. Cyrila a Metoda v Nitre (1991-2001))

V oboch prípadoch je obom zložkám bigramu, adjektívu aj substantívu, priradená správna morfológická značka. Vzhľadom na kontext je však substantívum v texte v nesprávnom tvare, v zodpovedajúcom tvare by malo inú značku.

6) Nesprávna gramatická väzba

Chyby tohto druhu sa vyskytujú najmä v publicistických textoch, ide o defektnú syntaktickú konštrukciu – kontamináciu. Dve slová sú spojené do jednej syntagmy, v našom prípade bigramu, pričom každé patrí do inej významom alebo formou asociovanej syntagmy (porov. Encyklopédia jazykovedy, 1993, s. 246).

Nielen on , ale aj ďalší prednášatelia sa venovali najmä ekonomickým a < sociálnym/sociálny/AAmp3x dôsledkoch/dôsledok/SSip6 > nášho vstupu do EÚ a čerpaniu prostriedkov z predvstupových a štrukturálnych fondov . (Východoslovenské noviny - Korzár 2003/04)

Váš otec pravdepodobne nepatrí k vyznávačom < najmodernejším/AAip3z ná-



KATARÍNA GAJDOŠOVÁ

zorov/SSip2 > . (Barbara Bickmorová - Návrat)

Kompletné systémy však ponúka aj < koncovým/AAmp3x spotrebiteľov/SSmp4 > , samozrejme s kompletným poradenstvom . (ASB - Architektúra. Stavebníctvo. Bývanie 2004/05)

Adjektívam je v týchto prípadoch priradená správna morfológická značka, okrem kategórie životnosti v prvom príklade. V druhom a treťom prípade odvodil analyzátor kategóriu životnosť/neživotnosť zrejme od nasledujúceho substantíva. Značka pridelená substantívam je vo všetkých príkladoch správna vzhľadom na ich tvar a na možnosť výberu z morfológického analyzátoru.

Špecifiká automatizovanej anotácie

Nie každý bigram tvorí v tejto kategórii jednu syntagmu, niekedy patrí adjektívum do predchádzajúcej syntagmy, substantívum do nasledujúcej. Na základe pádov druhého člena bigramu – substantíva môžeme analyzované bigramy rozčleniť takto:

1) Nominatívne

a) adjektívum ako postponovaný atribút

Aj v Starej zmluve , kde bol ľudom < Božím/boží/AAmp3x Izrael/izrael/SSis1 > , platilo , že súd sa začína od domu Božieho (Ez . 9 , 6 a 1 . Pet . 4 , 17) . (www.referaty.sk)

Deti z domova i obyvateľov obce prišiel pozdraviť aj prvý podpredseda Matice < slovenskej/AAfs3x Igor/SSms1 > Kovačovič . (Práca 1998/12)

Programu odovzdania cien sa zúčastnili predseda NR SR Jozef Migaš , predseda Matice < slovenskej/AAfs3x Jozef/SSms1 > Markuš a ďalší hostia . (Pod Vihorlatom 1999/05)

Autor slávnej básne Smrť Jánošíkova urobil aj nesmierne užitočný opis písomnosti zo súdu z marca 1713 , ktorý neskôr odovzdal do zbierok Matice < slovenskej/AAfs3x Pavol/SSms1 > Dobšínský . (Týždeň-2007)

Väčšina výskytov postponovaného adjektíva v tejto pozícii patrí syntagme *Matica slovenská*. Vo všetkých prípadoch ide o nesprávne priradený pád v morfológickej značke adjektíva: pri tvare *slovenskej* to mal byť genitív. Morfológická anotácia substantív je správna.

b) vlastné meno – na mieste adjektíva je postponovaný atribút

Keď na druhý deň prišiel k slečne < Marplovej/AFfs3x inšpektor/SSms1 > Craddock , vyzeral unavený a sklúčený . (Agatha Christie - Puknuté zrkadlo a Mŕtva v knižnici)

V liste < Rímskym/AAmp3x Pavol/SSms1 > doslovne zatvára dvere myšlienke , že by človek bol naďalej schopný „ rozhodnúť “ sa v otázke duchovných vecí po páde Adama . (Joseph M. Bianchi - Boh sa rozhodol zachrániť)

„ *Nedosvedčili ste pred niekoľkými minútami , slečna Martínezová , že pán Sabich povedal pani < Polhemusovej/AFfs3x anjelik/SSms1 > ?* “ (Scott Turow - *Pre nedostatok dôkazov*)

Vo všetkých bigramoch je morfológická značka priradená správne, keďže ide o kombináciu postponovaného atribútu z predchádzajúcej syntagmy a nasledujúce substantívum patriace do ďalšej syntagmy. Značkovanie automatickým nástrojom prebehlo korektne, v texte sa pozične prekrývajú jednotky z dvoch rozdielnych syntagiem.

c) adjektívum s adjektívnou paradigmou a iné substantívum

c1) priezviská skloňované podľa vzoru *pekný* alebo *cudzí* a substantívum v nominatíve

Kamila preto náročky zhodila svoj kúsok na tanieri , aby mama nemusela žiarliť . Erikovi < Malému/malý/AAms3x torta/torta/SSfs1> spadla na zem . (Anders Jacobsson, Sören Olsson - *Bertove patálie*)

V 33 . min ušiel J . < Čiernemu/čierny/AAms3x puk/puk/SSis1> v sóle tesne pred zamýšľaným zakončením . (Denník Šport 07/02)

V príkladoch sú adjektívam aj substantívam pridelené správne morfológické značky.

c2) geografické názvy

Ani jeden geografický názov nepatrí do dvoch rôznych syntagiem: adjektívom do predchádzajúcej syntagmy a substantívom do nasledujúcej syntagmy. Hoci ide vo väčšine prípadov o dobre známe vlastné mená, tager nedokázal správne určiť spolupríslušnosť adjektív a substantív v týchto bigramoch. Tento problém sa perspektívne vyrieši zaradením geografických názvov aj s ich paradigmami do slovníka morfológického analyzátora, s ktorým tager pri automatickej anotácii spolupracuje.

