

KOMPIUTERINĖ LINGVISTIKA/ COMPUTATIONAL LINGUISTICS

Lietuvių kalbos leksemų morfologinis anotavimas: ypatumai ir sunkumai

Erika Rimkutė, Vidas Valskys, Jolanta Vaskelienė

Anotacija. Straipsnyje rašoma apie lietuvių kalbos morfologinio anotatoriaus veikimo principus, automatinės morfologinės analizės specifiką. Didžiausias dėmesys skiriamas vienam iš 2007–2008 m. Valstybinio mokslo ir studijų fondo remto projekto „Internetiniai ištekliai: anotuotas lietuvių kalbos tekstynas ir anotavimo priemonės (ALKA2)“ įgyvendintų darbų – lietuvių kalbos morfologinio anotatoriaus leksinės duomenų bazės pildymui. Išsamiai aprašoma į morfologinio anotatoriaus leksikos duomenų bazę įtrauktinų žodžių atranka, morfologinio anotavimo etapai, sunkumai, su kuriais susidurta atliekant šį darbą.

Morfologinio anotatoriaus leksikos bazė padidinta 24 000 žodžių (daugiausia tikrinių ir bendrinių daiktavardžių), todėl tikimasi, kad gana žymiai pagerės morfologinio anotatoriaus kokybė ir bus išvengta daugybės neatpažintų žodžių.

Šiame straipsnyje norėta parodyti anotavimo procesą; atskleisti, kad kyla sunkumų ne tik vertinant, ar nauji žodžiai teiktini, ar neteiktini, reikalingi lietuvių kalbai ar galima apsieiti be jų; sudėtinga ne tik nustatyti naujų žodžių reikšmes, bet taip pat ir analizuoti morfologiškai: nustatyti linksniavimo paradigmą, giminę, kaitymą skaičiais, darybinius vedinius ir pan.

Reikšminiai žodžiai: tekstynas; automatinė morfologinė analizė; morfologinis anotatorius; daiktavardžiai.

Įvadas

2007–2008 m. Valstybinio mokslo ir studijų fondo remtame projekte „Internetiniai ištekliai: anotuotas lietuvių kalbos tekstynas ir anotavimo priemonės (ALKA2)“ dirbusių mokslininkų atlikta nemažai lietuvių kalbos kompiuterizavimo darbų: papildytas „Dabartinės lietuvių kalbos tekstynas“ (toliau – DLKT) (<http://donelaitis.vdu.lt>), sukurti eksperimentiniai sintaksiškai ir semantiškai anotuoti tekstynai.¹ Dar vienas svarbus projekte numatytas ir įgyvendintas darbas yra susijęs su morfologine analize: parengtas morfologiškai anotuotas tekstynas, internete pateikta visiems prieinama morfologinio anotavimo programa, padidintas morfologinio anotatoriaus atpažįstamų žodžių kiekis ir papildyta leksinė duomenų bazė. Šiame straipsnyje trumpai aptarti automatinės morfologinės analizės veikimo principai (žr. 1 dalį), o didžiausias dėmesys skiriamas lietuvių kalbos žodžių morfologinio anotavimo ypatumams aparti (žr. 2 dalį).

1. Automatinė morfologinė lietuvių kalbos analizė

Lietuvių kalbos automatinės morfologinės analizės programą daugiau nei prieš 20 metų sukūrė Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centro programuotojas Vytautas Zinkevičius. Jis sukūrė „kompiuterinę žinių apie lietuvių kalbos leksiką ir gramatiką bazę, lingvistinės informacijos paieškos šioje bazėje ir jos panaudo-

jimo automatiško lemavimo proceso metodiką bei programinę įrangą“ (Zinkevičius, 2000).

Ilgą laiką pagrindinis šios programos trūkumas buvo negalėjimas išspręsti morfologinio daugiareikšmiškumo problemų. Nustatyta, kad automatinės morfologinės analizės metu išryškėja didelis lietuvių kalbos morfologinis daugiareikšmiškumas, t. y. reiškinys, kai tam pačiam žodžiui ar žodžio formai galima priskirti kelis antraštinius žodžius (lemas) (pvz.: *laužo* – daiktavardis *laužas* ir veiksmažodis *laužyti*) arba kelias gramatines formas (pvz.: *padaryti* – bendratis ir neveikiamosios rūšies vyriškosios giminės dalyvio *padarytas* daugiskaitos vardininkas). Beveik pusė lietuvių kalbos žodžių formų yra morfologiškai daugiareikšmės (plačiau žr. Rimkutė, 2006), o tai sukelia daug problemų norint automatiškai analizuoti kalbą ne tik morfologiškai, bet ir sintaksiškai, semantiškai, kuriant automatinio vertimo sistemas.

Įvade minėto projekto metu patobulinta morfologinės analizės programa², sukurta morfologinio vienareikšminimo programa. Pritaikius statistinius modelius ir panaudojus 1 mln. žodžių pusiau automatiškai parengtą morfologiškai anotuatą tekstyną (apie šį tekstyną plačiau žr. Rimkutė, 2006), sukurta vienareikšminimo įrankis, kuris įdiegtas morfologiniame anotatoriuje (plačiau apie vienareikšminimo programą žr. Rimkutė, et al., 2007).

¹ Tekstynai sudaryti iš 20 tūkst. žodžių. Šie tekstynai nėra prieinami internete. Projekto metu norėta parengti sintaksinio ir semantinio anotavimo principus.

² Šiuos darbus atliko V. Zinkevičius, V. Daudaravičius ir E. Rimkutė.

Morfologiniame anotatoriuje pritaikius vienareikšminimo funkciją, pasirenkant vieną iš kelių morfologiškai daugiareikšmių formų pasiekiamas apie 94 proc. tikslumas. Antrašinių lietuvių kalbos žodžių formų (lemų) nustatymo tikslumas net 99 proc. Iki šiol sunkiausiai vienareikšminami linksnių sinkretizmo atvejai (pvz., *gėlės* – vns. kilmininkas ir dgs. vardininkas); ne visada tiksliai nustatomos mažai ortografiškai besiskiriančių žodžių (pvz., *tona* – *tonas*, *būdas* – *būda*, *tėvas* – *tėvai*) lemos.

Morfologinis anotatorius yra prieinamas internete visiems vartotojams (žr. http://donelaitis.vdu.lt/main.php?id=4&nr=7_1). Norint gauti morfologiškai anotuotą ir vienareikšmintą tekstą, internete esančiame anotatoriaus lange reikia įrašyti arba įklijuoti žodžius ar tekstą. Anotuojami ne ilgesni kaip 50 000 simbolių tekstai. Rezultatą galima matyti internetinės naršyklės lange, paspaudus mygtuką „rezultatas puslapyje“, arba anotuotas tekstas bus įrašytas į tekstinį failą, paspaudus mygtuką „rezultatas faile“.

