



faculty of social
sciences, arts
and humanities

SAL 39/2021

Research Journal
Studies about Languages

pp. 71-84

ISSN 1648-2824 (print)

ISSN 2029-7203 (online)

DOI 10.5755/j01.sal.1.39.29193

LINGUISTICS / KALBOTYRA

Automatizuotas arbitraliųjų kolokacijų atpažinimas:
būdvardžių ir daiktavardžių kolokacijos

Received 05/2021

Accepted 09/2021



<http://dx.doi.org/10.5755/j01.sal.1.39.29193>

HOW TO CITE: Kovalevskaitė, J., Rimkutė, E., & Vaičenonienė, J. (2021). Automatizuotas arbitraliųjų kolokacijų atpažinimas: būdvardžių ir daiktavardžių kolokacijos. *Studies about Languages / Kalbų studijos*, 39, 71–84. <http://doi.org/10.5755/j01.sal.1.39.28665>

Automatizuotas arbitraliųjų kolokacijų atpažinimas: būdvardžių ir daiktavardžių kolokacijos

Automatic Recognition of Arbitrary Adjective-Noun Collocations

JOLANTA KOVALEVSKAITĖ, Vytauto Didžiojo universitetas, Lietuva

ERIKA RIMKUTĖ, Vytauto Didžiojo universitetas, Lietuva

JURGITA VAIČENONIENĖ, Vytauto Didžiojo universitetas, Lietuva

Santrauka

Straipsnyje analizuojamas vienas kolokacijų tipas – arbitraliosios kolokacijos. Jos laikomos nemotyvuotomis (tuo skiriasi nuo kito kolokacijų tipo – trivialiųjų, arba motyvuotųjų, kolokacijų, pvz.: *graži diena, naujas tyrimas*), joms būdingas ribotas leksinis junglumas: pagal panašią reikšmę teoriškai galima rinktis daugiau negu vieną žodį, bet linkstama vartoti būtent kurį nors vieną, pvz., *platus akiratis*, bet ne *didelis akiratis*, *stipri sveikata*, bet ne *galinga sveikata*.

Analizuotos iš „Lietuvių kalbos pastoviųjų žodžių junginių duomenų bazės“ atrinktos visos kolokacijos, sudarytos iš būdvardžių ir daiktavardžių (apie 5000 junginių). Nustatyta, kad iš jų apie 650 yra arbitraliosios kolokacijos. Atrenkant šias kolokacijas, taikytas sinonimų pakeičiamumo testas: jeigu kolokacijos būdvardiškasis dėmuo (būdvardis ir būdvardiškai vartojamas dalyvis) nėra pakeičiamas kitu artimos reikšmės sinonimu, tokia kolokacija laikoma arbitraliaja. Šis arbitraliųjų kolokacijų atrankos procesas buvo automatizuotas: automatiškai atrenkant artimus būdvardžių sinonimus junginiuose su daiktavardžiais, pritaikytas žodžių vektorių modelis. Automatiškai būdvardžiai ir dalyviai sugrupuoti į maždaug 800 vektorių eilučių. Straipsnyje aprašyta, kaip vyko duomenų analizė, vertinimas ir tikslinimas, kokios kolokacijos ir kiek jų pripažintos arbitraliosiomis naudojant šį metodą.

RAKTINIAI ŽODŽIAI: arbitraliosios kolokacijos, DELFI.lt tekstinis, kompiuterinė lingvistika, vektorių metodas, ribotas leksinis junglumas, būdvardžiai, daiktavardžiai.

Įvadas

Pastaraisiais metais pastoviesiems žodžių junginiams atpažinti pritaikoma vis naujų metodų (plg. Mandravickaitė et al., 2016). Nemažai darbų parengta apie kolokacijas, jų ypatybes, vartoseną (kaip vienas iš svarbesnių tyrimų – Marcinkevičienė, 2010), kolokacijų ir frazeologizmų skirtumus (plg. Bielinskienė et al., 2017). Vis dėlto pasigendama darbų apie skirtingų tipų kolokacijas, jų požymius. Šie leksikos vienetai įvairiai apibrėžiami, tyrėjai pateikia gana skirtingus kolokacijų požymius.

Šiame straipsnyje kolokacijas skirstome į trivialiąsias (motyvuotas), pvz., *graži aplinka, linksmas žmogus*, kurios sudaromos be didelių leksinių apribojimų, ir arbitraliąsias (nemotyvuotas), pvz., *pilkas gyvenimas, nepatogus klausimas*, kurioms būdingas ribotas leksinis junglumas. Ir nors aiškios ribos tarp to, kas motyvuota, o kas ne, brėžti ne visada įmanoma, vis dėlto manome, kad kai kurių kolokacijų motyvuotumą be gilesnių tyrimų paaiškinti sudėtinga ir jos gali būti laikomos daugiau ar mažiau arbitraliomis leksinių ar semantinių ribojimų kontinuumu.

Šio straipsnio tyrimo objektas – arbitraliosios kolokacijos (toliau AK), iki šiol lietuvių darbuose sulaukusios gana nedaug dėmesio. Šios kolokacijos analizuojamos vykdant projektą „Arbitraliosios lietuvių kalbos kolokacijos: atpažinimas, aprašymas, vartojimas“ (ARKA) (projektą finansuoja Lietuvos mokslo taryba (Nr. S-LIP-20-18)). Projekto tikslas – aprašyti lietuvių kalbos AK ir parengti joms skirtus leksikografinius bei mokomuosius resursus.

Straipsnio tikslas – pristatyti arbitralumą kaip kolokacijų požymį ir paaiškinti, kaip atpažinti AK taikant sinonimų pakeičiamumo testą. Sinonimams nustatyti naudotas vektorių metodas (angl. *Word Embedding Approach*). Tyrimo šaltiniai: 1) „Lietuvių kalbos pastoviųjų žodžių junginių duomenų bazė“¹, joje pateikiama apie 12 tūkst. kolokacijų, tyrime panaudotos visos kolokacijos, sudarytos iš būdvardžių ir daiktavardžių (apie 5000); 2) DELFI.lt tekstynas², sudarytas iš 70 mln. žodžių; jame analizuoti susiję žodžiai, t. y. sudaryti žodžių vektoriai.

Antrame skyriuje apžvelgta, kaip skirtingi tyrėjai supranta arbitralumo terminą ir AK būdingas ypatybes, kokie atlikti AK tyrimai ir kaip jie pritaikomi šiame tyrime. Trečiame skyriuje pristatomas žodžių vektorių metodas, jo pritaikymas tyrimo tikslams, nuosekliai aprašoma AK atpažinimo eiga ir skirtingi vektorių metodo taikymo etapai, aptariami rezultatai ir tolesnių tyrimų gairės.

Arbitralumo samprata kolokacijų tyrimuose

Kadangi tyrimo objektui įvardyti pasirinkome AK terminą, svarbu aptarti arbitralumo sampratą ir jos taikymą kolokacijoms apibūdinti ir tirti. Pastebima, kad nors arbitralumas kaip kolokacijų požymis dažnai minimas tyrėjų darbuose, pats terminas plačiau nekomentuojamas kaip savaime suprantamas (Martines, 2017, p. 121). Chrestomatiniuose kolokacijų darbuose pateiktus apibrėžimus automatiškai perima kiti tyrėjai. Dažnai kolokacijų apibrėžimuose arbitralumas tiesiog įvardijamas kaip kolokacijų ypatybė:³

- kolokacija – tai arbitralus ir pasikartojantis žodžių junginys (Benson, 1990);
- kolokacijos yra arbitralios struktūros, t. y. tradiciniai / įprasti derinių variantai, kuriuos vartoja gimtakalbiai (Henriksen, 2013, p. 32);
- kolokacijos žymi leksikos derinius, kurie yra: a) gatavi, b) arbitralūs, c) nenuspėjami, d) pasikartojantys ir e) neriboto ilgio (Seretan, 2011, p. 26).

D. Liu (2010, p. 7) pabrėžia, kad kaip ir varijuojantys kolokacijų apibrėžimai, arbitralumas taip pat gali būti suprantamas ir apibrėžiamas keleriopai, priklausomai nuo tyrimo srities ir tikslų:

1) pirmu, tarpkalbiniu, požiūriu, arbitralumas reiškia tarpkalbinius kolokacijų raiškos skirtumus: skirtingose kalbose kolokacijos gali nesutapti (pvz., angl. *to have a dream* – liet. *sapnuoti sapną*);

2) antru, intrakalbiniu, požiūriu, arbitralumas konkrečioje kalboje suprantamas kaip nenuspėjami / nemotyvuoti sintaksiniai ir semantiniai žodžių ryšiai (pvz., *stipri kava*, bet ne *galinga kava*).

D. Liu skiria (2010, p. 8–9) dvi tyrėjų grupes: manančius, kad visoms kolokacijoms būdingas vienoks ar kitoks arbitralumas, ir teigiančius, kad ne visos kolokacijos yra arbitralios. Pavyzdžiui, antrosios kalbos tyrimų kontekste ši požiūrių skirtis svarbi, nes lemia kolokacijų mokymo metodiką – remiantis pirmuoju, tradiciniu, požiūriu, kolokacijos dažniausiai mokomos kaip gatavi fraziniai vienetai nesigilinant į sudedamąsias dalis. Plačiai diskutuojama apie tai, kad besimokantiesiems kalbų kolokacijos kelia daug iššūkių, nes jų dažnai negalima išversti pažodžiui iš savo gimtosios kalbos, t. y. jos yra arbitralios. Antrojo požiūrio šalininkai teigia, kad intrakalbiniu požiūriu arbitraliomis laikytos kolokacijos ne visada yra nemotyvuotos (Liu, 2010, p. 8–9), ir tai pagrindžia kognityviniai ir tekstynais remiami tyrimai. Pavyzdžiui, I. Bosque (2011, p. viii) teigia, kad kolokacijos gali būti arbitralios, tik jeigu

¹ Žr. <http://resursai.pastovu.vdu.lt/paieska/paprastoji>

² Žr. <https://klc.vdu.lt/pastovuSearch.html>

³ Vertimas iš anglų kalbos – straipsnio autorių.

terminas vartojamas F. de Saussure'o suteikta reikšme, t. y. iš tarpkalbinių skirtumų pozicijų (taip pat žr. Williams, 2001, p. 4).