<Veľkým/AAmp3x Lovciam/SSms1 > stačil na víťazstvo polčas , ale u susedov v Dolnom Ohaji zavládol smútok , lebo body putovali do Bešeňova . (Castrum Novum 1999/22)

Končiaci prezident Schuster k < Jaslovským/AAmp3x Bohuniciam/SSms1 > povedal , že Slovensko odstavi dva bloky tak , ako sa zaviazalo v prístupovej zmluve . (SME 2004/05)

Naopak , z víťazstva sa doreraz netešil len Trebišov , ktorý doma podľahol < Zlatým/AAmp3x Moravciam/SSms1 > , hoci do polčasu viedol . (SME 2004/08)

Keď po niekoľkých dňoch dosiahli hlavné mesto Soul , prišli < Južnej/AAfs3x Kórey/SSfp1 > na pomoc vojenskej oddiely USA a OSN . (<http://www.referaty.sk>)

Adjektívam bola morfológická značka pridelená správne, pri morfológickej značke substantíva už nástroj úspešný nebol, keďže v morfológickom analyzátore sa dosiaľ nenachádzajú tvary názvov geografických objektov, ako sú názvy miest a obcí. Problematickým pri priradení morfológickej značky je aj nesprávny pravopis (posledný príklad Kórey).



KATARÍNA GAJDOŠOVÁ

O atmosféru sa im postarali folklórne súbory z Hranovnice , Šuňavy , < Liptovskej/Aafs3x Tepličky/SSip1 > , Štrby , Spišskej Starej Vsi a zahraniční účastníci z poľských Gorlíc a českých Kunovic . (Východoslovenské noviny - Korzár 2001/07)

V skanzene vystúpili folklórne súbory z Dlhého nad Cirochou a < Kalnej/Aafs3x Roztoky/SSip1 > , ako aj folklórny súbor Kalina s čepčením nevesty . (Pod Vihorlatom 2002/20)

Ani v týchto príkladoch nie sú zemepisné názvy ani v jednom prípade značkové správne. Nástroj napr. v prvom a druhom prípade nedokázal spojiť predložku s niekoľkými nasledujúcimi adjektívami a substantívami a pridelil bigramu nesprávnu morfológickú značku.

d) nominatívne s elipsou

Elipsa substantíva je vždy medzi skúmaným adjektívom, s ktorým elidované substantívum patrí do jednej syntagmy, a substantívom, ktoré je súčasťou nasledujúcej syntagmy.

„ Dal mi ju ešte < šesťročnému/šesťročný/Aams3x pán/pán/SSms1 > Zíšek , Čáčovčan , ktorý ma zasvätil do koní... (Pravda 1994-apr)

Obdobne je fakultatívnu oproti predtým < obligatórnej/obligatórny/Aafs3x povinnosť/povinnosť/SSfs1 > prispieť do zdrojov fondu zo štátneho rozpočtu . (http://www.referaty.sk

Teraz máme v intendantnej službe svojho človeka , “ oznámil < ostatným/Aamp3x Richard/SSms1 > a začal rozdávať chlieb . (Colleen McCulloughová - Morganovo šťastie)

Všetky morfológické značky vo vyššie uvedených príkladoch sú priradené správne.

...s onou rekonštrukciou autentického príbehu detstva Georgea Pereca – príbehu dieťaťa židovsko - poľského pôvodu , ktorému ako štvorročnému zahynie otec vo vojne , ako < šesťročnému/šesťročný/Aais3x matka/matka/SSfs1 > v koncentračnom tábore . (Georges Perec - Čo je to tam na dvore za moped s chrómovaným kormidlom?)

Pánom s týmto menom prístane modrá . Ich rastlinným totemom je pŕhlava , < živočíchnym/živočíchny/Aamp3x kôň/kôň/SSms1 > . (Východoslovenské noviny - Korzár 2003/07)

V prvom príklade určil tager nesprávnu rodovú príslušnosť adjektíva (malo byť správne Aams3x), v druhom príklade je chyba nesprávnej rodovej príslušnosti rovnaká, avšak dá sa vysvetliť prihliadnutím nástroja na nasledujúce substantívum, ktoré sa v singulári označuje ako životné a v príklade má správnu anotáciu.

e) substantivizované apelatívne adjektívum a substantívum

K slovu sa dostali aj Vranovčania , keď Kozár zoči - voči Šantovi nepochodil . Aj keď < domácim/domáci/Aafp3x snaha/snaha/SSfs1 > nechýbala , proti



sformovanej defenzíve hostí sa nedokázali presadiť . (Východoslovenské noviny - Korzár 2002/04)

Keď zaznel zvon a ohlasoval , že sa poniesie < **chorému/chorý/AAns3x Viatikum/viatikum/SSns1** > , odobral sa do kostola , vzdialeného tri kilometre , obliekol si rochetku , vzal slnečník a sprevádzal Najsvätejšieho . (Eugenio Ceria - Don Bosco s Bohom)

Preto chcú Ďurka oženiť so Zuzkou Kamenských a vziať si ich za svojich . Zo začiatku Žofa a Lavko váhajú , no keď počujú , že < **mladým/mladý/AAm-p3x Palčík/palčík/SSms1** > prepíše svoj majetok , menia nôtu a Lavko hovorí (www.referaty.sk)

Vo všetkých prípadoch je na mieste adjektíva substantivizované adjektívum, ktoré morfológický analyzátor dosiaľ neobsahoval. Po ich doplnení do slovníka analyzátoru im bude perspektívne priradená kategória substantíva s adjektívnou paradigmou. V prvom a druhom prípade prihliadal tager pravdepodobne na nasledujúce substantívum a podľa neho určil aj príslušný rod. V treťom prípade sú všetky ostatné pozície v morfológickej značke prvého aj druhého člena bigramu správne.

2) Genitívne

V prípadoch bigramov adjektívum v datíve a substantívum v genitíve ide vždy o determinatívnu syntagmu, v ktorej je substantívum podradené nadradenému substantivizovanému adjektívu. V celom korpuse sa nachádzajú z tohto typu len nasledujúce tri výskyty:

... zasalutoval podľa predpisov službukonajúcemu pánu poručíkovi , pánu intendantnému dôstojníkovi , < **dozornému/AAms3x kuchyne/SSfs2** > , správcovi budov , práporčíkovi , slobodníkovi , desiatníkovi a strážnym pri vchode... (Georges Perec - Čo je to tam na dvore za moped s chrómovaným kormidlom?) Nedlho pred vystriedaním veliteľom roty – nadporučík Klepáč sa nezjavoval – urobil Jurkovič nevyhnutný poriadok na stole , poznačoval zmeny , desiatníkovi Čapounovi , novému < **dozornému/AAms3x roty/SSfs2** > , prikázal vyvetrať operačnú miestnosť od cigaretového dymu a sám sa pobral na svoju izbu . (Ladislav Ballek - Lesné divadlo)

Tager priradil obom zložkám bigramu správne značky. Keďže v Slovníku súčasného slovenského jazyka A - G (2006) je lexéma *dozorný* považovaná v týchto prípadoch za substantivizované adjektívum, bude potrebné doplniť do morfológického analyzátoru novú paradigmu pre túto lemu tak, aby analyzátor dokázal priradiť takýmto použitiam lemy *dozorný* morfológickú značku substantíva s adjektívnou paradigmou (SSms3).