Anotuotą tekstą reikia skaityti vertikalčiai (žr. toliau pateiktą pavyzdį). Pirmiausia nurodoma anotuoto žodžio forma (pažyma „word“, pvz., *tikroji*), toliau – tos žodžio formos lema (anraštinis pavaldas, pvz., *tikras*) (pažyma „lemma“) ir morfologinė informacija apie anotuotą žodžio formą (pažyma „type“, pvz., *bdv.*, *teig.*, *nelygin. l.*, *įvardž.*, *mot. g.*, *vns.*, *V.*).

Nurodomi ir tarpai tarp žodžių (pažyma „space“), ir skyrybos ženklai (pažyma „sep“, pvz., taškas pažymimas taip: *sep="."*).

```
<word="Tikroji" lemma="tikras" type="bdv., teig. nelygin. l., įvardž., mot. g., vns., V.">
<space>
<word="kalbos" lemma="kalba" type="dkt., mot. g., vns., K.">
<space>
<<word="vartoseną" lemma="vartoseną" type="dkt., mot. g., vns., V.">
<space>
<word="atsiskleidžia" lemma="atsiskleisti(-džia,-dė)" type="vksm., teig., sngr., tiesiog. n., es. l., vns., 3 asm.">
<space>
<word="tik" lemma="tik" type="jng.">
<space>
<word="tekste" lemma="tekstas" type="dkt., vyr. g., vns., Vt.">
<sep=".">
<p>
```

2. Lietuvių kalbos žodžių morfologinio anotavimo ypatybės

Rašyta, kad vienas iš morfologinio anotatoriaus trūkumų – negalėjimas nustatyti, kuri morfologiškai daugiareikšmė forma pavartota. Šią problemą pavyko išspręsti sukūrus vienareikšminimo programą ir ją pritaikius anotatoriuje. Kitas morfologinio anotatoriaus trūkumas yra tai, kad neatpažįstami kai kurie tikriniai žodžiai, nemaža dalis naujažodžių (žinoma, čia neturima mintyje kasdien atsirandantys nauji žodžiai), užsienio kalbų intarpai. Morfologinis anotatorius neatpažįsta apie 6 proc. anotuojamų žodžių, o tai trukdo automatiškai analizuoti lietuvių kalbą kitais lygmenimis.

Pagrindinė morfologinio anotatoriaus duomenų bazės leksika buvo imta iš „Dabartinės lietuvių kalbos žodyno“ (1972) ir „Tarptautinių žodžių žodyno“ (1985) (Zinkevičius, 2000), todėl, norint, kad ji geriau atspindėtų dabartinės lietuvių kalbos vartoseną, leksikos duomenų bazę reikėjo papildyti naujais žodžiais.

Taigi vienas iš projekto „Internetiniai ištekliai: anotuotas lietuvių kalbos tekstynas ir anotavimo priemonės (ALKA2)“ darbų ir buvo papildyti morfologinio anotatoriaus leksikos duomenų bazę naujais žodžiais ir juos morfologiškai anotuoti. Įtraukti tik tikriniai ir bendriniai daiktavardžiai.³ Iš viso suanotuota apie 24 tūkst. naujų bendrinių bei tikrinių daiktavardžių ir šiais žodžiais papildyta morfologinio anotatoriaus leksikos duomenų bazė.

Anotavimo ir duomenų bazės pildymo darbas vyko dviem etapais: pirmiausia buvo atrinkti morfologinio anotatoriaus neatpažinti žodžiai (tikriniai ir bendriniai daiktavardžiai)⁴ (žr. 2.1 skyrių), antro etapo metu atrinkti žodžiai buvo pusiau automatiškai morfologiškai anotuojami (žr. 2.2 skyrių).

2.1. Naujų žodžių atranka

2.1.1. Tikrinių daiktavardžių atranka

Iš pirminio neatpažintų žodžių sąrašo (jį sudarė 15 000 žodžių) buvo atrinkti dažniausi daiktavardžiai ir sudarytas 8 000 tikrinių daiktavardžių sąrašas. Atrenkant neatpažintus žodžius buvo atskirti bendriniai daiktavardžiai, taip pat žodžiai, kurie gali būti ir bendriniai, ir tikriniai. Ši užduotis pasirodė ne tokia paprasta, nes labai daug žodžių buvo nežinomi (pvz., *Arkona*, *Arekipa*, *Asinovskaja*, *Mesenija*, *Melila*, *Miliva*, *Monca* (vietovardžiai), *Medžė* (moters asmenvardis), *Genova* (vietovardis ir moters asmenvardis), *Ankaja*, *Haka* (pavadinimai) ir kt., taigi jų ieškota DLKT, paieškos sistemoje *Google*, enciklopedijose⁵ ir kitur.

Šiuose šaltiniuose tikrinta didžioji dalis žodžių, norint įsitikinti, ar žodis realiai vartojamas, ar jis tikrinis, ar taisyklai parašytas (o klaidingai parašyti daiktavardžiai buvo ištaisomi ir įtraukiami į sąrašą). Tikrinant paaiškėjo, kad dalis iš pirmo žvilgsnio bendriniais laikytų žodžių kartais taip pat būna ir tikriniai (pavadinimai); tokiu atveju laikyta, kad žodis gali būti ir bendrinis, ir tikrinis (vėliau jie anotuoti skyrium)⁶.

Atrenkant tikrinius daiktavardžius ir juos anotuojant susidurta su keliomis leksinio daugiareikšmiškumo rūšimis:

³ Anotuojant pasitaikė keli būdvardžiai, jie taip pat įtraukti į leksikos duomenų bazę, bet plačiau apie tai šiame straipsnyje nėra šoma.

⁴ Dalis naujų žodžių buvo automatiškai išrinkta iš „Dabartinės lietuvių kalbos tekstyno“ ir pateikta lingvistams. Kitą dalį analizuotinų žodžių iš automatiškai sudaryto sąrašo turėjo atsirinkti patys lingvistai (neatpažintų žodžių atrankos ir morfologinio anotavimo darbus atliko V. Valskys ir J. Vaskelienė). Šiame straipsnyje rašoma apie sunkumus, kurių iškilo lingvistams peržiūrint sąrašą ir iš jo atrenkant tinkamus daiktavardžius.

⁵ Ypač dažnai remtasi „Dabartinės lietuvių kalbos tekstynu“, paieškos sistema *Google*, *Vikipedija*, „Vietovardžių žodynu“ (2007 metų internetiniu leidimu), „Visuotinė lietuvių enciklopedija“, K. Kuzavino ir B. Savukyno „Lietuvių vardų kilmės žodynu“ (2007), „Lietuvių pavardžių žodynu“ (1 t. 1985; 2 t. 1989) ir kt.

⁶ Žinoma, pavadinimu gali pasidaryti kone kiekvienas bendrinis žodis, bet šiuo atveju pasikliauta realios vartosenos pavyzdžiais, o ne potenciškai galimais.