Galima skirti ir trečiąjį požiūrį, kai į arbitralumą vienu metu žiūrima iš tarp- ir intrakalbinių pozicijų. Liu (2010, p. 8) teigia, kad daug antrosios kalbos įsisavinimo tyrėjų arbitralumą suvokia abiem reikšmėmis: kolokacijos semantiškai nemotyvuotos, jų raiška skiriasi skirtingose kalbose, todėl jas sunku išmokyti / mokyti. Pavyzdžiui, B. Henriksen (2013) teigia, kad kolokacijų kompetencija ypač svarbi mokantis užsienio kalbos, ji suteikia galimybę besimokančiajam laisvai vartoti frazeologinę raišką, kalbėti realiu laiku, atpažinti žodžių daugiareikšmiškumą, konotacijas ir kt. Tokio pobūdžio kolokacijų tyrimams pasitelkiami įvairūs metodai ir šaltiniai, pvz., rašytinės kalbos duomenys, esė, tekstynai, žodinės kalbos tyrimai, vertimo užduotys, tinkamo atsakymo iš kelių pasirinkimas, praleistų žodžių įrašymo testai, akių judesių matavimas ir kt. Apžvelgdama tyrimus, B. Henriksen (2013, p. 33) akcentuoja, kad būtent AK yra problemiškos besimokantiems užsienio kalbos (t. y. tarpkalbinis arbitralumo požiūris), jos įsisavinamos kitaip negu motyvuotosios, skiriasi jų mokymo(si) metodai.

Taip pat semantiškai kolokacijos skiriamos į motyvuotas ir arbitralias – pastarųjų negalima paaiškinti remiantis propozicine žodžio reikšme (t. y. intrakalbinis požiūris į arbitralumą). Nustatyta, kad tiek negimtakalbių, tiek besimokančiųjų svetimiosios kalbos kolokacijų įvairovė mažesnė, daroma klaidų (Henriksen, 2013, p. 33). Kolokacijų arbitralumą siekiama paaiškinti mokant užsienio kalbos, plg. Lewis (2001, p. 26) teigia, kad kolokacijos arbitralumas atsiskleidžia šiais pavyzdžiais: angl. *high / tall building* (liet. *aukštas pastatas*), angl. *tall boy* (liet. *aukštas berniukas*), bet ne angl. **high boy*.

Savo tyrime D. Liu (2010, p. 10) svarsto, ar įvairiuose žodynuose ir vadovėliuose įtrauktos semantiškai arbitralios (nemotyvuotomis) laikomos kolokacijos iš tiesų yra tokios, jeigu jos tiriamos tekstynų lingvistikos ir kognityviosios analizės metodais, ir kaip tai gali paveikti kolokacijų mokymą. Intralingvistinės semantinės / kognityviosios analizės metodas, D. Liu (2010, p. 11) nuomone, gali padėti besimokantiems svetimiosios kalbos sąmoningai suprasti, kodėl kolokaciją sudarantys dėmenys vartojami kartu, ir taip lengviau juos įsiminti. Pavyzdžiui, kolokacijų arbitralumui tirti D. Liu (2010) kaip vieną iš grupių pasirinko desenantizuotus veiksmažodžius. Dauguma daiktavardžių *ability* (liet. *gebėjimas*), *work* (liet. *darbas*) ir *trip* (liet. *kelionė*) veiksmažodinių kolokatų yra plataus junglumo, desenantizuoti (pvz., angl. *make, take, have a trip*), dažnai vartojami ir kitose kolokacijose, vadinasi, jie tinkami tiriant arbitralumą. Manoma, kad daiktavardžių junginiai su desenantizuotais veiksmažodžiais yra semantiškai nemotyvuoti.

Išrinkus kiekvieno daiktavardžio veiksmažodinius kolokatus, paaiškėjo, kad, išskyrus kelis desenantizuotus veiksmažodžius, visi jie gali būti suskirstyti į aiškias semantines kategorijas, taigi, nėra arbitralūs. Pavyzdžiui, daiktavardis *ability* gali būti sinonimiškai vartojamas su tokiais veiksmažodžiais, kaip *demonstrate, display, show* ar *exhibit*, o tai rodo, kad šios kolokacijos yra motyvuotos. Desenantizuoti veiksmažodžiai toliau tirti kognityviuoju požiūriu pagrindžiant jų motyvuotumą – lingvistinė raiška perteikia reiškinio konceptualizavimą (Liu, 2010, p. 20). Pavyzdžiui, iš pirmo žvilgsnio tokie nemotyvuoti frazinių veiksmažodžių junginiai, kaip angl. *go up* (liet. *kilti, padidinti, išaugti*) ar *cut down* (liet. *sumažinti*) gali būti interpretuojami per konceptualiųjų metaforų raišką – *up* semantiškai persidengia su *daugiau*, o *down* su *mažiau* reikšmėmis (čia D. Liu (2010) minimas Kovesces ir Szabo (1996) tyrimas). Apibendrinamas savo tyrimą, D. Liu (2010) siūlo pergaltoti kolokacijų mokymo metodiką, įtraukti ir semantinę / kognityvinę analizę grįstus pratimus.

Kolokacijų arbitralumo neapibrėžtumą aptaria ir N. Nesselhauf (2005). Tyrėjos nuomone, sunku brėžti aiškią ribą tarp arbitralumo ir semantinio motyvuotumo jau vien dėl to, kad žodžių reikšmės taip pat susijusios. Teoriškai gal ir būtų įmanoma nustatyti visas žodžio reikšmes ir pagrįsti kiekvieno pavartojimo semantinę motyvaciją. Pavyzdžiui, jeigu manytume, kad veiksmažodžio *shake* (liet. *kratyti*) reikšmė konkrečiame junginyje *shake one's head* yra „sukti galvą nuo vieno peties prie kito parodant nepritimą“, ši kolokacija būtų laikoma semantiškai motyvuota (Nesselhauf, 2005, p. 31). Kita vertus, kaip teigia tyrėja, tam tikras arbitralumo laipsnis būdingas beveik kiekvienam žodžių junginiui, todėl tam tikro veiksmažodžio junglumo apribojimai gali būti aiškinami kaip daugiau semantiškai motyvuoti ir minimaliai arbitralūs arba daugiau arbitralūs negu semantiškai motyvuoti (Nesselhauf, 2005, p. 31).

Kaip matome, skirtingos tyrėjų grupės ir mokyklos įvairiai suvokia ir tiria arbitralumą: kaip reiškinį, būdingą visoms kolokacijoms, kaip ginčytiną kolokacijų savybę, kuri neturėtų būti laikoma universalioja, kaip galimą tarpkalbinę

kolokacijų raiškos skirtį, kaip semantinį ir sintaksinį nemotyvuotumą ar kaip griežtai nenustatomą savybę, kuri, atrodo, daugiau ar mažiau akivaizdi iš pirmo žvilgsnio, bet kelia daug klausimų tiriant kolokacijas gilumine analize. Apibendrinant *arbitralumo* sąvokos kolokacijų tyrimuose apžvalgą, galima skirti šiuos AK požymius: dažna, įprasta vartoseną; paviršinis semantinis nemotyvuotumas, kuris gali (arba negali) būti paaiškinamas taikant giluminę semantinę, kognityvinę analizę, t. y. intuityviai ryšys nenuspėjamas, reikalinga gilesnė analizė; galima skirtinga tarpkultūrinė raiška (nors ne visada, priklauso nuo kalbos ir kt.); semantinis daugiareikšmiškumas, kai tas pats žodis skirtingomis reikšmėmis vartojamas su skirtingais kolokatais – tokiu atveju viena reikšme vartoseną gali būti arbitrali, o kita motyvuota ir ši skirtis matoma tik kontekste. Semantiškai daug AK sudaroma su abstrakčiaisiais daiktavardžiais (Marcinkevičienė, 2010), plataus junglumo veiksmožodžiais (pvz., Liu, 2010; Nesselhauf, 2005).

AK atpažinimo eksperimentas taikant vektorių metodą

Šiame tyrime atliktas eksperimentas, kaip gana didelės apimties medžiagoje, pasirinkus tam tikros struktūros (būdvardis + daiktavardis) kolokacijas, atpažinti AK. Arbitralumą suvokiame taip, kad, esant pasirinkimui panašią reikšmę perteikti keliais žodžiais, konkrečioje kolokacijoje vartojamas tik vienas kuris iš tų panašios reikšmės žodžių. Tokiu atveju arbitralumas yra semantiškai nemotyvuotas leksinis pasirinkimas, rodantis konvenciją, tam tikrą kalbėtojų polinkį. Taigi arbitralumas susijęs su ribotu leksiniu junglumu. Tikrosios kolokacijos yra arbitralios, t. y. kuris nors vienas iš kolokacijos dėmenų turi būti riboto leksinio junglumo (angl. *restricted*). Pagal N. Nesselhauf (2003, p. 226), kuri remiasi F. Hausmannu (1998) ir I. Melčuku (1998), riboto junglumo yra tas kolokacijos žodis, kuris priklausomas nuo kolokacijos pagrindinio žodžio, t. y. pagrindinis žodis parenkamas nepriklausomai, pagal savo reikšmę ir lemia, koks bus parinktas kitas žodis. Parenkant tam tikros reikšmės priklausomąjį žodį, atrodytų, semantiškai galima būtų rinktis iš kelių, ypač kai yra sinonimų, tačiau dėl riboto leksinio junglumo konkrečioje kolokacijoje įprastai pasirenkamas tik vienas kuris nors sinonimas.