Tento pápež mal odvahu hovoriť nepríjemnú pravdu aj < **mocným/AAip3x sveta/SSis2** > . (www.tkkbs.sk 05/04)



Podobným príkladom ako *dozorný* je aj prípad lemy *mocný*, avšak táto lexéma nie je v dostupných slovníkoch klasifikovaná ako substantivizované adjektívum. Aj vzhľadom na jej výskyt v textoch SNK však predpokladáme, že pri spracúvaní v novom výkladovom slovníku bude prehodnotená. Tager priradil leám správne morfológické značky.

3) Akuzatívne

Škála tohto druhu bigramu adjektíva v datíve a substantíva v inom páde ako datíve sa do istej miery prekrýva s vyššie uvedeným nominatívnym členením 1a) a 1d) (postponovaný atribút, adjektívne substantíva).

a) postponed atribút

Na kríži Otca prosil si za vrahov , čo ňa tupili ; a tvoje ústa lotrovi < kajúcemu/AAis3x raj/SSis4 > sľúbili . (Blumentál 2001/11)

Za týchto okolností Literárny odbor Spolku sv . Vojtecha pokladal za svoju povinnosť podať Matici < slovenskej/AAfs3x návrh/SSis4> na úpravu slovenského spisovného jazyka . (Kultúra slova 2000/2)

V prvom prípade určil automatický nástroj nesprávnu rodovú príslušnosť adjektíva, štatisticky zrejme vychádzal z bigramu s nasledujúcim substantívom *raj*, ktoré je neživotné. Morfológická značka substantíva je priradená správne. Druhý príklad má obe morfológické značky správne, nástroj bez pomoci predložiek správne dešifroval jednotlivé formy dvojitej väzby.

b) substantivizované adjektíva a substantívum

MIKULÁŠ : Koláčiky nemôžem od svojej osemnástky , keď som vtedy zožral skoro celú perníkovú chalúčku , dostal som k < sladkému/AAms3x odpor/SSis4> a ani zo zdravotných dôvodov nemôžem teraz sladké . (www.birdz.sk)

Hodí <chudobnému/AAis3x balíček/SSis4 > cigariet , ten si jednu zapáli a začne fajčiť . (Agota Kristof - Veľký zošit)

Príbuzní potom tancujú okolo ohňa , ukazujúc < mŕtvemu/AAis3x predmety/SSip4 > , ktoré mu patrili . (Andrej Pauliny - Pokračovatelia dedičstva)

Pred zápasom ste si trúfali sťažiť < belasým/AAfp3x cestu/SSfs4> za tromi bodmi , lenže už po polhodine hry bolo rozhodnuté o vašej prehre . (Východoslovenské noviny - Korzár 2001/04)

Ako tréner priviedol k trom olympijským < zlatým/AAmp3x Viktora/SSms4> Sanejeva (1968 , 1972 , 1976) a trénoval aj Vaľukevičovho otca Gennadija (zlatý na HME 1979) . (Denník Šport 07/01)

V dokladoch nedokázal analyzátor určiť správny rod v paradigme substantivizovaného adjektíva v prvom člene bigramu, pri určovaní morfológickej značky substantívneho člena bigramu bol stopercentne úspešný. Chyby v značke adjektivizovaných substantív sú spôsobené faktom, že morfológický analyzátor v čase automatizovanej anotácie nedisponoval potrebnými informáciami o týchto substantivizovaných adjektívach.

Na rozkaz božstva podrezal Madávi svojim < **najbližším/blízky/AAmp3z hrdlá/hrdlo/SSnp4** > a ich krv nechal vytečť na obrazy „ krvilačnej “ bohyně , rozložené po zemi . (Pravda 1994-jan)

... zosobnenie sviežosti , pôvabu a ľúbivosti mladosti Homér (gr . Homéros) – grécky epický básnik , ktorému pripisujú autorstvo dvoch najslávnejších antických eposov Ílias a Odyseia Solón (asi 640 – 558 pred n . l .) – aténsky zákonodarca , ktorý dal mestu liberálnejšiu ústavu , < **chudobným/AAmp3z úľavy/SSfp4** > a posilnil národného ducha Aténčanov... (Anatole France - Ostrov tučniakov)

Rovnaké prípady zloženia bigramu a postavenia jeho jednotlivých členov ako v predchádzajúcom príklade rozpoznal tager tentokrát bezchybne.

c) elipsa

Na rozdiel od podskupiny b) nasledujúce adjektíva nemajú tendenciu k substantivizácii (substantivizované adjektíva *mladý* , *starý* sa používajú v iných významoch) a pociťujeme pri nich absenciu nadradeného substantíva.

Ježiš sľubuje < **tichým/AAmp3z dedičstvo/SSns4** > zeme (porov . Mt 5 , 5) , zeme modelovanej silou kráľovstva : čiže bude im patriť nebeské kráľovstvo . (Angelo Vigano - Kto je Ježiš Kristus?)

„ Nazdávam sa , že to budú veľmi dobré hry , ktoré dajú < **ostatným/AAfp3x lekcii/SSfs4** > , ako ich usporadúvať na úrovni a neničiť prírodu . (Pravda 1994-feb)

Vyústilo to do takej škiepky , že mladšia vylepila < **staršej/AAfs3y zaucho/SSns4** > . (Anton Hlinka - Slovo do týždňa)

V prvých dvoch príkladoch je nesprávne určený rod adjektíva. Substantíva vo všetkých príkladoch a adjektívum v poslednom doklade sú označované bezchybne.