1) homonimais (tos pačios fonetinės išraiškos vienodai tariamais žodžiais), pvz.: *Fortas* (tekstyno pavyzdžiai rodo, kad tai pavardė) ir bendrinis daiktavardis *fortas*; *Kitas* (miestas) ir įvardis *kitas*; *Krantas* (miestas) ir bendrinis daiktavardis *krantas*; *Tironas* (vietovardis) ir bendrinis daiktavardis *tironas*; *Gegužis* (pavardė) ir bendrinis daiktavardis *gegužis*; *Tarimas* (Kinijos regionas) ir bendrinis daiktavardis *tarimas*; *Šturmanas* (pavardė) ir bendrinis daiktavardis *šturmanas*; *Vyšniavas* (vietovardis) ir neteiktinas hibridas *vyšniavas*, analogiški *Morkavas* ir *morkavas*⁷ ir pan.;

2) homografais (skirtingai tariamais, bet taip pat rašomais žodžiais), pvz.: *Gėdės* (moters vardo kilmininkas) ir *ged* (veiksmažodžio būsimąjo laiko 3 asmuo); *Barà* (moters, vyro asmenvardis, pavadinimas) ir veiksmažodžio *barti* es. laiko 3 asmuo *bàra*; vyro asmenvardžio *Dr* dgs. vard. *Drąsia* ir priev. *dr*; *Bėrnas* (vietovardis) ir bendrinis daiktavardis *bėrnas*; *Durys* (vyro pavardė) ir bendrinis daiktavardis *dūrys*; *Malajà* (vietovardis) ir *Mālaja* (moters asmenvardis ir pavadinimas); *mėdija* (terpė) ir *medijà* (kuopinis medžių pavadinimas) ir t. t.;

3) homoformomis (atsitiktinai sutampančiomis skirtingų žodžių formomis), pvz.: tikrinio daiktavardžio *Marčius* dgs. vard. *Marčiai* sutampa su bendrinio dkt. *marti* vns. naudininku *marčiai*.

2.1.2. Kitų kalbos dalių atranka

Pildant morfologinio anotatoriaus leksikos duomenų bazę kitais nei tikriniais daiktavardžiais žodžiais, peržiūrėta per 7000 žodžių, daugiausia sudurtinių ar hibridinių daiktavardžių. Reikėjo sužiūrėti jų rašybą (ar taisyklą), pakeisti ar iš pateiktųjų palikti tik pagrindines, nežymėtąsias formas – vyriškosios giminės (jeigu žodis kaitomas giminėmis) vienaskaitos (daugiskaitos, jeigu nevartojama vienas-kaita) vardininkus, kad toliau šiuos žodžius būtų galima gramatiškai anotuoti. Kadangi akiratyje buvo dabartinėje kalboje labai retai vartojamų (pvz., *žirgvaldys*, *baisatis*) ar tik tam tikrų sričių profesinėje kalboje tepasitaikančių žodžių (pvz., *bakteriocenoze*, *neurooftalmologinis*), kilo nemažai abejonių dėl žodžių antraštinių morfologinių charakteristikų, todėl neretai buvo tikrinamasi DLKT, žodynuose, žinynuose, enciklopedijose, internete (pasitelkiant *Google* paieškos sistemą). Tačiau net ir papildomos priemonės ne visada gelbėjo nustatant nežymėtąsias ir ortografiškai tinkamas žodžių formas.

Pasitaikė nemažai rūgštis žyminčių pavadinimų, pvz.: *acetamidobenzoinė*, *aminokaproninė*, *antisensinė*, *arachidinė* ir daugelis kt. Kilo klausimas, ar šių būdvardžių vyriškoji giminė realiai egzistuoja, ar ji tik potenciali. Vis dėlto pagal analogiją su kitais būdvardžiais nuspręsta kaip antraštinę teikti vyr. g. formą, nors jų (bent patikros šaltiniuose) ir nepasitaikė. Tiesa, buvo ir tokių vardažodžių, kurių vyriškoji giminė kažin ar įmanoma, pvz.: *didžiakrūtė*, *nuogakrūtė*, *statpapė*, *kotrynietė* (vienuolių ordino narė), *podbielskinė* (tokia vyšnių veislė), *edinburginė* (bulvių veislė) ir kt.

Iškilo rūpestis dėl dalyvių formų. Tradicinėje gramatikoje jos laikomos išvestinėmis veiksmažodžių formomis, bet tik bendratį ar asmenuojamąsias formas palikti būtų buvę ne-

tikslinga, nes anotavimo programa, ko gero, dalyvius būtų „ignoravusi“. Kadangi dalyvių morfologiniai parametrai labai skiriasi nuo veiksmažodžių finitinių formų, dėl to jie buvo palikti kaip atskiros leksemos (pvz.: *agresyvėjantis*, *besislapstantis* ir t. t.).

Tam tikrų apmąstymų sukėlė bendrinei kalbai neįprastų lyčių vertinimas. Pavyzdžiui, analizuotame sąraše buvo pateikti net du žinomos svetimbės variantai *baranka* ir *baronka*. Kadangi dabartinėje vartosenoje abi lytys labai gajos, abidvi ir paliktos. Įtraukta žargonybė *tūsintis*, o tik žemaičiuose žinomos *dirmavonės* atsisakyta.

Radosi nemenkas pluoštas žodžių, kurių pateiktis „Lietuvių kalbos žodyne“ įvairuoja – atskirais straipsniais nurodyti darybiniai sinonimai (*atitvara atitvaras*; *dumplūnė dump-lainis*) arba darybiniai variantai⁸ (*darbymetė darbymetis*; *daugmintis daugiamintis*; *didnosis didžianosis*; *daugkalbystė daugiakalbystė*; *skoliaraštis skolarastis*; *garsaeilis garsiaeilis* ir t. t.). Kadangi tekstyno duomenys rodo dabartinėje kalboje esant abu žodžių variantus, nuspręsta juos anotuoti skyrium. Tiesa, išbrauktas *planšetas*, nes DLKŽ ir TŽŽ yra tik *planšetė*.

Iškilo klausimas, ar atskirais leksikos vienetais laikytinos raidinės santrumpos (pvz., *BSSE* (rinkmenoje viskas buvo teikiama mažosiomis raidėmis), neįprastai užrašytas ir neįprastai skambantis braškių veislės pavadinimas *honeyoe*, nuo vienuolikos iki devyniolikos metų įvardijimas *niolik-metis*. Liko iki galo neaišku, kaip vertinti pramanytus, realiai nesančius, todėl kabutėse rašomus žodžius *šokoterapi-nis* ar *cibukinė*, rašytojų vaizduotės sukurtą *tiputapę* (V. Žilinskaitės apysakos pavadinimas), dirbtinai sukonstruo-tus *urkizmą* (matyt, iš pavardės *Urkis*), bendrinei kalbai neteiktiną *viltietį* (iš pavadinimo „Viltis“) ir pan. Plg. teks-tyno pavyzdį *Beje, daugelis ir mūsų žodžių vėl aitėmis bei ytėmis patapo*. Neaiškumų sukėlė ir sudėtinių vietovar-džių nekaitomi dėmenys, pvz., *Porto Alegre*, *San Žermen de Prė*. Neaišku, ar nekaitomus dėmenis reikia anotuoti atskirai nuo pirmojo.