Riboto leksinio junglumo nustatymas

Bielinskienė et al. (2017, p. 90), nurodo, kad ribotas leksinis junglumas gali būti kolokacijų vartojimo pastovumo kriterijus, o nustatyti ribotą leksinį junglumą galima taikant pakeičiamumo testą. N. Nesselhauf (2003; 2005) tyrimuose, kuriais remiamės, analizuojamas tik vienas veiksmožodžio ir daiktavardžio struktūrinis kolokacijų tipas, tačiau, kaip teigia pati tyrėja, jos siūlomas kolokacijų apibrėžimas ir juo grįsta žodžių junginių klasifikacija gali būti tinkama ir kitiems gramatiniams kolokacijų tipams (Nesselhauf, 2003, p. 225). Mūsų tyrime ji pritaikoma būdvardžio ir daiktavardžio junginiams.

N. Nesselhauf (2003) teigia, kad kolokacijai būdingas arbitralus pakeičiamumo ribojimas (angl. *arbitrary restriction on substitutability*): laisvuosiuose žodžių junginiuose žodžiai parenkami pagal semantinį suderinamumą, pavyzdžiui, sakoma *skaityti laikraštį*, bet ne *gerti laikraštį*. O kolokacijose pasirinkimai apriboti leksiškai, tarkim, *surinkti* (angl. *dial*) yra riboto junglumo, nes gali būti vartojamas tik su kai kuriais daiktavardžiais: *surinkti galima numerį*, bet ne *radijo stotį*. Skaitant įvairių autorių darbus, matyti, kad juose vietoj *arbitralumo* sąvokos sinonimiškai vartojami tokie kolokacijų apibūdinimai, kaip angl. (*semantically*) *unmotivated* (liet. nemotyvuotos), *semantically opaque* (liet. semantiškai neaiškios), *non-motivated* (liet. nemotyvuotos), (*semantically/cross-culturally*) *unpredictable* (liet. nenuspėjamos), *non-deducible* (liet. neišvedamos) (plačiau žr. Liu, 2010; Bosque, 2011; Henriksen, 2013). Tai rodo tikrųjų kolokacijų sąsają su leksinio, o ne semantinio pobūdžio apribojimais.

Tirdama iš veiksmožodžių ir daiktavardžių sudarytas kolokacijas, N. Nesselhauf aprašo, kad veiksmožodis (daiktavardis) parinkta reikšme negali būti vartojamas su visais sintaksiškai ar semantiškai galimais daiktavardžiais (veiksmožodžiais). Šią reikšmę tyrėja įvardija kaip ribotą reikšmę (angl. *restricted sense*). Žodžio reikšmė laikoma ribota, jeigu atitinka bent vieną iš dviejų kriterijų (Nesselhauf, 2003, p. 225):

1 kriterijus: veiksmožodžio (daiktavardžio) reikšmė tokia specifinė, kad junglumas įmanomas tik su nedideliu daiktavardžių (veiksmožodžių) skaičiumi (tyrime apsiribota 5 žodžių riba);

2 kriterijus: veiksmožodis (daiktavardis) parinkta reikšme negali būti vartojamas su visais sintaksiškai ar semantiškai galimais daiktavardžiais (veiksmožodžiais).

Pavyzdžiui, žodis *norėti* pirmine reikšme neatitinka nė vieno kriterijaus, nes gali jungtis su daugybe daiktavar-

džių (*norėti žaislų, vaiko, gerti, mašinos* ir kt.) ir neturi jokių junglumo ribojimų, nebent semantinius. Tačiau anglų kalbos veiksmažodis *dial* (*surinkti*) yra riboto junglumo, nes gali būti vartojamas tik su kai kuriais daiktavardžiais (*number* ir gal dar keliais) (Nesselhauf, 2003, p. 226). Savo tyrimo pagrindu N. Nesselhauf (2005, p. 29–30) anglų kalboje išskiria penkias veiksmažodžio ir daiktavardžio dermes: 1) veiksmažodis dera su beveik visais daiktavardžiais; 2) veiksmažodis dera su didele daiktavardžių grupe; 3) veiksmažodis dera su nedidele, bet aiškiai semantiškai apibrėžiama daiktavardžių grupe; 4) veiksmažodis dera su nemaža daiktavardžių grupe, bet yra išimčių; 5) veiksmažodis dera tik su keletu daiktavardžių (plačiau žr. Nesselhauf, 2005, p. 29–30).

Taikydama savo metodologiją praktiškai, N. Nesselhauf (2003, p. 228) pabrėžia, kad nustatyti, ar junginyje su daiktavardžiu veiksmažodžio reikšmė yra ribota ar ne, buvo ypač sudėtinga. Pirmiausia, tikrinti žodžių aprašymai dviejuose žodynuose: jeigu abiejuose žodynuose veiksmažodžio apribojimai neįvardyti, veiksmažodis buvo priskiriamas laisviesiems žodžių junginiams. Jeigu žodyno apibrėžime nurodyti konkretūs daiktavardžiai, su kuriais vartojamas veiksmažodis, žodžių junginys laikytas kolokacija. Nustatytos kolokacijos suskirstytos į kelis pogrupius pagal tai, kokį kriterijų atitinka: 1) kolokacijos, kurios, nors jų junglumas ir ribotas, turi daugiau laisvės sudaryti įvairesnius junginius; 2) kolokacijos, pasižyminčios stipriais apribojimais.

N. Nesselhauf (2003) atliko tyrimą su nedideliu kolokacijų skaičiumi. Mūsų tyrime apdorojamos medžiagos kiekis yra daug didesnis, todėl ribotam junglumui patikrinti pakeičiamumo testą taikėme kiek kitaip: panaudodamos tam tikro būdvardžio sinonimus, tikrinome, ar tas pats daiktavardis yra užfiksuotas kolokacijoje su daugiau negu vienu sinonimu iš tos pačios sinonimų grupės. Laikėmės požiūrio, kad jeigu būdvardžio reikšmė yra ribota, jis bus vartojamas ne su visais galimais daiktavardžiais.

Artimų sinonimų tyrimai rodo, kad tokiems žodžiams, nors jie ir yra panašių (artimų) reikšmių, bet būdinga skirtinga vartosena, t. y. skirtingos kolokacijos (Xiao & McEnery, 2006, p. 108). Autoriai teigia, kad dažniausiai artimi sinonimai turi skirtingą semantinę prozodiją, taigi dėl to skiriasi ir jų kolokatai; įvairiose kalbose tokie artimi sinonimai negali būti pakeičiami be tam tikrų prasmės poslinkių (Xiao & McEnery, 2006, p. 124–125). Tai yra dėl leksinio junglumo skirtumų, tos pačios grupės sinonimai netinka vieni kitų kolokacijose, pvz., Xiao & McEnery (2006) mini Halliday (1976, p. 73) pavyzdžius: arbata (angl. *tea*) įprastai būna stipri (angl. *strong*), bet ne galinga (angl. *powerful*), o automobilis (angl. *car*) dažniau apibūdinamas kaip galingas, o ne stiprus, nors abu būdvardžiai ir turi panašias reikšmes.

Taigi mūsų pakeičiamumo kriterijaus taikymo schema rėmėsi tokia logika: 1) jeigu įrodome, kad tam tikros kolokacijos būdvardis turi sinonimų, iš to seka, kad yra pasirinkimo reikalaujanti situacija, vadinasi, prielaida AK; 2) remdamiesi savo turima kolokacijų duomenų baze, automatiškai patikriname, ar analizuojamas daiktavardis jungiasi tik su vienu tos sinonimų grupės būdvardžiu. Tokiu būdu mūsų eksperimentas paremtas artimų sinonimų nepakeičiamumu, kuris ir įrodo, kad skirtingos kolokacijos su (artimais) sinonimais yra laikytinos AK. Pavyzdžiui, turime sinonimų eilę *deramas, tinkamas, teisingas, pageidaujamas*. Įvade minėtoje „Lietuvių kalbos pastovijų žodžių junginių duomenų bazėje“ turime kolokaciją *tinkamas sprendimas*, vadinasi, galimybė vartoti *deramas sprendimas* nerealizuota, nors teoriškai galima (gal realizuojama kaip laisvasis junginys, o ne kolokacija), *deramas* randamas kitoje kolokacijoje – *deramas elgesys*. Vadinasi, turime dvi AK – *tinkamas sprendimas* ir *deramas elgesys*.

Tačiau siekdamas atlikti tokį eksperimentą, turėjome išspręsti problemą, kaip gauti duomenų apie (artimus) sinonimus. Nutarta tam panaudoti vektorių metodą (apie jį rašoma poskyryje „Vektorių metodo taikymas (artimiems) sinonimams nustatyti“).

Vektorių metodo taikymas (artimiems) sinonimams nustatyti

Vektorių metodo pristatymas

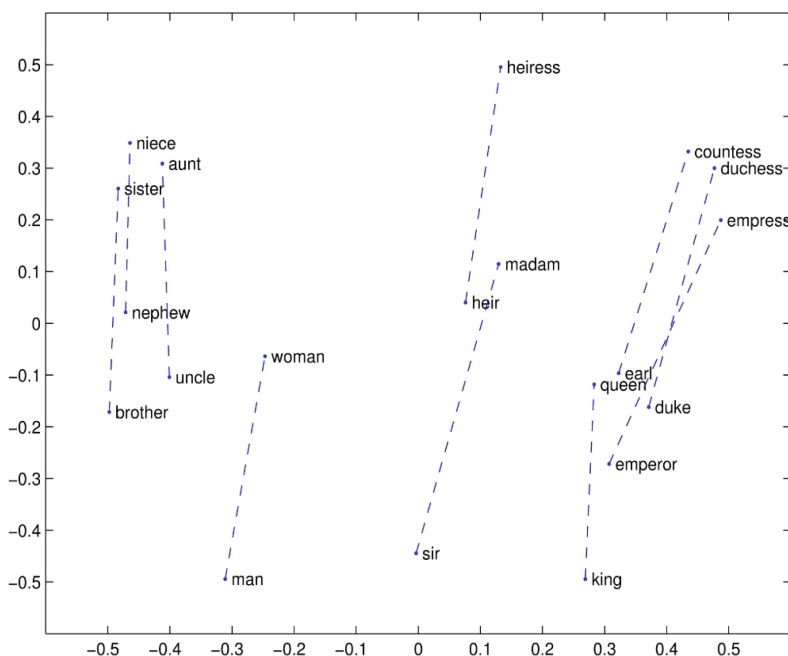
Vektorių metodas (angl. *Word Embedding Approach*) priklauso grupei metodų, skirtų automatiškai analizuoti žodžius pagal jų distribucijos, arba konteksto, panašumus (angl. *distributional similarity measures*). Įvairiems ryšiams (hiperonimijai, sinonimijai, antonimijai, partonimijai) automatiškai atpažinti tinka skirtingi distribucijos panašumus analizuojantys metodai (plačiau žr. Grigonytė, 2010, p. 77). Tokie metodai paremti hipoteze, kad artimos reikšmės ir (ar) semantiškai susiję žodžiai yra vartojami panašiam kontekste (Grigonytė, 2010, p. 70).