4) Vokatívne

Korpus *prim-4.0-snk-all* obsahuje len 2 výskyty dvojíc slov, ktoré sú automaticky označované ako spojenie adjektíva v datíve a nasledujúceho substantíva vo vokatíve. Pravidlá morfológickej anotácie musia byť formulované tak, aby sa v nich objavili všetky tvary, ktoré sa reálne vyskytujú alebo sa potenciálne môžu vyskytnúť v textoch. Paradigma substantív preto musí obsahovať aj vokatív, ktorého značku nástroj použije podľa štatistických pravidiel. Obe uvedené príklady sú však automaticky označované nesprávne, pretože ide o substantíva v datíve. Nesprávnu automatizovanú anotáciu v druhom prípade spôsobilo zle rozdelené slovo v pôvodnom texte: tvar *panej* je rozdelený na dve samostatné „slová“ *pane* a *j*.

Prišli sme až z Brazílie a toto ti nesieme “ - prihovárali sa Milostivému < **pražskému/AAms3x Jezuliatku/SSms5** > , pričom v rukách držali špeciálny prístroj . (Adsum 2005/03)



Bol tu aj úryvok stredovekej rytierskej piesne o láske Arna k Adelaide , < urodzenej/AAfs3x pane/SSms5 > j z Toulouse . (neznámy zdroj)

5) Lokálové

Spojenie adjektíva v datíve a substantíva v lokáli sa v zvolenej verzii korpusu nachádza 249-krát. Vo väčšine prípadov ide o spojenie slov v nesprávnej gramatickej väzbe – kontaminácii, ako sme o tom hovorili vyššie.

K významným < technologickým/AAfp3x novinkách/SSfp6 > Punta II bude patriť vznetrový motor 1 , 9 JTD (59 kW) a automatická prevodovka CVT s plynulou zmenou prevodového pomeru . (Profit 1999/09)

K dispozícii sú síce zasvätené doslovy ku < knižným/AAip3x vydaniach/SSnp6 > ich diel , niektoré dobové kritické ohlasy či príležitostné hodnotiace príspevky , ucelený teatrologický pohľad na ich tvorbu však chýba . (Slovo 2002-12)

Napriek < ťažkým/AAfp3x podmienkach/SSfp6 > pri ich „ výrobe “ v štúdiách . (Literárny týždenník 2000/37)

Všetkým bigramom pridelil nástroj správne morfológické značky.

6) Inštrumentálové

Ukrajinské vedenie vystupuje aj naďalej proti < medzinárodným/AAip3x spoločenstvom/SSns7 > požadovanému odstaveniu jadrovej elektrárne v Černobyle . (Pravda 1994-jul)

Ten je vďaka < automatickým/AAip3x elektromotormi/SSip7 > ovládaným spojкам v rozvodovej prevodovke a nábojoch kolies možné zapínať ovládačom na palubnej doske , a to do rýchlosti až 100 km / h . (Profit 1998/12)

Spojenia datívneho adjektíva a inštrumentálového substantíva v skúmaných bigramoch adjektíva v datíve a substantíva v inom páde ako datíve dokladujú príklady s datívnymi predložkami *proti*, *vďaka*, ktoré sa viažu až na substantíva *odstaveniu*, *spojkám* na štvrtom mieste od predložky. Na prvom (*medzinárodným*, *automatickým*) a druhom (*spoločenstvom*, *elektromotormi*) mieste od predložky stojí skúmaný bigram adjektívum a substantívum, pričom ten bigram je doplnením adjektíva (*požadovanému*, *ovládaným*), ktoré rozvíja na predložku sa viažuce, avšak pozíčne ďalej umiestnené substantívum (*odstaveniu*, *spojkám*). Pri takýchto komplikovaných vzťahoch vo vete je zrejme, že analyzátor nedokáže správne prideliť všetky morfológické značky. Napriek náročnosti jednotlivých syntagiem priradil tager substantívam v oboch bigramoch správnu morfológickú značku. Adjektíva analyzoval na základe datívnej predložky a priradil im v prvom prípade nesprávne kategóriu rodu, čísla a pádu a v druhom prípade kategóriu pádu.

Nádvorie bolo vydláždené tesanými mramorovými kockami a široké schodisko , vedúce k < dvojkrídlovej/AAfs3x železom/SSns7 > obitej bráne v úzadí , postavili z toho istého drahého materiálu . (Robert van Gulik - Vražda v Kantone)

Substantívum s predložkou *k bráne* bližšie určujú postupne rozvité adjektívne prívlastky *dvojkřídlovej* a *obitej*. Druhé adjektívum *obitej* je rozvité ešte príslovkovým určením vyjadreným substantívom *železom*. Adjektívum, ktoré je postponovaným prívlastkom predchádzajúceho predložkového datívu (súčasne tiež väzobným adjektívom), sa ocitlo v bigrame s nasledujúcim substantívom, ktoré je obligatónnym doplnením ďalšieho adjektíva. Na prvý pohľad skratová konštrukcia je v skutočnosti naozaj funkčná a používaná. Napriek syntaktickej náročnosti viacsovného spojenia priradil automatický nástroj morfológické značky jednotlivým členom bigramu správne.

K miske < plnej/AAfs3x reklamou/SSfs7 > ospevovaných výrobkov ho totiž doženie až skutočný , nefalšovaný hlad . (InZine 2001-jun)

Správnou automatizovanú morfológickú anotáciu skúmaných bigramov môžeme vidieť aj v prípade väzobného adjektíva *plnej*, ktoré sa spája s pozične vzdialeným substantívom *výrobkov*, pričom toto substantívum rozvíja zhodný prívlastok *ospevovaných*, ktorý bližšie určuje príslovkové určenie *reklamou*. Ide opäť o celú sériu navzájom sa prekrývajúcich vzťahov

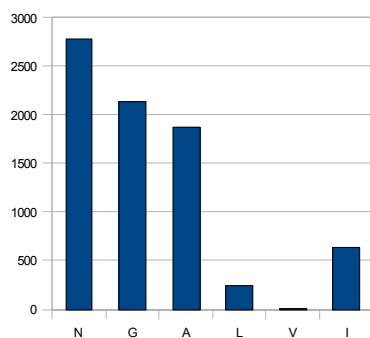
a) elipsa

Napriek mnohým odporcom má Mateus úspech podnes .K < ružovému/AAis3x rokmi/SSip7 > pribudol i biely , ponúkali sa dokonca i série tichého Matea . (Fedor Malík - Tisícročné víno)

Oveľa viac však tušil o spojení ľudského s detským , s detskou vôľou vzlietnuť k < božskému/AAms3x skratkou/SSfs7 > - ako bájnny vták . (Literárny týždenník 1998-dec)

V prvom prípade analyzátor určil morfológické značky správne, v druhom prípade nesprávne identifikoval rod adjektíva, hoci substantívum má prideľnú správnu morfológickú značku.