Kai kurių svetimos kilmės daiktavardžių dorojimą apsun-kindavo nevienoda tų pačių žodžių perraša, pvz., *kapucinė* ir *cappuccinė*, *Donicetė* ir *Donizetė*. Nuspręsta teikti tik vieną, lietuvių rašybos taisyklės atitinkantį, variantą.

2.2. Naujų žodžių anotavimas

Atrinkus analizuotinus žodžius, kitas etapas – atrinktų tik-rinių ir bendrinių daiktavardžių anotavimas. Anotavimas vyko pusiau automatiškai: naudotas morfologinio anota-vimo „sufleris“. Kiekvienas žodis iš sudaryto žodžių sąrašo buvo analizuojamas atskirai pagal tokius morfologinius požymius: priskiriami kuriai nors linksniavimo paradig-mai, giminei, skaičiui (vienaskaitinis, daugiskaitinis, abis-kaitinis ar nekaitomas daiktavardis), nustatoma tikriniš-kumo rūšis (bendrinis ar tikrinis, jei tikrinis, reikėjo nuro-dyti tikriniškumo rūšį), darybos galimybės: ar gali turėti neigiančių priešdėlių, deminutyvų (jei taip, tai kokio pobū-džio), priesaginių vedinių (žr. priede pateiktas lentelės).⁹

⁸ Apie darybinius sinonimus ir darybinius variantus žr. Vaskelienė, 2000, pp.5-13.

⁹ Šios gramatinės kategorijos ir kai kurie jų terminai paimti iš V. Zinkevičiaus morfologinio anotatoriaus.

⁷ *Google* duomenys rodo, kad priesagos *-avas* vediniai yra vietovardžiai.

Kai kuriais atvejais anotavimo programa pasiūlydavo variantą (kaitymo paradigmos, giminės, skaičiaus kategorijų, dažniausiai jis būdavo tinkamas, nors kartais reikėdavo rinktis kitokį), kitais atvejais analizuojantys lingvistai patys turėdavo pasirinkti iš nurodytų gramatinių kategorijų.

Taigi nauji žodžiai buvo analizuojami nurodant anksčiau minėtus gramatinius ir semantinius požymius. Toliau paminėtos problemos, su kuriomis susidurta anotuojant tikrinius (žr. 2.2.1.)¹⁰ ir bendrinius (žr. 2.2.2.) daiktavardžius.

2.2.1. Tikrinių daiktavardžių anotavimas

Nes visada buvo lengva nustatyti daiktavardžių su galūne *-a giminę*¹¹, nes labai daug tokių žodžių yra nelietuviškos kilmės. Reikėjo ieškoti tekstų, kuriuose vartojami šie tikriniai daiktavardžiai, ir iš konteksto išsiaiškinti tų žodžių giminę, pvz., taip nustatyta daiktavardžių *Dubna*, *Bohmová* ir kt. moteriškoji giminė; daiktavardžių *Alcala*, *Manjara*, *Gura*, *Ata*, *Bodmova*, *Jama* ir kt. vyriškoji giminė. Radosi ir tokių žodžių, kurie gali būti interpretuojami ir kaip vyriškosios, ir kaip moteriškosios giminės daiktavardžiai, tada žodis anotuotas du kartus, pvz., *Bula* (mot. g. vietovardis) ir *Bula* (vyr. g. pavardė), *Luka* (mot. g. vardas ir ežeras) ir *Luka* (vyr. g. asmenvardis), *Baskija* (mot. g. vietovardis) ir *Baskija* (vyr. g. asmenvardis), *Davila* (mot. g. pavadinimas) ir *Davila* (vyr. g. pavardė) ir pan.

Panašių problemų kilo ir su *-ė* galūnę turinčiais daiktavardžiais. Pavyzdžiui, *Heinė*, *Brontė*, *Ševardnadzė*, *Abuladzė*, *Mugabė* gali būti interpretuojami ir kaip vyrų, ir kaip moterų pavardės. Anotuojant šiuos žodžius svarstyta, kad gal vertėtų įsivesti kokią naują tokius žodžius tiksliau apibūdinančią pažymą, nes anotacija *abigimininis* (pagal analogiją su *abiskaitinis*) nėra tikslinga – morfologinės analizės programa tiesiog bus nepajėgi atsirinkti, kurios giminės žodis pavartotas. Nuspręsta tokiais atvejais anotuojant vartoti bendrosios giminės sąvoką, suteikiant jai gerokai sąlygišką ir platesnę sampratą nei įprasta tradicinėje morfologijoje. Gal giliau padiskutavus ir apmąsčius rastųsi ir kokia kita tokių asmenvardžių traktavimo išeitis.

Gerokai sunkiau pasidavė anotuojami žodžiai, kurie gali būti suvokiami tiek kaip vietovardžiai, tiek kaip asmenvardžiai ir priskiriami tiek vyr., tiek ir mot. giminei. Pavyzdžiui: *Sula* vyro pavardė upė; *Kama* vyro vardas upė; *Balduonė* Latvijos kurortas vyr. pavardė; *Griežė* vyr. pav. kaimas (mot. g.), *Valeta* argentiniečio vyro pav. Maltos sostinė ir t. t. Tokius žodžius nuspręsta įrašyti dukart ir anotuoti dvejopai – ir kaip vyriškosios giminės, ir kaip moteriškosios giminės skirtingo skaitiškumo tikrinius daiktavardžius.

Nustatant anotuojamų žodžių **skaičių**¹², laikyta, kad vietovardžiai (miestai, kaimai, upės, kalnai ir kt.) paprastai yra vienaskaitiniai (pvz.: *Bodhgaja*, *Borovskojė*, *Bretanija*, *Chodynka*, *Chortica*, *Doha*, *Dusnyčia*, *Džerba*, *Bakuba*, *Bielica*) daiktavardžiai. Žinoma, pasitaikė ir keletas daugiskaitinių vietovardžių pavyzdžių (*Trakai*, *Justiniškės* ir

pan.). Itin retais atvejais, kai vartojama ir vienaskaita, ir daugiskaita, žodis laikytas abiskaitiniu, pvz., bažnytkaimis Baltarusijoje vadinamas *Bokštas* ir *Bokštai*.

Kadangi vienaskaitiniai asmenų vardai ir pavadinimai (tą patį pavadinimą kartais turi keletas klubų, įmonių ir pan.) tam tikrais atvejais gali būti vartojami ir daugiskaitos forma (pvz., *Mūsų miestelyje buvo dvi Eidukienės*, *Kuris iš trijų Dunauskų turėjo atvykti? Heda tuo metu buvo populiarus moters vardas, taigi neretai mokykloje būdavo keletas Hedų*, *Atneškite man visų metų Lituanisticas*), jie buvo anotuojami kaip abiskaitiniai (t. y. turintys ir vienaskaitą, ir daugeliu atvejų potencinę daugiskaitą). Tiesa, tai šiek tiek kertasi su morfologų teiginiais, kad tikriniai žodžiai žymi individualius daiktus, todėl negali būti kaitomi skaičiais (Paulauskienė, 1994). Apsvarsčius, kaip reikėtų elgtis su metonimijos atmainos antonomazijos (apie šią tropą rūši plačiau žr. Koženiauskiene, 2001) atvejais (*Nusibodo man tie šekspyrai* ar *Nereikia man tų nicų* ir pan.), nutarta nuo tokių atvejų atsiriboti, nes taip vartojamas tikrinis žodis virsta bendriniu (pvz., *Ar kas mokate morzę?*).