Žodžių vektoriai naudojami įvairiuose kalbos technologijų uždaviniuose, pvz., klasifikavimo: panašių dokumentų radimas, įvardytųjų esybių identifikavimas ir pan. (Gutiérrez & Norambuena, 2019). Kadangi žodžiai yra charak-

terizuojami vektoriais, tai žodžių panašumą gana neblogai iliustruoja geometrinė interpretacija, žodžiai gali būti įsivaizduojami kaip taškai (vektoriniai) n -matėje erdvėje⁴. Kuo tie taškai yra artimesni (galima naudoti klasikinį euklidinį, kosinuso ar kitus atstumo skaičiavimo metodus), tuo jie panašesni. Tokiu atveju pagal kontekstą miestų pavadinimai, merai, krepšininkai sukrinta į panašių žodžių grupes (pvz., žodžiui *Kaunas* panašiausi žodžiai yra *Vilnius*, *Panevėžys*, *Alytus*, *Klaipėda*, *Šiauliai*, *Marijampolė*...; žodžiui *Lietuva* – *Lenkija*, *Estija*, *Europa*, *Latvija*, *Rusija*, *Ukraina*, *valstybė*, *kaimyninė*, *šalis*, *Gruzija*...). Tai reiškia, kad tokie analizuojami žodžiai dažniausiai būna tokia pat kontekste kaip ir kiti žodžiai.

Šiame tyrime naudotas GloVe žodžių vektorių modelis (Pennington et al., 2014), kuris suteikia galimybę suformuoti žodžių vektorius atsižvelgiant į bendrą statistiką tekстыne ir artimų žodžių kontekstą. Vektoriai apmokyti panaudojant *StanfordNLP GitHub* saugykloje pateiktą GloVe vektorių generavimui skirtą programinį kodą⁵, mokymo tekstai – 70 mln. žodžių DELFI.lt tekstynas. Pirmiausia tekstynas sujungtas į vieną failą, skrybos ženklai nepašalinti, visos raidės paverstos mažosiomis. Toliau vektorių mokymas susidarė iš kelių etapų: sudarytas žodynas, jame nuspręsta atmesti žodžius, kurie pasitaikė rečiau nei penkis kartus; tada, panaudojant numatytas komandas, sudarytas žodis-žodis buvimo kartu statistikos failas, pasirinktas lango dydis 5, o tai reiškia, kad analizuojamo žodžio kontekstas formuojant vektorius yra po penkis žodžius iš kairės ir penkis iš dešinės pusės. Šis failas atsitiktinai sumaišytas. Galiausiai sugeneruotas žodžių vektorių failas, generavimui pasirinkta 100 iteracijų, vektorių dimensija – 200 (plačiau žr. Bumbulienė, 2018), t. y. kiekvienas žodis aprašomas 200 skaitinių reikšmių.⁶

Dvimatėje erdvėje atvaizduoti žodžių vektorius galima pritaikius PCA (principinių komponentių analizės) metodą (angl. *principal component analysis*). Šis metodas suteikia galimybę sumažinti duomenų dimensiją išlaikant kuo didesnę duomenų variaciją (Jolliffe, 2002). Vizualizavus žodžius dvimatėje erdvėje, galima suprasti jų tarpusavio ryšius, matoma žodžius siejanti geometrija (žr. 1 ir 2 pav.) – žodžių porų kryptys yra panašios, jei šios poros sudarytos pagal tokią pačią tvarką.

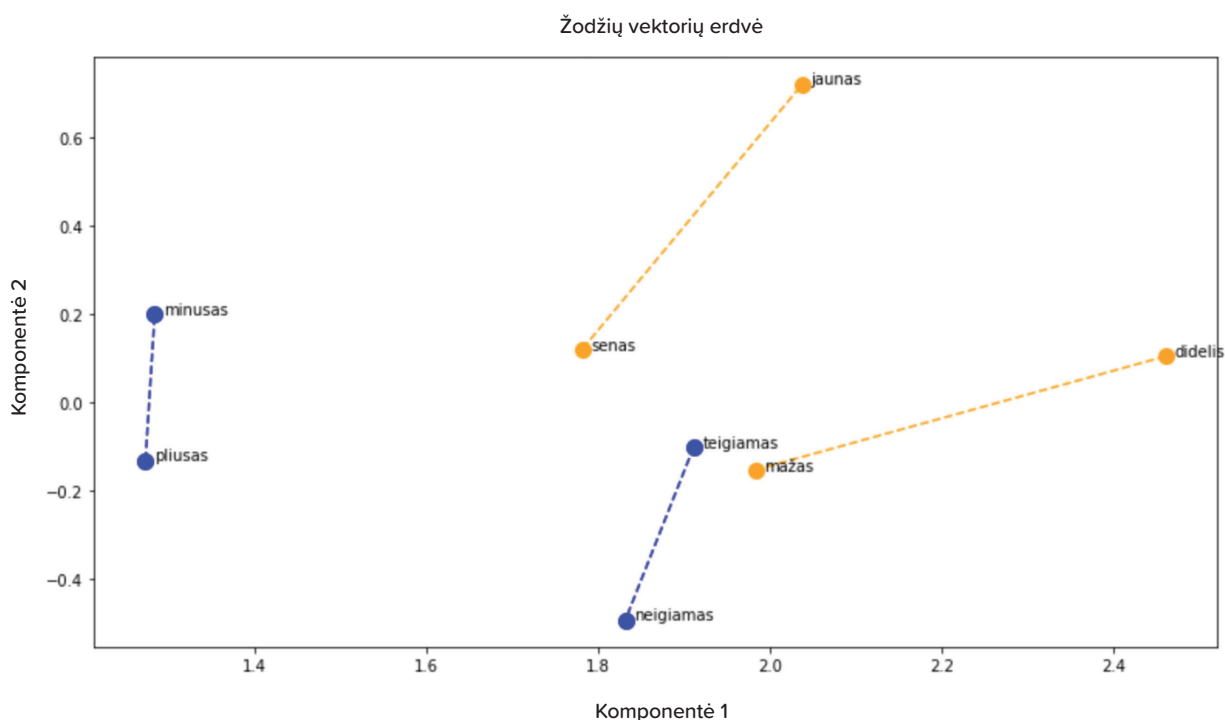


1 pav. Moteriškosios ir vyriškosios lyties asmenis įvardijančių žodžių porų geometrija (šaltinis – <https://github.com/stanfordnlp/GloVe>)

⁴ n – žodžius charakterizuojančio vektoriaus dydis.

⁵ Žr. <https://github.com/stanfordnlp/GloVe>

⁶ Už duomenų parengimą dėkojame kolegei Ievai Bumbulieni.



2 pav. Žodžių, žyminčių amžių ir teigiamumą / neigiamumą žodžių porų geometrija

Skirtingas kontekstas padeda atpažinti perkeltines reikšmes, todėl vektorių metodai taikomi ir perkeltinės reikšmės junginiams atpažinti, pvz., siekiant atpažinti frazeologinius junginius, kurių struktūra veiksmazodis + daiktavardis (Gharbieh et al., 2016). Mūsų eksperimente vektorių metodas naudotas todėl, kad AK pripažinimo eksperimentui reikėjo sinonimų. Kadangi vektorių metodu parengti duomenys jau buvo prieinami kaip vienas iš PASTOVU projekto⁷ rezultatų, nutarta sinonimams nustatyti panaudoti būtent vektorių metodu apdorotus duomenis.

Lietuvių kalbai yra ir kitų šaltinių, kuriuose pateikiami sinonimai: „Sinonimų žodynas“ (2015), lietuviškasis *LitWordNet* tinklas (Vitkutė-Adžgauskienė et al., 2015). Šiame projekte analizuojama dabartinės lietuvių kalbos medžiaga, o medžiaga „Sinonimų žodynui“ rinkta iš senesnių „Lietuvių kalbos žodyno“ ir „Dabartinės lietuvių kalbos žodyno“ leidimų, todėl imti sinonimus iš DELFI.lt tekstinio nutarta dėl duomenų naujumo. *LitWordNet* tinklas jau integruoja ne tik žodynų, bet ir tekstinę medžiagą, jeigu žodis turi kelias reikšmes, jis patenka į kelis sinsetus⁸ (Vitkutė-Adžgauskienė et al., 2015, p. 108). Dėl šios priežasties *LitWordNet-as* iš pradžių buvo svarstytas kaip tinkamas sinonimų šaltinis, tačiau ir jo atsakyta, nes, kaip pastebėta tiriant duomenis, apie kai kuriuos dažnus daugiareikšmius būdvardžius (pvz., *siauras*) *LitWordNet-e* informacijos nepateikiama, ne visų kalbos dalių semantiniai ryšiai baigti aprašyti (plg.: „veiksmazodžių semantiniai ryšiai šioje versijoje dar neišplėtoti“ (Vitkutė-Adžgauskienė et al., 2015, p. 108), be to, būtų reikėję atskirai dirbti ieškant būdų, kaip šioje bazėje esančius duomenis panaudoti mūsų projekto uždaviniams.