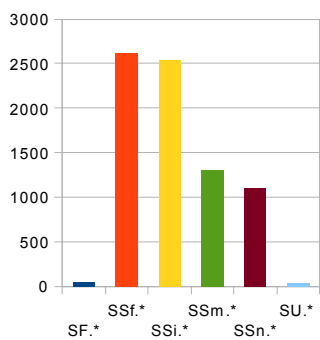
Grafické zobrazenie jednotlivých atribútov na mieste substantíva v bigrame s predchádzajúcim adjektívom v datíve



Graf znázorňuje zastúpenie automaticky označovaných pádov substantíva vo všetkých nájdených výskytoch bigramov adjektívum v datíve a substantíva v inom páde ako datíve. Najviac sa vyskytuje nominatív (2776), hneď za ním genitív (2130) a akuzatív (1870). Štvrtým v poradí je inštrumentál (632), posledným je lokál (240). Výskyt vokatívu (2) v bigrame sme vysvetlili v bode 6).



KATARÍNA GAJDOŠOVÁ



V grafe možno vidieť rozloženie jednotlivých substantívnych paradigiem (2. miesto v morfolologickej značke) skúmaných substantív v bigrame adjektívum v datíve a substantívum v inom páde ako v datíve tak, ako boli pridelené automaticky anotáčnym nástrojom. Najvyšší výskyt má substantívna paradigma v ženskom rode (SSf.* - 2618), po nej substantívna paradigma v mužskom rode - neživotné substantíva (SSI.* - 2549), nasleduje substantívna paradigma v mužskom rode - životné

substantíva (SSm.* - 1306), substantívna paradigma v neutre (SSn.* - 1103). Posledné miesta patria zmiešanej paradigme (SF.* - 41) a neúplnej paradigme (SU.* - 31)

Ďalšie skupiny skúmaných bigramov už nevyčleňujeme na základe pádu substantíva, ale na základe iných faktorov a súvislostí.

7) Cudzie slová končiace sa na -cia

Vyčlenenú skupinu tvoria v bigrame s adjektívom tie feminína, ktoré sa končia na sufix *-cia* a v textoch sa v prípade datívu singuláru vyskytujú s nesprávnym zakončením na dlhú samohlásku (*situácií*). Správne zakončenie tohto druhu feminín v datíve singuláru je krátke (*situácii*). Dlhú samohlásku majú tieto feminína v genitíve plurálu, preto dochádza pri morfolologickej anotácii k výberu morfolologickej značky s týmto pádom aj číslom. Nesprávny zápis feminín v texte teda spôsobuje inú anotáciu, správnu v rámci morfologických paradigiem, avšak nesprávnu v rámci kontextu. V konečnom dôsledku teda nejde o chybu automatického priradovania morfologických značiek, ale o chybu tvorcu textu tak, ako sme o tom hovorili už vyššie.

Gajdoš z *Humenného sa venoval* < **politickéj/politický/AAfs3x situácií/situácia/SSfp2** > v *HZDS ... (Východoslovenské noviny - Korzár 2003/03)*

Môžu sa tiež vymurovať z lícových tehál do kotevných profilov pripevnených na úrovni stropných konštrukcií alebo kotvených pomocou spôn priamo k < **nosnej/AAfs3x konštrukcií/SSfp2** > (obr . 3 . 235) . (Anton Puškár, Jozef Fučila, Ivan Řehák, Boris Vavrovič - *Obvodové plášte budov - fasády*)

Ešte po ceste na snem sa dohodli , že k < **správnej/správny/AAfs3x formulácií/formulácia/SSfp2** > *cirkevného učenia je potreba použiť filozofický výraz , , súpodstatný - gr . homousios “ , čo bolo na sneme aj urobené . (http://www.referaty.sk)*



Prednotifikačná fáza Ide o neformálnu fázu , ktorej cieľom je informovať Komisiu o pozadí novej resp . pravdepodobnej fúzie , prejednať načasovanie oznámenia a následného rozhodnutia , poskytovať účastníkom fúzie technickú i právnu podporu a rovnako spresniť druh a rozsah údajov , ktoré budú potrebné k < samotnej/AAfs3x notifikácii/SSfp2 > . (Beáta Marková - Hospodárska súťaž v SR po vstupe do EÚ)

Vo všetkých štyroch príkladoch morfológická značka automaticky pridelaná substantívam zodpovedá ich tvaru.

8) Viac adjektív za sebou

K automatickému určeniu datívneho adjektíva a substantíva v inom páde ako v datíve došlo aj v prípadoch, keď za sebou nasledujú dve adjektíva.

Vo vzťahu k našej téme , ale v istom zmysle k < celému/celý/AAis3x Kuzmányho/kuzmány/SUis2> životnému pôsobeniu sa javia ako rozhodujúce činitele podstatne ovplyvňujúce jeho osobnostný vývin a konanie v sledovaných oblastiach najmä dve skutočnosti . (Kultúra slova 1997/1)

„ Účinkuje “ Karel Kryl - to bol názov spomienkového podujatia , ktoré k < nedávnym/nedávny/AAip3x nedožitým/nedožitým/SSis1> 50 . narodeninám známeho a i na východe SR stále obľúbeného skladateľa , básnika a speváka Karla Kryla usporiadali v Múzeu moderného umenia rodiny Warholovcov v Medzilaborciach . (SME 1994/04)

Tager dokázal priradiť správnu morfológickú značku adjektívu, ktoré nasleduje po datívnej predložke *k*, ale nesprávne priradil značku nasledujúcemu adjektívu, ktoré označoval ako substantívum v prvom prípade s neúplnou paradigmou (*U*), v druhom prípade so substantívnou paradigmou.

a) vlastné mená v pozícii adjektív

Emil Žiga dal spätnú na strelu Jožkovi Kostelníkovi , ten strieľal krížnu strelu , brankár to nejako zle vyrazil a ja som už nemal problém s dopravením lopty do prázdnej brány , „ zhodnotil to , čo predchádzalo siedmemu < Puhákovmu/AFns3x zásahu/SSis2> v sezóne , samotný strelec . (Východoslovenské noviny - Korzár 2002/11)