Iškilo klausimas, kokį skaičių nurodyti, kai tas pats daiktavardis būna vienaskaitinis (pvz., vietovardis) ir abiskaitinis (pvz., asmenvardis, pavadinimas); nutarta nurodyti, kad jis – abiskaitinis, pvz.: *Dambrovica* (vietovardis ir mot. g. asmenvardis) dėtas į abiskaitinių grupę; *Džida* (vietovardis ir upė) – į vienaskaitinių grupę (nes ir upė, ir vietovardis skaičiais nekaitomi žodžiai).

Daugiausia keblumų sukėlė tikrinių daiktavardžių daugia-reikšmiškumo atvejai. Pastebėta, kad griežtos ribos tarp asmenvardžių ir vietovardžių nėra – nemažai analizuotų daiktavardžių buvo ir asmenų (vardas, pavardė), ir vietos (šalies, regiono, vandens telkinio, salos ir pan.) pavadinimas, pvz., *Kriaunelis* pavardė ežeras, *Ūla* vardas upė, *Rusnė* mot. vardas miestas upė, *Elmė* mot. vardas upė, analogiški *Bergas*, *Lichtenbergas*, *Ofenbachas*, *Krachas*, *Moronis*, *Vankuveris*, *Lembas*, *Chemnicas*, *Sandvičas*, *Raigardas*, *Herefordas*, *Braunšveigas*, *Ilgas*, *Dergas*, *Liu-nenburgas*, *Pelekas*, *Dubininkas*, *Žuvininkas* ir t. t. Kadangi vardai ir pavardės, kaip jau minėta, laikyti abiskaitiniais daiktavardžiais, tokius dvejopai traktuojamus žodžius nuspręsta ir anotuoti kaip abiskaitinius (jeigu tekste pasitaikys vietovardis, jis turės tik vienaskaitos formą, jei asmenvardis – galimi ir daugiskaitos linksniai).

Tapačiai buvo pasielgta su žodžiais, kurie reiškia tiek vietos pavadinimą, tiek ir nepriesaginę moters pavardę (*Jauniškė* (iš *Jauniškis*), *Kukarskė*, *Alksninė*, *Pilvė*, *Bukantė*, *Bijotė*, *Padegimė*, *Paberžė*, *Bražuolė*, *Raudonė*, *Sidabrė* ir pan.). Tiesa, buvo ir tokių, kurie laikyti vienaskaitiniais, pvz.: *Kreta* graikų deivė sala; *Troja* deivė vietovė. Dievų pavadinimai laikyti vienaskaitiniais daiktavardžiais (dievų ar dievių vardų daugiskaita negalima), pvz., *Erebas*, *Gabjauja*, *Šiva*, *Prometėjas*, *Himenėjas*, *Andrajas*, *Bachas*, *Asklepijas*, *Baubis*, *Anubis*, *Apis*, *Višnus*, *Svaikstikas*, *Jumis*, *Ūsinis*, *Žvaigždikas*, *Panas*, *Mardukas* ir kt., bet kai greta kai kurių dievų pavadinimų vartojami tokie pat asmenvardžiai, daiktavardis laikomas abiskaitiniu, pvz., *Morfėjas*, *Helijas*, *Bubilas*, *Eolas* ir pan.

¹⁰ Apie tikrinių daiktavardžių anotavimo ypatumus taip pat žr. Valskys, 2008.

¹¹ Giminės kategorijos pasirinkimo ypatybės pateiktos 3 lent.

¹² Skaičiaus kategorijos pasirinkimo ypatybės pateiktos 2 lent.

Buvo nelengva nustatyti ir **tikriniškumo rūšį** (žr. priedo 1 lent.). Tikriniams daiktavardžiams reikėjo parinkti rūšį, nurodyti, kas konkrečiai (geografinis pavadinimas, asmenvardis, firmos ar kt. vardas) yra anotuojamas daiktavardis. Taigi nustatant tikriniškumo rūšį, neretai reikėdavo peržiūrėti daug vartosenos pavyzdžių. Tais atvejais, kai analizuoti daiktavardžiai buvo ir asmenų, ir vietų ar kitokie pavadinimai ir nesiskyrė jų giminė, nutarta įvairių tikriniškumo rūšių daiktavardžius dėti į grupę *Kiti tikriniai vardai*, pvz., tokie būtų moteriškosios giminės tikriniai daiktavardžiai *Dambrovica* (asmenvardis ir vietovardis), *Estela* (vardas ir vietovardis), *Elija* (asmenvardis, vietovardis, pavadinimas), *Ilka* (asmenvardis ir vietovardis), *Krasnaja* (upė ir kaimas), *Kaja* (asmenvardis ir pavadinimas), *Girona* (vietovardis ir pavadinimas), *Katniava* (vietovardis ir pavadinimas) ir t. t.

Nustatant daiktavardžių pobūdį (žr. priedo 4 lent.) (ar jis įprastas daiktavardis, ar vyriškosios giminės moteriškai linksniuojamas, ar fleksiškai nekaitomas ir pan.) didelių sunkumų nekilo.

Anotuojamiems žodžiams reikėjo parinkti ir kai kuriuos **darybos** aspektus (žr. priedo 6, 7 lent.): daugeliu atvejų nurodyta, kad anotuojami daiktavardžiai gali turėti įprastus deminutyvus, nes „reikalui esant, deminutyvų <...> gali būti pasidaroma kone iš kiekvieno daiktavardžio ir paprastai ne tik su kuria nors viena, bet su keliomis priesagomis“ (LKG, 1965). Spręsta, prie kurių iš anotuojamų daiktavardžių gali būti pridėdama vardažodinės ypatybės priesagos pavadinimų pati daro priesagą *-ininkas*, *-ė*. Tokių vedinių itin daug sudaroma iš įvairiausios semantikos daiktavardžių. Nutarta, kad galimi priesagos *-ininkas* vediniai iš vietas ir įvairius pavadinimus reiškiančių daiktavardžių (pvz., *mitrovininkas*, *moncininkas*, *edininkas*, kuriais pavadinamas žmogus, kaip nors susijęs su vietovėmis *Mitrovica*, *Monca* arba dirbantis *Edos* firmoje), o iš asmenvardžių – tik kai kuriais atvejais (pvz., *muravjovininkas*: *Šis žmogus yra tikras muravjovininkas – perskaitęs viską apie tą istorinį veikėją*). Nustatyta, kad daugeliu atvejų bent jau potenciškai galimi vediniai su priesagomis *-iškas* ir *-inis*, o kitas priesagas anotuojami daiktavardžiai turi gerokai rečiau. Anotuojamiems daiktavardžiams nebūdinga ir prefiksinė daryba.