Pirmasis eksperimento etapas

Vektorių parengimo procesas vyko taip: pirmiausia iš „Lietuvių kalbos pastoviųjų žodžių junginių duomenų bazėje“ esančių kolokacijų, kurių struktūra būdvardis + daiktavardis, išrinkti visi į tas kolokacijas įeinantys būdvardžiai. Jiems iš DELFI.lt tekstinio atrinkti kiti būdvardžiai ir dalyviai naudojant vektorių metodą. Būdvardžiai ir dalyviai

⁷ žr. <http://pastovu.vdu.lt/>

⁸ Vitkutė-Adžgauskienė et al. straipsnyje (2015, p. 107–108) terminas *sinsetas* paaiškintas taip: „Sinsetas susideda iš vieno arba keleto žodžių, kurie gali vienas kitą pakeisti tame pačiame kontekste, kitaip sakant, sinsetas susideda iš sinonimų.“

į vieną vektorių eilutę sudėti pagal panašų junglumą (arčiau analizuojamo žodžio pateikiami tie, kurių didesnis panašumo įvertis). Vadinasi, duomenys atrinkti ne tik pagal kolokacijas iš minėtos duomenų bazės, bet pagal jų kontekstinę aplinką DELFI.lt tekстыne.

Pateiktos susijusių būdvardžių ir dalyvių lemos (t. y. antraštinės / žodyninės formos), o ne kaitybinės formos, kad būtų lengviau aprėpti duomenis. Gautos 770 vektorių eilutės, kuriose pagal konteksto panašumą sugrupuoti būdvardžiai ir dalyviai. Nuspręsta tam tikram būdvardžiui pateikti ne daugiau kaip po 20 susijusių būdvardžių ir dalyvių, kurių panašumo indeksas ne mažiau kaip 0,3 (maksimalus įvertis yra 1).

Pasirinkta naudoti kosinuso panašumo (angl. *Cosine similarity*) matą, kuris dažnai taikomas panašumo skaičiamams būtent žodžių vektorių erdvėje (Schnabel et al., 2015). Šis panašumo rodiklis parodo, kiek ta pačia ar priešinga kryptimi yra nukreipti vektoriai: kuo vektorių kryptis (kampas tarp jų) erdvėje artimesnė / vienodesnė, tuo panašumo metrikos vertė artimesnė 1, vadinasi, tuo panašesni vektorių eilutėje (toliau VE) nurodyti žodžiai.

Rezultatai. Prie kiekvienos vektorių eilutės buvo nurodytos tiriamos daiktavardis + būdvardis kolokacijos, kurių būdvardžiai, būdvardiškai vartojami dalyviai yra iš tos VE. Pavyzdžiui, prie VE *įdomus; didelis; paskutinis; mažas; pirmas; svarbus; stiprus; paprastas; pastaras; didis; vienintelis; susyti; geras; antras; išskirtinis; gražus; populiarus; užfiksuoti; keistas; įprastas* buvo priskirtos kolokacijos su daiktavardžiu *klausimas* (*panašus klausimas, įdomus klausimas, vienintelis klausimas, paprastas klausimas, keistas klausimas, geras klausimas*), su daiktavardžiu *aplinka* (*panaši aplinka, įprasta aplinka, graži aplinka, gera aplinka*) ir kolokacijos su dar 18 daiktavardžių.

Atkreipiame dėmesį, kad aprašomoje VE yra veiksmazodis *susyti*, iš tiesų tai dalyvis *susijęs*, vartojamas būdvardiškai, todėl pateikiamas prie būdvardžių. Dėl tokių atvejų reikia pasakyti, kadangi duomenys yra automatiškai morfologiškai anotuoti, pasitaiko automatinio anotavimo klaidų, t. y. ne visus VE nurodomus žodžius galima laikyti būdvardžiais arba dalyviais.

Kaip minėta, 770 VE buvo pateikiami pagal panašų kontekstą susieti būdvardžiai ir dalyviai. Peržiūrėjus šiuos duomenis, paaiškėjo, kad į tą pačią VE pateko artimesnių ir labiau nutolusių sinonimų, taip pat pateko ir antonimų, hiperonimų. Vadinasi, VE yra nemažai „triuškmo“⁹. Pavyzdžiui, 14-oje¹⁰ VE *didžiulis; mažas; stiprus; didis; jausti; skirti; rimtas; svarbus; reikalingas; milžiniškas; panašus; brangus; sunkus; realus; geras; gražus; nedidelis; bendras; tikras; išskirtinis* yra aiški sinonimų grupė *didis, didžiulis, milžiniškas*, bet pasitaiko ir antonimų *mažas, nedidelis*, kurie gali sudaryti antrą sinonimų porą. Kiti VE žodžiai taip pat gali būti grupuojami smulkiais grupėmis (pvz., *svarbus, brangus, reikalingas; didis ir stiprus; panašus ir bendras*), tačiau jie jau nėra artimi sinonimai grupei *didis, didžiulis, milžiniškas*. Taip išryškėja, kad vektoriai iš esmės įvertina konteksto panašumą, bet ne konkrečiai atrenka sinonimus, kitaip sakant, VE yra sinonimų, bet ne visi būdvardžiai yra sinonimai. Todėl, vertinant, pavyzdžiui, šiai vektorių eilutei priskirto daiktavardžio *dalis* kolokacijas (plg. *didelė dalis, svarbi dalis, bendra dalis, maža dalis, nedidelė dalis*), pagal pakeičiamumo kriterijų kaip AK galėtume pripažinti kolokaciją *didelė dalis* (nes savo medžiagoje neturime užfiksuotos kolokacijos *milžiniška dalis*, plg. „Lietuvių kalbos pastovijų žodžių junginių duomenų bazėje“ esančią kolokaciją su *milžiniškas – milžiniškas krūvis*).

Jeigu atidžiau pasižiūrėtume į daiktavardžius, kurie sudaro kolokacijas su minėtos VE būdvardžiais, tai tarp jų yra: *dalykas, darbas, galimybė, kaina, karas, klausimas, komanda, miestas, mokestis, pagalba, pajėgos, projektas, reikalas, rinka, turtas, valstybė, įvykis, šalis, ūkis*. Beveik visi įeina į kolokacijas su *didelis* (*didelis klausimas, didelė kaina, didelis miestas, didelė pagalba*), tiesa, kai kuriose kolokacijose šis būdvardis vartojamas netiesiogine reikšme (plg. *didelė pagalba* ir *didelis miestas*). AK nelaikytinos tokios kolokacijos kaip *didelis miestas*, nes tokį junginį nesunku sudaryti semantiškai derinant abiejų žodžių reikšmes, vadinasi, tik tos kolokacijos, kuriose *didelis* vartojamas netiesiogine reikšme (kaip *didelis klausimas, didelė kaina, didelė pagalba*), priskirtinos prie AK, kaip ir anksčiau minėtosios *dalis* kolokacijos *maža dalis, nedidelė dalis*.

Apibendrinus peržiūrėtų pirmųjų 50 VE informaciją, matyti, kad 31 eilutėje yra sinonimų, tačiau dažna situacija, kai vienoje VE yra daugiau negu viena sinonimų pora, plg.: 1-oje VE yra a) *blogas, prastas*; b) *rimtas, tvirtas*; 8-oje VE yra a) *pavojingas, beprotiškas*; b) *istorinis, pirminis*; c) *viešas, oficialus*; 11-oje VE: a) *realus, tikras*; b) *būtinus*,

⁹ Tai pažodinis anglų kalbos termino *noisy data* arba *noise*, reiškiančio perteklinius, neaktualius duomenis, vertimas.

¹⁰ Čia ir toliau nurodytos vektorių duomenų, kuriais remtasi šiame straipsnyje, vektorių eilučių numeriai.

reikalingas; c) *panašus, bendras*. Gali būti ir atvirkščiai, kad du sinonimiški būdvardžiai yra vienoje VE, o dar vienas jiems tinkamas – visai kitoje VE (pvz., 48-oje VE yra *retas*, o ankstesnėje VE yra *išskirtinis* ir *ypatingas*).

16-oje VE yra: *lengvas; sudėtingas; ilgas; nustatyti; nelengvas; didelis; malonus; rimtas; geras; brangus; įtemptas; greitas; reikalingas; intensyvus; padaryti; prastas; gražus; didžiulis; įdomus; paskutinis*. Artimi sinonimai yra *įtemptas* ir *intensyvus*, bet kaip ir jau aprašytoje 14-oje VE, ir šioje eilutėje yra *didelis* ir *didžiulis, malonus, geras, įdomus, gražus*. Pastebėta, kad tokie būdvardžiai, kaip *malonus, geras, įdomus, gražus*, itin dažnai kartojasi skirtingose VE. Tai įrodo, kad iš vektoriams sudaryti naudojamo būdvardžių sąrašo neišėmus tokių bendros reikšmės ir dėl to itin junglių būdvardžių, kaip *geras, naujas, gražus, svarbus*, VE suvienodėja. Be to, kaip jau minėta, toje pačioje VE dažnai randame gana skirtingų būdvardžių (plg. 14 VE esančius: *didžiulis, reikalingas, panašus*), ir dėl to susidaro įspūdis, kad vienoje vektorių eilutėje yra per daug „triukšmo“.

Siekiant geresnių VE rezultatų, nutarta kitaip paruošti eksperimente naudojamus duomenis, kad VE gautume daugiau informacijos apie artimus sinonimus:

- 1 neprasminga į VE sudarymą įtraukti plataus junglumo būdvardžių, ir ne tik dėl to, kad jie pasikartoja beveik visose VE, bet ir dėl to, kad šie būdvardžiai sudarant junginius paprastai tiesiog semantiškai derinami su tuo žodžiu, su kuriuo vartojami. Vadinasi, išskyrus retus atvejus (pvz., *geras klausimas, gerą valandą*) tokį būdvardį kaip, pavyzdžiui, *geras*, galima be jokių leksinių ribojimų jungti su daiktavardžiais. O jeigu taip, tai tokie būdvardžiai, kaip *geras, blogas, mielas, gražus* ir pan., nesudaro AK ir nėra aktualūs mūsų tyrimui (žr. poskyrį „Antrasis eksperimento etapas“, kur paaiškinta, kokių būdvardžių grupių atsisakyta);
- 2 kaip jau minėta, pirmuoju eksperimento etapu buvo nuspręsta tam tikram būdvardžiui pateikti ne daugiau kaip po 20 susijusių būdvardžių ir (ar) dalyvių, kurių panašumo indeksas ne mažiau kaip 0,3. Kadangi maksimalus įvertis yra 1, tai, siekiant geresnės duomenų kokybės, nutarta padidinti panašumo kriterijų nuo 0,3 iki 0,5.