Podľa neho bolo < Šindlerovej/AFfs3x vystúpenie/SSns1> s kapelou tak skvelé , že by mohlo ísť hneď na CD . (www.birdz.sk)

Aj keď som nejedol len suché kôrky ako < Čenkovej/AFfs3x deti/SSnp4> , lebo náš chlieb mal aj striedku , nielen kôrku , predsa som bol hladný ako ony , ako všetky deti u nás , čo obyvatelia úrodnejších krajov nevedia vždy pochopiť . (Ladislav Ťažký - Pred potopou)

V prvom príklade priradil nástroj temer všetky atribúty v tagu správne, chybná je len kategória rodu, ktorá sa má zhodovať s nasledujúcim substantívom. Tomu je zas nesprávne priradená kategória pádu. V druhom a treťom príklade je pri adjektíve nesprávne určený rod a pád. Substantívum v druhom



príklade je označované správne, v treťom príklade je chybný pád. Všetky chybné určené pády sú zrejme spôsobené pádovou homonymiou a nazdáva-me sa, že aj nevhodnými štatistickými postupmi tagera.

9) *Nezhoda vo vlastnom mene*

V prípade cudzích vlastných mien, ako sú názvy filmov alebo názvy spoločností, inštitúcií či vlastné mená osôb, nedokáže analyzátor správne určiť príslušnosť k jednotlivým paradigmám v rámci slovného druhu. Je to prirodzené, keďže cudzie vlastné mená nie sú súčasťou databázy morfológického analyzátor. Tager tak nemá z čoho vyberať príslušné morfológické značky a riadi sa len podľa zakončenia slova. Často si tak síce zvolí správnu paradigmu v rámci zakončenia, ale nesprávnu z hľadiska reálneho ohýbania (neraz cudzieho slova).

Po filmových drámach ako je snímok Pod povrchom , Nebezpečné známosti či < akčnej/akčný/Aafs3x Catwoman/catwoman/SSms1> z Batmanovho návratu sa Pfeifferová zjavne rozhodla vrátiť ku komédii . (Východoslovenské noviny - Korzár 2003/04)

V príklade určil tager správne slovný druh, paradigmu, rod aj číslo adjektíva, pád však priradil nesprávne. Pri substantíve bol úspešne automaticky určený len slovný druh, paradigma a číslo.

< Novému/nový/Aams3x Squadron/squadron/SSis1> Leaderovi sa podarilo získať si rešpekt medzi zvereným mužstvom a 242 . squadrona sa pod jeho velením vypracovala na slušnú úroveň . (http://www.referaty.sk)

Anotácia bigramu bola viac-menej úspešná, adjektívu bol nesprávne priradený rod životného maskulína miesto neživotného, substantívu bol nesprávne priradený pád.

Trebišovčania v nedeľu vycestujú do Lipian , kde proti < miestnej/Aafs3x Odeve/SSis6> odohrajú ďalší z kategórie nepríjemných duelov . (Východoslovenské noviny - Korzár 2002/03)

Problém pri automatizovanom anotovaní vlastných mien sa ukazuje niekedy aj pri domácich vlastných menách, ako to možno vidieť v uvedenom príklade. Adjektívum bolo označované správne, pri substantíve došlo k tvarovej homonymii s lokálom singuláru lemy *odev* (*o odeve*), pričom v príklade ide o datív singuláru lemy vlastného mena *Odeva*.

10) *Rekcia*

Prehľad väzobných adjektív a ich kolokátov, ktoré obsahoval materiál skúmaných bigramov adjektíva v datíve a substantíva v inom páde uvádzame v tabuľke. Vo všetkých prípadoch (okrem *blízky Kremľu*) ide o genitívnu väzbu.

plný	násilia	6	blízky	Kremľu	1
	lásky	5	neschopný	súcit	1
	hviezdy	3		vďačnosti	1
	fantázie	3		práce	1
	sily	3	nedôstojný	Rimana	1
	nádeje	3		človeka	1
hodný	štúdia	1	nehodný	lásky	1
	odmeny	1		úcty	1
	povolania	1			
	života	1			
	odhalenia	1			
	úcty	1			
	zaznamenania	1			
	ocenenia	1			

Vracajú sa z parketu , zadychčaní , ovievajú sa . Vracajú sa k stolu < **plnému/Aais3x zvyškov/SSip2** > . (Adalet Ağaoğluová - Svadobná hostina)

Ak nechabne vo vytrvalosti , postupne sa zaradí do zástupu závislých osôb a dostane súhlas , aby mohol horlivo a hlúpo lichotiť pyšnému patrónovi , < **neschopnému/Aams3x vďačnosti/SSfs2** > či priateľstva , a takmer lahostajnému k prítomnosti , odchodu a návratu cudzinca . (Edward Gibbon - Úpadok a zánik Rímskej ríše)

Hlásiť sa k pokornému a tichému Srdcu < **plnému/Aans3x lásky/SSfs2** > a pravdy , znamená nechať sa ňou pretvoriť , a nechať sa až zbožštiť pravdou . (Blumentál 2000/6)

zdalo sa jej , že počuje tie triumfálne tóny , ktoré ju sprevádzali ; boli slávnostné , ale plné horkého , nesmierného žiaľu , akoby takto oblečená a zavesená do svojho muža nekráčala v ústrety radostnej budúcnosti , ale naopak , v ústrety ťažkej obeti , životu < **plnému/Aais3x strachu/SSis2** > a neprekonateľných ťažkostí . . . (Alberto Moravia - Lahostajní)

V uvedenom príklade bral tager pravdepodobne do úvahy substantívum, ktoré predchádza skúmanému adjektívu, a nie nasledujúce substantívum, ako je to bežnejšie. Automatické určenie morfolologickej značky prebehlo vo všetkých prípadoch adjektíva aj substantíva bezchybne, hoci pre nástroj ide naozaj o zložitý vzťahy vo vete.