Dar vienas iš analizės etapų – tinkamos **paradigmos** pasirinkimas. Anotuojant reikėjo pasirinkti iš 37 paradigmų. Toks kaitymo paradigmų skaičius pasirinktas dėl to, kad norėta, jog paradigmoje nebūtų išimčių, taip pat atsižvelgta į kamiengalio minkštumą, priebalsių kaitą tam tikruose linksniuose, nes priešingu atveju automatinės morfologinės analizės metu būtų daroma klaidų. Paradigmų pavadinimais pasirinkti konkretūs žodžiai, nes manyta, kad anotuojantiems lingvistams žodžius lengviau priskirti vienai ar kitai paradigmai, kai paradigma pavadinta ne skaičiumi, o konkrečiu žodžiu. Toliau pateikti anotavimo procese naudotų paradigmų pavadinimai: 1) *taksi* (0¹³), 2) *vyras* (1), 3) *elnias* (1), 4) *kelias* (2), 5) *vėjas* (2), 6) *svečias* (2), 7) *brolis* (3), 8) *akimojis* (3), 9) *peilis* (3), 10) *užantis* (3), 11)

briedis (3), 12) *arklys* (3), 13) *vabzdys* (3), 14) *paprotys* (3), 15) *žvejys* (3), 16) *turgus* (4), 17) *žmogus* (4), 18) *koridorius* (5), 19) *pavojus* (5), 20) *jūra* (6), 21) *sauja* (7), 22) *žinia* (7), 23) *marti* (7), 24) *gervė* (8), 25) *cigaretė* (8), 26) *dildė* (8), 27) *buržujė* (8), 28) *krošnis* (9), 29) *grandis* (9), 30) *ietis* (9), 31) *žvėris* (10), 32) *dantis* (10), 33) *debesis* (10), 34) *akmuo* (11), 35) *šuo* (11), 36) *sesuo* (12), 37) *duktė* (12).

Dėl daugiareikšmiškumo kai kuriais atvejais tą pačią leksemą reikėjo analizuoti du kartus. Pavyzdžiui, du kartus reikėjo užrašyti ir skirtingai pažymėti žodžius *Maris* ir *Siesartis*, nes (priklausomai nuo reikšmės) skiriasi jų visi morfologiniai parametrai: *Maris* vyro vardas pavardė mot. pavardė (nelinksniuojama), todėl anotauta *Maris* – paradigma *Peilis*, vyriškoji g., abiskaitinis ir *Maris* – paradigma *Taksi*, moteriškoji g., nekaitomas; *Siesartis* (-ies) upė ir *Siesartis* (-čio) ežeras, taigi fiksuota *Siesartis* – paradigma *Ietis*, moteriškoji g., vienaskaitinis ir *Siesartis* – paradigma *Užantis*, vyriškoji giminė, vienaskaitinis.

Susidurta ir su žodžiais, kurie kirčiuojami paskutiniame skiemenyje ir paprastai nekaitomi, pvz., *Olivj* □ ar *Mej* □. Jie priskirti nekaitomų *taksi* tipo daiktavardžių paradigmoms.

Sudėtiniai vietovardžiai kartais sudaryti iš kaitomų ir nekaitomų dėmenų. Kadangi tai neskaidomi pavadinimai, anotavimo programoje jie buvo pateikti su brūkšneliu (*Gran-Kanarija*, *Saudo-Arabija* (plg. *Arabija* (pusiasalis); *Bisau-Gvinėja*; *Kazlų-Rūda*, *Rio-Saladas*, *Las-Vegas*, *Rod-Ailandas* ir pan.) ir anotuojami kaip viena leksema, kurios giminę ir skaičių rodo antras, kaitomas, dėmuo.

2.2.2. Bendrinių daiktavardžių anotavimas

Kadangi bendrinių daiktavardžių į morfologinio anotatoriaus leksikos duomenų bazę įtraukta mažiau nei tikrinių, todėl kilo ir mažiau sunkumų anotuojant bendrinius daiktavardžius. Kaip ir anotuojant tikrinius daiktavardžius, dažnai reikėdavo bendrinei kalbai būdingus reikalavimus derinti su praktine analize, todėl kai kurie anotavimo aspektai gali prieštarauti tradicinės kalbotyros nuostatoms. Pavyzdžiui, prie daiktavardžio *pseudokalba* buvo pažymėta, kad jis gali turėti priesagos *-ininkas* vedinių, nors tai ir neatitinka žodžių darybos principų – *pseudokalbininkas* yra vedinys iš *kalbininkas*, pridėjus lotyniškąjį elementą *pseudo-*, o ne iš *pseudokalba*.

Neaiškumų anotuojant žodžius kilo ir dėl to, kad įtrauktinų žodžių sąrašė buvo ne tik bendrinės kalbos žodžių, bet ir žargonybių, tarmybių, pvz., žodis *auklyba* gali būti suvokiamas ne tik kaip auklėjimo proceso pavadinimas, bet ir kaip moksleivių darybinė žargonybė (daiktavardžio *auklėtoja* darybinis sinonimas, sudarytas su veikėjų pavadinimams nebūdinga priesaga), todėl šį žodį reikėtų skirtingai anotuoti.

Kai kada tekdavo kliautis tik nuovoka ir tekstyno duomenimis, pvz., analizuojant žodį *orgtechstatyba* – jis buvo priskirtas prie firmų vardų, bet gal gali būti ir vartojamas ir kaip bendrinis daiktavardis.

Kitos problemos, pvz., paradigmos nustatymas, buvo bendros ir tikriniais, ir bendriniais daiktavardžiams, todėl išsamiau čia neminimos.

¹³ Skliaustuose nurodytos „Dabartinės lietuvių kalbos gramatikoje“ (1996) pateiktos paradigmos. Išimtis tik paradigma *taksi*, kuri nenurodyta gramatikoje, o šioje analizėje pažymėta kaip nulinė paradigma.

Išvados

Morfologinės analizės programos leksinė duomenų bazė padidinta 24 000 žodžių (daugiausia tikrinių ir bendrinių daiktavardžių), todėl tikimasi, kad gana žymiai pagerės morfologinio anotatoriaus kokybė ir bus išvengta daugybės neatpažintų žodžių. Dabar morfologinis anotatorius neatpažįsta apie 6 proc. anotuojamų žodžių, tikimasi, kad įtraukus projekto metu anotuotus naujus žodžius neatpažintų žodžių sumažės iki 3 proc.

Šiame straipsnyje norėta parodyti anotavimo procesą; atskleisti, kad kyla sunkumų ne tik vertinant, ar nauji žodžiai teiktini, ar neteiktini, reikalingi lietuvių kalbai ar galima apsieiti be jų; sudėtinga ne tik nustatyti naujų žodžių reikšmės, bet taip pat ir analizuoti morfologiškai: nustatyti linksniavimo paradigmą, giminę, kaitomumą skaičiais, darybinius vedinius ir pan.