Antrasis eksperimento etapas

Antrajame tyrimo etape nuspręsta vektorių sudaryti su visais DELFI.lt tekстыne pavartotais būdvardžiais, t. y. ne tik su tais, kurie sudarė kolokacijas su 97 daiktavardžiais, pateiktais „Lietuvių kalbos pastovijų žodžių junginių duomenų bazėje“. Išanalizavus dažninį būdvardžių lemų sąrašą (jame buvo 9442 lemos), nuspręsta 1) tolesnei analizei imti tik 10 ir daugiau kartų minėtame tekстыne pavartotus būdvardžius; 2) atsisakyti tam tikrų grupių būdvardžių.

Kaip jau minėta ankstesniame poskyryje, vektorių analizei trukdo plačios reikšmės ir plataus junglumo būdvardžiai, jie paprastai yra labai dažni, gali jungtis su įvairiais žodžiais, todėl atmesti šie būdvardžiai: *naujas, geras, įvairus, įdomus, puikus, gražus, blogas, nuostabus, malonus, neblogas, nenuostabus, negražus, nebeįdomus, nenaujas, nemalonus*. Atkreiptinas dėmesys, kad su kai kuriais būdvardžiais (pvz., *ydingas*) tokie būdvardžiai, kaip *blogas*, gali tikti, tačiau tokių atvejų yra nedaug palyginti su „triukšmu“, kurį gavome, kai šie junglūs būdvardžiai buvo vektorių sąrašė.

Be plačios reikšmės būdvardžių grupės, pasirinktos dar kelios grupės būdvardžių, kurie neįtraukti į naujų VE rengimą:

- 1 rūšį rodantys būdvardžiai (daugiausia su priesagomis *-inis, -iškas*), pvz., *politinis, ekonominis, finansinis, teisminis, psichomotorinis, kultūrologinis, tarptautinis, pietinis, vidurinis, atsarginis, savaiminis, šalutinis, mobilus, nekilnojamasis, valgomasis, asocijuotas, ginkluotas, lietuviškas, lenkiškas, prorusiškas, vakarietiškas, katalikiškas, lotyniškas, vakarykštis, pirmalaikis* ir kt. Didelė būdvardžių grupė su *-inis* atmesti, nes yra kolokacijose, kurios yra dvižodžiai terminai, plg. *nevyriausybinių organizacijų, struktūrinė reforma, atsarginis išėjimas, kvalifikacinis laipsnis, mineralinis vanduo, cukrinis diabetas*.

Tiesa, kai kurie būdvardžiai su *-inis* palikti, nes gali sudaryti AK, kuriomis nepasakoma rūšis ar sritis, plg. *gynybinis elgesys* (kas kita nei *gynybinis pajėgos*), *filosofinis klausimas* (kas kita nei *filosofinis traktatas*), *raktinis žodis*, arba jie neturi rūšies reikšmės, pvz.: *išskirtinis dėmesys* (plg. *ypatingas*), *kosmetinės priemonės* (plg. *paviršutiniškos*), *foninės žinios* (plg. *bendrosios*). Tokie *-inis* būdvardžiai reiškia ypatybę, bet ne rūšį (plg. *bulvarinis, policinis, hipotetinis, nišinis, užkulisinis, apsimestinis, butaforinis, anekdotinis, marginalinis*). Šie pavyzdžiai rodo, kad ne visi būdvardžiai, turintys rūšiniams būdvardžiams būdingų darybinių požymių, laikytini rūšiniais (angl. *relational*), pagal turinį jie aprašomieji (plg. Heyvaert, 2010, p. 1312–1313);

- 2 įvardžių funkciją atliekantys būdvardžiai bei dalyviai (*pastaras, vienas, ankstesnis, sekantis, minėtas, nepaminėtas, dažnas*) ir kelintiniai skaitvardžiai (*pirmas, antras, trečias...*), kurie automatiškai anotojuojant priskirti prie būdvardžių;
- 3 kai kurie spalvas žymintys būdvardžiai, pvz., *raudonas, mėlynas, oranžinis*, bet keletas žodžių, pvz., *žilas, žalias, juodas, baltas, rožinis*, palikti sąrašė, nes gali sudaryti AK, plg. *žila senovė, rožiniai akiniai*;
- 4 modaliniai būdvardžiai ar dalyviai: *galimas, būtinai*;
- 5 kai kurie vediniai: a) priesagos *-okas*, nes rodo mažesnį nei pamatiniu žodžiu pasakytą kiekį, pvz.: *grubokas, kraupokas, primityvokas, ilgokas*; b) atsisakyta deminutyvinių būdvardžių, pvz.: *švarutis, švarutėlis, pilnutėlis, liesutis, naujutėlaitis, girtutėlis*. Analizuoti šių vedinių pamatiniai žodžiai;
- 6 fizinės asmens ypatybės arba būseną žymintys būdvardžiai ar būdvardiškai vartojami dalyviai, pvz.: *suaugęs, nevedęs, girtas, kairys, ištisas, pavienis, neščia, neįgalus, pagyvenęs, heteroseksualus, pėsčias, asocialus, benamis, vedinas, nešinas, tenykštis, ilgaplaukis*;
- 7 fizinius objektų požymius žymintys būdvardžiai, pvz.: *kalnuotas, sultingas, medinis, degus, sproguš, medžiaginis, skardinis, lininis, birus, pailgas, ovalus, parduotas, panaudotas, varinis, vienspalvis, languotas, akmenuotas*.

Vis dėlto 7-oje grupėje yra išimčių, pvz., *lakus*: šis būdvardis neaktualus AK sudaryti, jei jis yra terminų dalis (pvz., *lakieji junginiai*), bet gali sudaryti AK *laki vaizduotė*, todėl paliktas prie būdvardžių, iš kurių bus sudaromos naujos vektorių eilutės, sąrašė.

Rezultatai. Atsakius būdvardžių iš aprašytų grupių (3234 lemų) ir tokių būdvardžių, kurių dažnis mažiau nei 10 kartų (4145 lemų), iš viso antrajam etapui atsirinkome apie 2000 būdvardžių lemų. Su šiais būdvardžiais sudarytos 875 VE.

Nors antrajame etape gavome geresnės kokybės duomenis, didelis naujų duomenų trūkumas yra tas, kad labai daug būdvardžių (1055), kuriems nesudaryta VE, pvz., *nelegalus, priešingas, reguliarus, slaptas, išankstinis, saldus, vienišas, skubus, elementarus, fizinis*. Žinoma, to buvo galima tikėtis, nes antrajam eksperimentui naudojome net 2000 būdvardžių, tačiau jei duomenyse neturime sinonimų, negalime pritaikyti pakeičiamumo testo.

Kaip minėta, antrajam eksperimento etapui buvo pasirinkta riba 0,5; pateikiama ne daugiau kaip 20 susijusių žodžių VE. Rezultatuose matyti, kad dėl to gavome skirtingo ilgio VE: nuo 2 žodžių iki 20 žodžių. Kuo ilgesnė VE, tuo mažiau sinonimų galima tikėtis. Pavyzdžiui, 1-oje VE yra tokie būdvardžiai: *svarbus; reikšmingas; reikalingas; naudingas; pagrindinis; didelis; aktualus; rimtas; artimas; vienintelis; ypatingas; aktyvus; paskutinis; esminis; išskirtinis; teigiamas; panašus; pavojingas; sudėtingas*.

Mažiau su *svarbus* atrodo susiję *rimtas, artimas, aktyvus, paskutinis*. Būdvardžiai *ypatingas, išskirtinis* gali sudaryti atskirą sinoniminę grupę, o labiausiai nuo *svarbus* pagal reikšmę nutolę atrodo *teigiamas, panašus, pavojingas, sudėtingas*. Nuspręsta, kad tolesniuose tyrimuose dėl plataus junglumo reikia atsisakyti būdvardžių *svarbus, reikšmingas*.

Kitose ilgesnėse VE taip pat matyti artimų ir (ar) susijusių reikšmių žodžiai, tačiau ne visi laikytini artimais sinonimais: tarkim, 15-oje VE sinonimais laikytini *aiškus, akivaizdus*, bet kiti šioje grupėje esantys būdvardžiai jau turi papildomų reikšmių požymių: *paprastas* artimas su *įprastas*, reikšminiais požymiais skiriasi *galutinis, prieinamas, teigiamas, ilgalaikis, tvirtas, realus*, plg.:

15 VE: *aiškus; akivaizdus; galutinis; teigiamas; esminis; paprastas; įprastas; tvirtas; normalus; realus; vienintelis; prieinamas; protingas; reikšmingas; ilgalaikis*.

Pastebėta, kad trumpose VE (2 žodžių) taip pat dažnai gali būti pateikti ne sinonimai, pvz., 40 VE: *žalias, šviežias*; 87 VE: *tolimas, ilgas*; 104 VE: *pilnas, kupinas*; 142 VE: *nepriklausomas, savarankiškas*; 520 VE: *prioritetinis, pirmaeilis*, o antonimai, plg. 41 VE: *atviras, uždaras*; 73 VE: *palankus, nepalankus*; 85 VE: *gausus, nedidelis*. Antonimų pasitaiko ir ilgesnėse VE, tačiau vis dėlto bendra tendencija – naujose VE surinkta daugiau pagal reikšmę artimų sinonimų, plg.: *tragiškas, skaudus, siaubingas, nelaimingas, liūdnas, baisus; nelengvas, sudėtingas, sunkus; klasikinis, tradicinis, tipiškas; žinomas, garsus, legendinis, populiarus, žymus*. Tokius rezultatus lėmė antrojo etapo sprendimas VE sudaryti su visais DELFI.lt tekстыne pavartotais būdvardžiais.