V divadle poézia nie je , alebo by aspoň nemala byť príčinou , ale následkom , a rozprávačská stavba a charakterová kresba urobila túto vrcholne intímnu hru prístupnou každému človeku < **schopnému/Aais3x citov/SSip2** > . (Arthur Miller - Zákruty času)



Adjektívu bol nesprávne priradený rod neživotného maskulína miesto správneho životného maskulína, keďže rod je závislý od predchádzajúceho životného maskulína *človek* a nie nasledujúceho správne označovaného neživotného maskulína *cit*. Nástroj priradil kategóriu rodu podľa bigramu s le-mou *cit*.

11) Nesklonné substantíva

Do pozičného spojenia v skúmanom bigrame vstupujú aj nesklonné substantíva a adjektíva. Ak nie sú súčasťou slovníka morfológického analyzátor, pri ich správnom určení vznikajú problémy.

Okolo piatej hodiny nasadáme do džípov spolu s vodičom a rangerom , ktorí nás budú sprevádzať < ranným/AAnp3x safari/SSmp1 > . (Cestovateľ 2002/5)

Bigram má nesprávne priradené morfológické značky. Substantívum *safari* malo byť v strednom rode, miesto datívu v prípade adjektíva a nominatívu v prípade substantíva mala byť obom členom bigramu priradená inštrumentálna kategória pádu.

Pikantériou je , že po remíze 1 : 1 na MS 1998 , ktoré usporiadali práve vo Švajčiarsku , naši vo zvyšných troch dueloch vyhrali vždy < rovnakým/AAnp3x skóre/SUns4 > 3 : 1 - vo štvrtfinále MS 2003 (góly SR : Šatan 2 , Kapuš) , MS 2004 (góly SR : Štrbák , Chára , Demitra) a v kvalifikačnej skupine MS 2005 (góly SR : Šatan , Marián Hossa , Pálffy) . (Denník Šport 06/02)

Morfológický analyzátor nepridelil ani jednému členovi bigramu komplexne správnu morfológickú značku. V oboch prípadoch ide o chybu na mieste pádu, v prípade adjektíva aj čísla. Správne mal byť označený singulár inštrumentálu. *Keď pritískala pery na matkinu nehybnú tvár , zaligotali sa jej < blond/AUfs3x vlasy/SSip4 > . (Harold Robbins - Park Avenue 79)*

Nesklonným je v tomto prípade adjektívum. Kategória slovného druhu a paradigmaty je správna, rod, číslo a pád sú nekorektné. Kategória pádu je nesprávne určená aj pri substantíve.

Na veľkoplšnej televíznej obrazovke v spoločenskej miestnosti premietali < video/AUip3x filmu/SSis2 > Monty Python. (Avery Corman - Najcennejšie, čo mám)

V niektorých prípadoch je pre nástroj veľmi náročné priradiť tvaru správnu morfológickú značku. Aj uvedený prípad je príkladom, keď nástroj priradil slovu *video* morfológickú značku adjektíva, hoci ide o nesklonné neutrum. Chybu je možné vysvetliť neprítomnosťou lexémy *video* v kategórii substantíva v aktuálnej verzii morfológického analyzátor.

12) Skratky

V materiáli skúmaných bigramov sa našli aj spojenia, kde druhým členom bigramu je skratka. Ide o 24 jedinečných výskytov, najfrekvencovanejšou

skratkou je skratka *DIČ*. Skratky a značky sú v koncepcii morfolologickej anotácie SNK označované morfológickou značkou *W*. Vzhľadom na pestrosť značiek používaných v písaných textoch neobsahuje v súčasnosti morfológický analyzátor všetky existujúce značky. Je teda prirodzené, že pri automatizovanej anotácii označil tager skratky ako substantíva.

Predpokladáme , že < novovytvorenému/AAis3x SFRK/SSis1 > bude na jeho činnosť pridelené jednorazovo 600 mil . Sk z prostriedkov Slovenskej konsolidáčnej a Ministerstva hospodárstva SR a ďalšie finančné zdroje z FNM SR , z predstupových , zo štrukturálnych a z ďalších rozvojových zdrojov EÚ , prípadne ďalšie zdroje , ktoré bude SFRK používať pri svojej činnosti revolvingovým spôsobom pri ozdravovaní jednotlivých hospodárskych subjektov . (Hospodárske noviny 02/06)

Schválená bola aj novela zákona o vlastníctve bytov a nebytových priestorov , ktorá má pri kúpe bytov zrovnoprávniť nájomcov bytov , ktoré patrili < bývalému/AAis3x ROH/SSis1 > s nájomcami iných bytov . (Hospodárske noviny 02/06)

Výzorom a spracovaním interiéru patrí XC 90 k < najelegantnejším/Aamp3z SUV/SSis1 > na trhu . (Plus 7 dní 2006/47)

V priebehu evolúcie sa metylácia DNA vyvinula v prokaryotických bunkách (baktériách) hlavne ako mechanizmus obrany proti < cudzorodej/AAfs3x DNA/SSfs1 > , pretože ju umožňuje odlišiť od vlastnej DNA na základe stupňa metylácie... (Eva Mišúrová, Zuzana Daxnerová - Molekulárno-biologické základy ontogenetického vývinu)

ADJEKTÍVUM V AKOMKOLIEK PÁDE OKREM DATÍVU A SUBSTANTÍVUM V DATÍVE

Korpus *prim-4.0-snk-all* obsahuje 35 542 výskytov opačného bigramu, teda adjektíva v inom páde ako v datíve a substantíva v datíve. Vyhľadávali sme pomocou príkazu: ([tag="A.*[124567].*"] [tag="S.*3.*"]). Vzhľadom na rozsiahlosť materiálu sme analyzovali takéto postavenie len v ručne morfológicky anotovanom subkorpuse *r-mak-3.0*, kde sa takýto bigram nachádza 65-krát. Najfrekvencovanejšou je na mieste adjektíva lema *podobný* (19), ďalej sú to lemy *blízky* (6), *verný* (5), *rovný* (5), *poslušný* (3), *úmerný* (3), *vďačný* (3), *prístupný* (2), *prospešný* (2). Na mieste substantíva sú to lemy *boh* (4), *rýchlosť* (4), *ježiš* (2), *rodina* (2).

Vo väčšine prípadov ide o adjektívum, ktoré je buď súčasťou menného prísudku (1. a 2. príklad), alebo je to atribút k inému substantívu a substantívum z bigramu je objektom.