Straipsnyje nesistengta pateikti atsakymų į visus anotuojant iškilusius klausimus. Ateityje automatinio morfologinio anotatoriaus duomenų bazė turėtų būti pildoma, galbūt tada vėl bus grįžta prie neaiškių anotavimo atvejų, bus išspręsti kai kurie diskusiniai dalykai.

Įtraukiant naujus žodžius į morfologinio anotatoriaus leksikos duomenų bazę, kaip ir atliekant kitokio pobūdžio

automatinę ar pusiau automatinę kalbos analizę, pastebėta, kad dažnai teoriniuose kalbiniuose darbuose negalima rasti aiškaus atsakymo; kartais reikia nusižengti tradicinėms teorinėms nuostatoms, atsižvelgiant į kompiuterinių programų specifiką.

Literatūra

1. Ambrazas, V. red., 1996. *Dabartinės lietuvių kalbos gramatika*. Vilnius: Mokslo ir enciklopedijų leidykla.
2. Koženiauskienė, R., 2001. *Retorika: iškalbos stilistika*. Vilnius: Mokslo ir enciklopedijų leidybos institutas.
3. Paulauskienė, A., 1994. *Lietuvių kalbos morfologija. Paskaitos lituanistams*. Vilnius: Mokslo ir enciklopedijų leidykla.
4. Rimkutė, E., 2006. *Morfologinio daugiareikšmiškumo ribojimas kompiuteriniame tekste*. Daktaro disertacija. Kaunas: Vytauto Didžiojo universitetas.
5. Rimkutė E., Daudaravičius V., 2007. Morfologinis dabartinės lietuvių kalbos teksto anotavimas. *Kalbų studijos*, nr. 11, pp.30-35.
6. Ulvydas, K. vyr. red., 1965. *Lietuvių kalbos gramatika*, 1 t. Vilnius: Mintis.
7. Valskys, V., 2008. Lietuvių kalbos tikrinių daiktavardžių morfologinio anotavimo ypatumai. *Valoda dažadu kultūru kontekstā*, pp.181–186.
8. Vaskelienė, J., 2000. *Darybiniai sinonimai*. Šiauliai.
9. Zinkevičius, V., 2000. *Lemuoklis – morfologinei analizei. Darbai ir dienos*, nr. 24, pp.245-274.

Erika Rimkutė, Vidas Valskys, Jolanta Vaskelienė

Annotation of Lithuanian Lexemes: Peculiarities and Problems

Summary

The article presents the principles of the morphological annotator and the peculiarities of automatic morphological analysis. The paper focuses on building the lexical database of the Lithuanian morphological annotator, which is one of the completed tasks of the project *Internet Resources: Annotated Corpus of the Lithuanian Language and Tools of Annotation (ALKA 2)*, implemented in 2007-2008 and sponsored by the Lithuanian State Science and Studies Foundation. The selection of words to be included into the lexical database of the morphological annotator is described in detail. The stages of morphological annotation and difficulties in this paper are also discussed.

The lexical database of the morphological annotator has increased by 24 000 words (mostly proper and common nouns). Therefore it is expected that the quality of the morphological annotator will improve considerably and many unrecognized words will be avoided.

The goal of the article is to show the process of annotation. It reveals that problems arise not only during the evaluation of acceptability of new words for the Lithuanian language and the identification of their meanings, but also during their morphological analysis. It is difficult to determine their declension paradigms, gender, number inflection, derivatives, etc.

Straipsnis įteiktas 2009 11
Parengtas spaudai 2009 11

Apie autorius

Erika Rimkutė, humanitarinių mokslų daktarė, Vytauto Didžiojo universiteto Lietuvių kalbos katedros docentė, Kompiuterinės lingvistikos centro mokslo darbuotoja.

Mokslinės veiklos sritys: tekstynų lingvistika, kompiuterinė lingvistika, automatinė morfologinė analizė ir sintezė, morfologinis daugiareikšmiškumas, automatinė sintaksinė analizė.

Adresas: Vytauto Didžiojo universitetas, Humanitarinių mokslų fakultetas, Kompiuterinės lingvistikos centras, K. Donelaičio g. 52–206, 44244 Kaunas
El. paštas: e.rimkute@hmf.vdu.lt

Vidas Valskys, humanitarinių mokslų daktaras, Vilniaus pedagoginio universiteto Lietuvių kalbos katedros docentas.

Mokslinės veiklos sritys: lietuvių kalbos sintaksė, kalbos kultūra, specialybės kalba, kompiuterinis morfologinis anotavimas.

Adresas: Vilniaus pedagoginis universitetas, Lituanistikos fakultetas, Lietuvių kalbos katedra, T. Ševčenkos g. 31, 03111 Vilnius.
El. paštas: vidasval@gmail.com

Jolanta Vaskelienė, humanitarinių mokslų daktarė, Šiaulių universiteto Lietuvių kalbotyros ir komunikacijos katedros docentė.

Mokslinės veiklos sritys: žodžių daryba, morfologija, sinonimija, teksto lingvistika.

Adresas: Šiaulių universitetas, Humanitarinis fakultetas, Lietuvių kalbotyros ir komunikacijos katedra, P. Višinskio g. 38, 76352 Šiauliai
El. paštas: v.jolanta@splius.lt

1 lentelė. Daiktavardžių tikriniškumo rūšys

Nr.	Tikriniškumo rūšies aprašymas
1	Bendriniis
2	Asmenvardis – vardas
3	Ne tikrinis, bet iš pagarbos gali būti traktuojamas (t. y. rašomas didžiąja raide) kaip tikrinis (pvz., <i>Dievas, Altingas, Koranas</i> ir pan.)
4	Tikrinis geografinis vardas – miestas, kaimas, šalis, gyvenvietė, t. y. toks, iš kurio galimi vediniai su <i>-ietis (-ė)</i> ir pan.
5	Kiti tikriniai vardai, pvz., firmų pavadinimai
6	Tikrinis geografinis vardas: upės, ežerai, kalnai ir pan.
7	Bendriniu virtęs vardas (gali būti rašomas ir mažąja, ir didžiąja raide), pvz., <i>Eskulapas – eskulapas</i>
8	Bendriniu virtęs tikrinis geografinis vardas (gali būti rašomas tiek mažąja, tiek didžiąja raide), pvz., <i>Kvirinalas – kvirinalas</i>
10	Bendriniu virtęs kitas tikrinis vardas (pvz., antikinių švenčių), pvz., <i>herėjos – Herėjos</i>

2 lentelė. Daiktavardžių skaičiaus kategorijos nustatymas

Nr.	Skaičiaus kategorijos aprašymas
0	Daiktavardis skaičiais nekaitomas
1	Turintis ir vienaskaitą, ir daugiskaitą ¹⁴ , pvz., <i>stalas</i>
2	Turintis ir vienaskaitą, ir daugiskaitą <i>analgetikų</i> tipo: nors žodynuose (dažniausiai) pateikiamas daugiskaitos forma, bet gali įgyti ir vienaskaitos formas, pvz., <i>analgetikai – analgetikas</i>
3	Daugiskaitinis, pvz., <i>miltai, žirklys</i>
4	Vienaskaitinis
5	Labiau vienaskaitinis nei abiskaitinis, nes tai tikrinis vardas, pvz., <i>Achilas</i>
6	Labiau daugiskaitinis nei abiskaitinis, pvz. <i>Parkos</i> (tikrinis iš TŽŽ)