Žinoma, ir naujose VE vis dar pasitaikė įsimaišiusių netinkamų duomenų, pvz., 114 VE pateikta: *pirminis; mįslingas; pradinis*, kur tik *pirminis* ir *pradinis* laikytini sinonimais. Iš 97 daiktavardžių, su kuriais sudarytos šiame tyrime analizuotos kolokacijos, prie šios VE matome *informacija, pagalba, produktas, projektas, rezultatas, rinkimai, tyrimas*, jie visi sudaro kolokacijas tik su būdvardžiu *pirminis*. Kokie kiti daiktavardžiai lėmė, kad buvo surinkta tokia VE, nematome, todėl nežinome, pagal kokių kitų daiktavardžių konteksto panašumą į vieną grupę šioje VE pateko būdvardis *mįslingas* (plg.: 28-oje VE yra trys būdvardžiai *pasaulinis, šaltas, didis*, ir nors tarp daiktavardžių yra tik *kaina, saugumas, tinklas, čempionatas*, todėl nesunku numatyti, kad būtent tokią šios VE sudėtį lėmė daiktavardis *karas*, plg. *pasaulinis karas, šaltasis karas, didysis karas*). Turime ir 167 VE tik iš dviejų sinonimų – *pradinis* ir *pirminis*, tačiau prie jos – jau kiti daiktavardžiai nei 114 VE: *duomenys* ir *kaina*. Tai rodo, kad tie patys būdvardžiai dažnai patenka į daugiau negu vieną VE. Kitas pavyzdys: 152 VE yra sudaryta iš būdvardžių *sausas, drėgnas, šiltas, šviesus, šaltas*, čia yra net dvi antonimų poros. Jie surinkti į vieną grupę vektorijų metodu įvertinus daiktavardžius, tarp kurių yra: *kelias, oras, produktas, rezultatas, vieta, įstatymas*. Visi šie daiktavardžiai sudaro kolokacijas su *sausas* (plg. *sausas oras* ir *sausas įstatymas*), o *vieta* sudaro kolokacijas su visais šios VE būdvardžiais, išskyrus *šiltas, oras* – su visais, išskyrus *šviesus*.

Taigi nors antrojo etapo duomenys kokybiškesni, tačiau lieka paties metodo ribotumų. Galima apibendrinti, kad pagrindinė priežastis, dėl kurios negalime panaudoti visos VE surinktos informacijos, – VE atspindėtas bendras tam tikros daiktavardžių grupės junglumo semantinis laukas, ir jame nebūtinai yra tik sinonimai. Tarkim, 9-oji VE, nors ir neilga, bet nepateikia reikalingų duomenų (nėra sinonimų): *jaunas; dažnas; stiprus; garsus; mažas; gyvas; sveikas; senas; neblaivus; laimingas*. Ji atspindi būdvardžius, vartojamus asmenims apibūdinti (amžius, fizinės savybės, būseną, savijautą ir pan.), o iš daiktavardžių kaip tik ir matome asmenų pavadinimus: *asmuo, grupė, karta, komanda, ministras, moteris, vadovas, vaikas, vyras, žmogus*. 43-ojoje VE yra *privatus; nuosavas; viešas; šiuolaikinis; mažas* – tai būdvardžiai, kurie paprastai reikalingi kalbant apie turtą, gyvenimą ir panašius dalykus (plg. su šia VE susietus daiktavardžius: *aplinka, automobilis, duomenys, gyvenimas, informacija, reikalas, turtas*). Šioje eilutėje yra ir sinonimų, ir antonimų: *privatus, nuosavas* ir *privatus – viešas*. 99-ojoje VE randame: *netikėtas; neįprastas; keistas; prastas; teisingas; teigiamas; greitas; neįtikėti; galutinis; įspūdingas; įprastas; reikšmingas; skandalingas; nesėkmingas; paprastas; staigus; akivaizdus; protingas; griežtas*. Tai yra būdvardžių rinkinys, kurie (visi arba tik kai kurie) reikalingi vartojant tokius daiktavardžius, kaip *klausimas, rezultatas, būdas, įvykis*, plg. *keistas būdas, paprastas klausimas, netikėtas rezultatas, skandalingas įvykis*.

Pastebėta gana nemažai tokių sinonimijos atvejų, kurie liko nesusieti: tarkim, yra *pasakiški turtai* (reikšmė „labai dideli turtai“), tačiau būdvardžio *pasakiškas* nė vienoje VE su *didelis, didis* arba *milžiniškas* nenurodyta. Kitas atvejis: *preliminarius* ir *išankstinis* – abu būdvardžiai yra, bet nepateko į vieną VE, galbūt dėl to, kad tarp jų daiktavardžių – tik vienas bendras – *informacija*.

Sudėtinga dirbti su labai ilgomis VE, pvz., 1-oje VE pateikti tokie žodžiai: *svarbus; reikšmingas; reikalingas; naujingas; pagrindinis; didelis; aktualus; rimtas; artimas; vienintelis; ypatingas; aktyvus; paskutinis; esminis; išskirtinis; teigiamas; panašus; pavojingas; sudėtingas*, joje nustatyti būdvardžiai, kurie sudaro kelias sinonimų grupes: a) *svarbus, reikšmingas*; b) *pagrindinis, esminis*; c) *artimas, panašus*; d) *vienintelis, ypatingas, išskirtinis*. Dėl šios priežasties VE lingvistai rankomis turėjo sužymėti sinonimus, kurie panaudoti nustatant AK. Iš 875 VE eilučių šiame VE buvo pažymėta nuo 2 iki 7 sinonimų. Likusiose VE nebuvo sinonimais laikytinų žodžių.

Pakeičiamumo testo taikymas

Kai VE sužymėjo sinonimų poras, lingvistai taikė poskyryje „Riboto leksinio junglumo nustatymas“ aprašytą sinonimų pakeičiamumo testą: prie kiekvienos sinonimų poros ar VE buvo pateiktos kolokacijos su tam tikru daiktavardžiu. Jeigu su tam tikru daiktavardžiu pateikta tik viena kolokacija (su vienu sinonimu iš sinonimų grupės ar poros), laikyta, kad analizuojamas daiktavardis jungiasi tik su vienu tos sinonimų grupės būdvardžiu, ir tai reiškia, kad ši kolokacija pripažįstama kaip arbitrali: pvz., su daiktavardžiu *prekyba* buvo pateikta tik kolokacija *intensyvi prekyba*, bet nebuvo *greita prekyba*, vadinas, iš sinonimų poros *greitas, intensyvus* yra tik vienas pasirinkimas, o tai reiškia, kad kolokacija *intensyvi prekyba* priskirta arbitraliosioms.

Sinonimų pakeičiamumo testavimas atliktas neautomatiškai, nes šiame procese reikia atsižvelgti į sinonimų reikšmę ir į testuojamo daiktavardžio reikšmę: tarkim, daiktavardžio *žmogus* kolokacijoms testuoti galima naudo-

ti sinonimų eilę *žymus, garsus, legendinis*, bet kolokacijai *žymi dalis* įvertinti sinonimai *žymus, garsus, legendinis* pagal reikšmę jau netinka. Be to, tikrinant pakeičiamumo testu tam tikrą konkretų daiktavardį, kartais prie sinonimų eilės reikėdavo pridėti dar vieną sinonimą iš tos pačios VE, pvz.: tikrinant daiktavardžio *diena* būdvardžių pakeičiamumą, prie sinonimų eilės *sudėtingas, sunkus, nelengvas, rimtas* buvo pridėtas *ilgas*, nes tiko pagal savo reikšmę tikrinant kolokacijos *ilga diena* arbitralumą (plg. *ilga diena* – „sunki, nelengva diena“).

Pritaikius pakeičiamumo kitu sinonimu testą, AK pripažinta apie 650 kolokacijų iš 5000 vertintų būdvardžio ir daiktavardžio kolokacijų.

Apibendrinamosios pastabos ir tolesnių tyrimų gairės

kolokacijas: 1) veiksmožodžio ir daiktavardžio, 2) daiktavardžio ir daiktavardžio.

Šiame straipsnyje aprašyta, kaip panaudojant vektorių metodu surinktus duomenis – sinonimus – buvo vertinamas iš būdvardžių (būdvardiškai vartojamų dalyvių) ir daiktavardžių sudarytų kolokacijų ribotas leksinis junglumas. Pritaikius pakeičiamumo kitu sinonimu testą, nustatyta, kas tik būdvardžio ir daiktavardžio AK, nors tokiu pačiu metodu planuojama apdoroti ir kitokios struktūros kolokacijas: 1) veiksmožodžio ir daiktavardžio, 2) daiktavardžio ir daiktavardžio.

Dėl didelio tyrimo duomenų kiekio (vien šiame tyrime analizuota 5000 kolokacijų) ir siekiamos daugiau objektyvumo vertinant kolokacijų arbitralumą išbandėme kiek kitokį leksinių ribojimų tikrinimo būdą nei tie, kurie buvo išbandyti su nedideliu duomenų kiekiu (plg. Nesselhauf (2003)). Taigi šiame straipsnyje aprašytas lietuvių kalbotyros darbuose mažai pritaikytas pusiau automatinės analizės metodas.

Žinoma, siekiant įvertinti šio metodo kokybiškumą ir efektyvumą, reikalingi tolesni tyrimai. Tačiau jau dabar aišku, kad vertinant reikia atsižvelgti į ribotumus, išryškėjusius per tyrimą: mus riboja pasirinkto analizuoti teksto duomenys (dydis, tipas), nes dirbame su publicistikos kalbai būdingomis kolokacijomis, o iš kitokios sandaros, kitokio dydžio teksto galbūt turėtume daugiau kolokacijų su tais pačiais analizuotais daiktavardžiais. Kitaip sakant, neatmestina tikimybė, kad kolokacijos gali būti sudarytos ir su kitais sinonimais, tik jų nėra mūsų analizuotoje medžiagoje (arba tokie žodžių junginiai nelaikomi kolokacijomis). Pavyzdžiui, mūsų medžiagoje buvo tik *ypatingas žmogus*, o *išskirtinis žmogus* nebuvo, nors teoriškai *išskirtinis žmogus* irgi būtų galima kolokacija.