Sme radi a < vďační/Aamp1x Bohu/SSms3 > , že brat Jozef mohol svojimi darmi slúžiť a pomáhať ECAV na rôznych postoch (Patmos3_1)



KATARÍNA GAJDOŠOVÁ

Sčasti je < podobný/AAis1x kauze/SSfs3 > Gaulieder . (SME-april2004)
Pravidlá o dňoch a čase , o povolených a zakázaných jedlách - pokrmoch sa
totiž vyvinuli v rituálny systém < podobný/AAis4x zákonom/SSip3 > farizejov
z dôb Ježišových . (Patmos)

Záver

Na to, aby mohol počítačový nástroj priradiť textovej jednotke správnu značku, je nevyhnutné, aby boli textové jednotky správne zapísané. Pravopisným, priekleповým a iným technickým chybám, ktoré sú súčasťou akéhokoľvek korpusu, sa tvorcovia takýchto textových databáz nikdy nevyhnú, keďže korpusy sa skladajú z reálne utvorených a vydaných textov. V štúdiu sme ukázali nielen na chybné texty od autorov či poskytovateľov textov, ktoré sú súčasťou databázy SNK, a na problémy pri ich automatizovanej morfolologickej anotácii, ale analýzou materiálu sme zistili aj zaujímavé konštrukcie spojenia adjektíva a substantíva v texte. Stoja vedľa seba, a predsa každé patrí do iného rodu či pádu, do inej syntagmy.

Pri automatizovanej anotácii tager pracuje so slovníkom morfolologickeho analyzátora, ktorý vyberá vhodné morfologické značky pre jednotlivé tvary slov v texte, a na základe ich zhody alebo podobnosti s ich okolím (bigramy, trigramy) a štatistických pravidiel ich priraduje jednotlivým tvarom. Ako ukázal realizovaný výskum, vo väčšine prípadov výskytu nesprávnych slov či tvarov v texte (časti Formálne chyby v texte, Rekcia, Cudzí slová končiace sa na -cia) im tager priradil správne morfologické značky zodpovedajúce danému tvaru. Nešlo teda o nekorektnosť tagera a morfolologickej značky, ale o chybu v pôvodnom texte. Tento druh chýb nemajú tvorcovia korpusov možnosť odstraňovať, pretože z hľadiska vyvíjania nástroja a pridelovania morfologických značiek jednotlivým tvarom v textoch ani nie je čo opravovať. Nástroje pracujú dobre, chyby sú v texte samotnom, a tie sa podľa koncepcie korpusov neodstraňujú.

Pri osobitnej analýze skúmaných bigramov adjektívum v datíve a substantívum v inom páde ako datíve sa ukázalo, že pri substantívach v nominatíve priraduje tager pády jednotlivým zložkám správne či už v prípadoch postponovaného adjektíva alebo vlastných mien. Vzhľadom na neúplnosť slovníka morfolologickeho analyzátora (ktorý sa však stále dopĺňa) sú v nominatívnom postavení zle označované vlastné mená. Problém s určením rodu, čísla či pádu (zväčša iba z týchto kategórií) vzniká pri elipsách v rámci skúmaného bigramu. Podobné problémy sú aj v bigrame, kde je substantívum v akuzatíve.

Genitívne substantíva stoja v bigrame so substantivizovanými adjektívami. Ich správne označenie ako substantív s adjektívnou paradigmou sa dosiahne po ich doplnení do slovníka morfolologickeho analyzátora.



Z pohľadu tagera považujeme za najnáročnejšie tie bigramy, kde je spojené adjektívum v datíve so substantívom v inštrumentáli. Ako sme ukázali v bode 6), ide o komplikované vnorené determinatívne konštrukcie. Tager síce dokáže správne určiť slovný druh a paradigmu, avšak pri určovaní rodu, čísla či pádu nie je úspešný. Podobným problémom sú cudzie mená či názvy v texte, ktorým tager priradí morfológickú značku na základe zakončenia, čo však nemusí byť správne z hľadiska ohýbania cudzieho mena v slovenčine.

Podrobná analýza materiálu prispeje k doplneniu morfológického analyzátoru o nové slová a ich paradigmy a k odstráneniu nesprávnych anotácií v textoch tam, kde prebehla automatizovaná anotácia nedôsledne. Perspektívne plánujeme analyzovať bigramy v kombinácii so všetkými pádmi, aby sa odhalili ďalšie súvislosti medzi členmi mennej skupiny a mohla sa skvalitniť morfológická anotácia aj v iných pádoch a ich kombináciách.

LITERATÚRA

Český národný korpus. Praha: Ústav Českého národního korpusu FF UK. Dostupný z WWW: <http://korpus.cz>

Encyklopédia jazykovedy. Bratislava: Obzor 1993. 513 s.

JELÍNEK, Tomáš: Neshoda adjektíva s následujícím substantivem. In: Korpusová lingvistika: Stav a modelové přístupy. Praha: Nakladatelství Lidové noviny - Ústav Českého národního korpusu 2006, s. 153 - 179.

Morfológický analyzátor. Bratislava: Jazykovedný ústav E. Štúra SAV 2009. Dostupný z WWW: <http://slovník.juls.savba.sk>.

Slovenský národný korpus - prim-4.0. Bratislava: Jazykovedný ústav E. Štúra SAV 2009. Dostupný z WWW: <http://korpus.sk>.

Slovenský národný korpus - r-mak-3.0. Bratislava: Jazykovedný ústav E. Štúra SAV 2008. Dostupný z WWW: <http://korpus.sk>.

LINGUISTIC ANALYSIS OF AUTOMATED ANNOTATION OF DATIVE OF ADJECTIVES IN HETEROGENEOUS NOMINAL GROUPS IN THE SLOVAK NATIONAL CORPUS

Key words: annotation, bigram, nominal groups, morphological markers, context

The aim of the paper is to analyze the annotation of the bigram adjective in the dative and substantive in a different case when they are located next to each other and in the Slovak National Corpus in version *prim-4-0-all* they are

KATARÍNA GAJDOŠOVÁ

not tagged in a unified way, i.e. when the following adjective or substantive is by the programme not assigned the same tag (morphological marker) in the category of gender, number or case. In the paper we analyze not only the situations in erroneous texts by the authors or providers of texts which form part of the Slovak National Corpus database, and the problems within their automated morphological annotation, but by analyzing the material we also reveal the interesting contextual positions of the adjective and the substantive in the nominal group or nominal groups. Although formally they stand next to each other, each of them belongs to a different gender or case depending on their appurtenance to another syntagm.