3 lentelė. Daiktavardžių giminės kategorijos nustatymas

Nr.	Giminės kategorijos aprašymas
1	Įprastas vyr. g. daiktavardis
2	Vyr. g. dkt., bet turi mot. g. atitikmenį, pvz., <i>informatikas – informatikė</i>
3	Daiktavardis nekaitomas; tradiciškai laikoma, kad tai vyr. g., nors tai ne visais atvejais akivaizdu, pvz., <i>kredo, taksi</i>
4	Dažniausiai laikoma, kad daiktavardis yra vyr. g., bet jis gali būti tiek vyr., tiek ir mot. giminės, pvz., kai kurie nekaitomi dkt.: <i>buržua, atašė, portjė</i> ar kaitomi: <i>junga, buožė</i>
5	Įprastas mot. g. daiktavardis
6	Įprastas bendrosios g. daiktavardis, pvz., <i>valkata, vėpla, mēmė</i>
7	Daiktavardis nekaitomas; gali būti laikoma, kad tai mot. g., nors tai ne visais atvejais akivaizdu, pvz.: <i>bežė</i>
8	Mot. g. daiktavardis; turi vyr. g. atitikmenį, pvz., <i>aulininkė – aulininkas</i>
9	Daiktavardis nekaitomas; apie jo giminę ką nors pasakyti sunku (dažniausiai tai užsienio kalbų žodžiai)

¹⁴ Dėl patogumo daiktavardžiai, turintys ir vienaskaitą, ir daugiskaitą, morfologinės analizės programoje vadinami *abiskaitiniais*.

4 lentelė. Daiktavardžių pobūdis

Nr.	Pobūdžio aprašymas
1	Paprastas daiktavardis
2	Etnonimas (tautietis)
3	Nekaitomas kitų kalbų bendrinis, pvz., <i>kredo</i>
4	Teksto vienetas – užsienio k. žodis (dažniausiai posakiuose, pvz., lotyniškuose, ir pan.), kuriam lietuvių k. gramatinės kategorijos (giminė, linksnis, kalbos d. ir pan.) neturi prasmės arba turi ją sąlyginai (priskyrimas daiktavardžiams dažniausiai irgi yra sąlyginis), pvz.: <i>allegro, coupe, entree</i>
5	Vyriškos giminės daiktavardis, linksniuojamas kaip mot. g., pvz., <i>dėdė, viršila</i>
6	Sudaiktavardėjęs vyr. g. būdvardis, laikytinas daiktavardžiu, pvz., <i>jaunasis</i>
7	Sudaiktavardėjęs vyr. g. dalyvis, laikytinas daiktavardžiu, pvz., <i>išrinktasis, palaimintasis, pažįstamas, laukiamasis, mylimasis, kaltinamasis, nežinomas, dirbantysis, suaugęs</i>
8	Sudaiktavardėjęs mot. g. būdvardis, laikytinas daiktavardžiu, pvz., <i>jaunoji, gulsčioji, baltosios, lygiosios</i>
9	Sudaiktavardėjęs mot. g. dalyvis, laikytinas daiktavardžiu, pvz., <i>palaimintoji, pažįstama, kertamoji, kuliamoji, mylimoji, dirbančioji, sprogtinoji</i>
10	Bendrosios giminės daiktavardis, kaitomas kaip vyr. g.
11	Bendrosios giminės daiktavardis, kaitomas kaip mot. g., pvz., <i>valkata</i>

5 lentelė. Daiktavardžių neiginiai

Nr.	Neigiamumo aprašymas
1	Daiktavardis žodynuose be neigiančio prefikso, neigiančių prefiksų turėti negali
2	Turi neigiantį prefiksą <i>a-</i> ; be šito prefikso jis nevartojamas, pvz.: <i>alogizmas</i>
3	Turi neigiantį prefiksą <i>an-</i> ; be šito prefikso jis nevartojamas, pvz.: <i>analfabetizmas</i>
4	Turi neigiantį prefiksą <i>a-</i> ; yra atitinkamas dkt. ir be šito prefikso, pvz.: <i>aneurinas</i>
5	Turi neigiantį prefiksą <i>be-</i> ; yra atitinkamas dkt. ir be šito prefikso, pvz.: <i>besotumas</i>
6	Turi neigiantį prefiksą <i>ne-</i> ; yra atitinkamas dkt. ir be šito prefikso, pvz.: <i>neakivaizdininkas</i>
7	Turi neigiantį prefiksą <i>ne-</i> ; be šito prefikso jis nevartojamas, pvz.: <i>nerimas</i>
8	Turi neigiantį prefiksą <i>be-</i> ; be šito prefikso jis nevartojamas, pvz.: <i>bedalis</i>
9	Daiktavardis žodynuose be neigiančio prefikso; gali turėti neigiančius prefiksus <i>ne-</i> , <i>nebe-</i> ir pan.; pvz.: <i>abejingumas</i> – <i>neabejingumas</i>

6 lentelė. Daiktavardžių deminutyviniai vediniai

Nr.	Deminutyvinių vedinių aprašymas
1	Deminutyvų neturi
2	Turi įprastus deminutyvinius vedinius, bet tik kurios nors vienos giminės (pvz., <i>stalas</i> – <i>stalelis</i>)
3	Turi abiejų giminių deminutyvus, pvz., <i>meška</i> – <i>meškutis</i> – <i>meškutė</i>
4	Daiktavardžiai nekaitomi, nors kai kurie jų, ypač dažniau vartojami ir ypač šnek. kalboje, gali būti linksniuojami; tokie dkt. kartais gali įgyti mažiškos formos, pvz., <i>atašė, buržua</i>
5	Kaip ir 2 atveju (turi įprastus deminutyvus; bet tik kurios nors vienos giminės); bet žodynuose jau pateikiama mažiškos forma, pvz., <i>balnelis</i> (dviračio) (dažniausiai tai terminai)

7 lentelė. Daiktavardžių priesaginiai vediniai

Derivacijos būdai
Daiktavardžio vedinių su <i>-ininkas (-ė)</i> galimumas
Daiktavardžio vedinių su <i>-ybė</i> galimumas
Daiktavardžio vedinių su <i>-ystė</i> galimumas
Būdvardžio vedinių su <i>-ingas</i> galimumas
Būdvardžio vedinių su <i>-iškas</i> galimumas
Būdvardžio vedinių su <i>-inis</i> galimumas
Prieveiksmio vedinių su <i>-in(i)ai</i> galimumas
Vedinių su <i>-ėtas</i> galimumas
Būdvardžio vedinių su <i>-(u)otas</i> galimumas

DOI: 10.5755/j01.sal.1.15.43348