Iš tam tikros VE atrinkus sinonimus, būtent pagal tuos sinonimus vertinamas kolokacijos arbitralumas, tačiau tokiu būdu liko nevertintos tokios kolokacijos, kurių būdvardis nesudaro jokios sinonimų poros ar eilės su tos VE būdvardžiais, pvz.: vertinant daiktavardžio *vaizdas* kolokacijų arbitralumą pagal sinonimų eilę *keistas, neįprastas, netikėtas, neįtikėtinas*, už borto liko kolokacija *liūdnas vaizdas*. Kitas pavyzdys – *klausimas* kolokacijos: pagal sinonimų eilę *efektyvus, veiksmingas* neįtraukta kolokacija *netikėtas klausimas*. Todėl, vertinant šiuo metodu gautas AK, svarbu palyginti šiuos duomenis su kitokiu (pvz., neautomatizuotu) metodu pripažintomis kolokacijomis.

Literatūra

- Benson, M. (1990). Collocations and general-purpose dictionaries. *International Journal of Lexicography*, 3, 23-34. <https://doi.org/10.1093/ijl/3.1.23>
- Bielinskienė, A., Kovalevskaitė, J., Rimkutė, E., & Vilkaitė, L. (2017). Kolokacijų ir frazeologizmų atpažinimo kriterijai. *Kalbų studijos / Studies about Languages*, 31, 83-101. <http://dx.doi.org/10.5755/j01.sal.0.31.18710>
- Bosque, I. (2011). Deducing collocations. In Boguslavsky, I., & Wanner, L. (Eds.), *Proceedings of the 5th International Conference on Meaning-Text Theory* (pp. 1006-1023).
- Bumbulienė, I. (2018). Automatinis lietuviškų pastoviųjų žodžių junginių atpažinimas. *Magistro darbas*, Vytauto Didžiojo universitetas.
- Gharbieh, W., Bhavsar, V., & Cook, P. (2016). A word embedding approach to identifying verb-noun idiomatic combinations. *Proceedings of the 12th Workshop on Multiword Expressions*. Berlin, 112-118. <https://www.aclweb.org/anthology/W16-1817.pdf> <https://doi.org/10.18653/v1/W16-1817>
- Grigonytė, G. (2010). *Building and Evaluating Domain Ontologies: NLP Contributions*. Berlin: Logos Verlag.
- Gutiérrez, L., & Keith, B. (2019). A systematic literature review on word embeddings. In Mejia, J., Muñoz, M., Rocha, Á., Peña, A., & Pérez-Cisneros, M. (Eds.), *Trends and Applications in Software Engineering. CIMPS 2018. Advances in Intelligent Systems and Computing*, vol 865. Springer, Cham. https://doi.org/10.1007/978-3-030-01171-0_12

- 8 Henriksen, B. (2013). Research on L2 learners' collocational competence and development - a progress report. In Camilla, B., Linquist, Ch., & Layfer, B. (Eds.), *L2 Vocabulary Acquisition, Knowledge and Use. New Perspectives on Assessment and Corpus Analysis*. European Second Language Association.
- 9 Heyvaert, F. (2010). An outline for a semantic categorization of adjectives. In Dykstra, A., & Schoonheim, T. (Eds.), *Proceedings of the XIV EURALEX International Congress*. Leeuwarden/Ljouwert, 1309-1318. https://www.euralex.org/elx_proceedings/Euralex2010/123_Euralex_2010_9_HEYVAERT_An%20outline%20for%20a%20semantic%20categorization%20of%20adjectives.pdf
- 10 Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer, New York.
- 11 Lewis, M. (2001). *Teaching Collocation: Further Developments in the Lexical Approach*. London: Language Teaching Publications.
- 12 Liu, D. (2010). Going beyond patterns: Involving cognitive analysis in the learning of collocations. *TESOL Quarterly*, 44(1), 4-30. <https://doi.org/10.5054/tq.2010.214046>
- 13 Mandravickaitė, J., Rimkutė, E., & Krilavičius, T. (2016). Hybrid approach for automatic identification of multi-word wpxpressions in Lithuanian. *Proceedings of the Seventh International Conference Baltic HLT 2016*. Amsterdam, Berlin, Tokyo, Washington, DC: IOS Press, 153-159.
- 14 Marcinkevičienė, R. (2010). *Lietuvių kalbos kolokacijos*. Mokslo monografija. Kaunas. <https://www.vdu.lt/cris/bitstream/20.500.12259/250/1/ISBN9789955126560.pdf>
- 15 Martinez, I. P. (2017). Arbitrariedad y motivación en las colocaciones (Arbitrariness and Motivation in Collocations). *RLA. Revista de Teórica y Aplicada*, 55(2), 121-142. <http://dx.doi.org/10.4067/S0718-48832017000200121>
- 16 Mooney, A., Peccei, J. S., LaBelle, S., Henriksen, B. E., Eppler, E., Irwin, A., Pichler, P., & Soden, S. (2011). *The Language, Society and Power Reader*. London/New York: Routledge. <https://doi.org/10.4324/9780203836545>
- 17 Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, 24(2), 223-242. <https://doi.org/10.1093/applin/24.2.223>
- 18 Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. John Benjamins: Amsterdam. <https://doi.org/10.1075/scl.14>
- 19 Pennington, J., Socher, R., & Manning, Ch. D. (2014). GloVe: Global Vectors for Word Representation. *Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543. <https://doi.org/10.3115/v1/D14-1162>
- 20 Reilly, J., Hung, J., & Westbury, Ch. (2017). Non-Arbitrariness in Mapping Word Form to Meaning: Cross-Linguistic Formal Markers of Word Concreteness. *Cognitive Science*, 41, 845-114. <https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.12361> <https://doi.org/10.1111/cogs.12361>
- 21 Schnabel, T., Labutov, I., Mimno, D., & Thorsten, J. (2015). Evaluation methods for unsupervised word embeddings. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 298-307). 10.18653/v1/D15-1036 <https://doi.org/10.18653/v1/D15-1036>
- 22 Seretan, V. (2011). *Syntax-based collocation extraction (Vol. 44)*. Springer Science & Business Media. <https://doi.org/10.1007/978-94-007-0134-2>
- 23 Lyberis, A. (2015). *Sinonimų žodynas. 2-asis patais. leid.* Vilnius: Lietuvių kalbos institutas. <https://lkiis.lki.lt/sinonimu.jsessionid=E7C8D-1CF040FD26A71F80E707D32F0A2>
- 24 Vitkutė-Adžgauskienė, D., Dainauskas, J. J., Amilevičius, D., & Utkā, A. (2015). *Lietuvių kalbos žodžių tinklas LitWordNet. Darbai ir dienos*, 64, 101-114 https://www.vdu.lt/cris/bitstream/20.500.12259/27751/1/ISSN2335-8769_2015_N_64.PG_101-114.pdf
- 25 Williams, G. (2001). Mediating Between Lexis and Texts: Collocational Networks in Specialized Corpora. *ASp*, 63-76. <https://doi.org/10.4000/asp.1782>
- 26 Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, 27(1), 103-129. <https://doi.org/10.1093/applin/ami045>

Abstract

Jolanta Kovalevskaitė, Erika Rimkutė, Jurgita Vaičenonienė. Automatic Recognition of Arbitrary Adjective-Noun Collocations

This article focuses on arbitrary collocations – a particular type of collocations which are characterized by unmotivated relations between the constituents (differently from trivial or motivated collocations as, for example, a *beautiful day, new research*). Typically, arbitrary collocations have a certain degree of lexical restrictedness, i.e., although there may be several close synonyms, a particular one is preferred in a certain word combination, for example, *broad/wide outlook vs. big outlook; strong health vs. powerful health*.

As a result of the analysis of 5000 adjective-noun collocations retrieved from the “Database of Lithuanian Multiword Expressions”, approximately 650 arbitrary collocations were identified using the synonym substitution test: if the adjectival component of the collocation (adjective or participle) could not be replaced by a close synonym, the collocation was considered arbitrary. The methods of computational linguistics, or Word Embedding Approach in particular, were used to automatically retrieve close synonyms of adjectives in adjective-noun collocations. Nouns and participles were automatically grouped into approximately 800 vector strings. The article describes in detail the steps in data processing and analysis as well as arbitrary collocation identification criteria and methods by using the Global Vector (GloVe) model.

Apie autores

JOLANTA KOVALEVSKAITĖ

Dr., Vytauto Didžiojo universitetas, Humanitarinių mokslų fakultetas, Kompiuterinės lingvistikos centras, Kaunas, Lietuva

Mokslinių interesų sritys

Tekstynų lingvistika, pastovieji žodžių junginiai, leksikografija

Adresas

V. Putvinskio g. 23-216, Kaunas, LT-44243, Lietuva

El. paštas

jolanta.kovalevskaite@vdu.lt

ERIKA RIMKUTĖ

Doc. dr., Vytauto Didžiojo universitetas, Humanitarinių mokslų fakultetas, Lituanistikos katedra, Kaunas, Lietuva

Mokslinių interesų sritys

Tekstynų lingvistika, kompiuterinė lingvistika, pastovieji žodžių junginiai, automatinė gramatinė analizė

Adresas

V. Putvinskio g. 23-216, Kaunas, LT-44243, Lietuva

El. paštas

erika.rimkute@vdu.lt

JURGITA VAIČENONIENĖ

Doc. dr., Vytauto Didžiojo universitetas, Humanitarinių mokslų fakultetas, Kompiuterinės lingvistikos centras, Kaunas, Lietuva

Mokslinių interesų sritys

Tekstynų lingvistika, vertimo studijos, skaitmeninė humanitarika

Adresas

V. Putvinskio g. 23-216, Kaunas, LT-44243, Lietuva

El. paštas

jurgita.vaicenoniene@vdu.lt

