

Kolokacijų ir frazeologizmų atpažinimo kriterijai

Towards the criteria for identification of idioms and collocations

SAL 31/2017

Kolokacijų ir frazeologizmų atpažinimo kriterijai

Received 07/2017

Accepted 11/2017

COMPUTATIONAL LINGUISTICS / KOMPIUTERINĖ LINGVISTIKA

Agnė Bielinskienė

Dr., Vytauto Didžiojo universitetas, Lietuva

Jolanta Kovalevskaitė

Dr., Vytauto Didžiojo universitetas, Lietuva

Erika Rimkutė

Doc. dr., Vytauto Didžiojo universitetas, Lietuva

Laura Vilkaitė

Dr., Vilniaus universitetas, Lietuva

 <http://dx.doi.org/10.5755/j01.sal.0.31.18710>

Straipsnyje analizuojami du pastoviųjų junginių tipai: kolokacijos ir frazeologizmai. Pateikiami šių leksikos vienetų apibrėžimai, samprata užsienio ir lietuvių autorių darbuose. Su pastoviųjų junginių nustatymo sunkumais susidurta vykdant projektą „Lietuvių kalbos pastoviųjų žodžių junginių automatinis atpažinimas (PASTOVU)“. Tiek rankomis, tiek automatiškai analizuojant iš delfi.lt portalo sudarytą tekstyną, iškilo problema, kad vis dar trūksta aiškių kriterijų, kuriais remiantis būtų galima atskirti kolokacijas ir frazeologizmus lietuvių kalboje, todėl šio tyrimo tikslas – aprašyti kriterijus, kurie leistų pirmiausia lingvistams kuo objektyviau tekste atpažinti kolokacijas bei frazeologizmus ir kuo aiškiau atskirti šiuos pastoviųjų junginių tipus. Kolokacijoms ir frazeologizmams bendri trys požymiai: vartojimo pastovumas, semantinis nedalomumas ir sustabarėjimas. Visų šių požymių laipsnis gali būti nustatomas pagal testus, paremtus kriterijais. Taikant atpažinimo testus, pateikiant atitinkamus pavyzdžius, bandyta teoriškai ir praktiškai atskleisti kolokacijų bei frazeologizmų skirtumus ir panašumus. Matyti, kad ir kolokacijų, ir frazeologizmų klasės nėra vienalytės: abiem atvejais yra tipišku ir ne tokiu tipišku junginių. Vartojimo pastovumą ir semantinį nedalomumą siūlytume laikyti esminiais požymiais, pagal jų kriterijus ir testus galima būtų atskirti, ar pastovusis junginys laikytinas kolokacija, ar frazeologizmu.

REIKŠMINIAI ŽODŽIAI: pastovieji junginiai, kolokacijos, frazeologizmai, frazeologija, tekstynas, anotavimas, lietuvių kalba.



Research Journal
Studies about Languages
No. 31/2017
ISSN 1648-2824 (print)
ISSN 2029-7203 (online)
pp. 83-101
DOI 10.5755/j01.sal.0.31.18710
© Kaunas University of Technology

Įvadas

Pastovieji žodžių junginiai, nors ir seniai tyrinėjama sritis, bet aktuali iki šiol. Lietuvoje parengta keletas frazeologijos ir posakių žodynų, tačiau kolokacijų žodynų pasigendama¹. Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centro mokslininkams kartu su kolegomis iš Baltijos pažangių technologijų instituto vykdant projektą „Lietuvių kalbos pastoviųjų žodžių junginių automatinis atpažinimas (PASTOVU)“², analizuojami klausimai, kaip atpažinti pastoviuosius junginius (toliau – PJ) tiek automatiškai, tiek žmogui. Projektas apima dviejų PJ klasių – kolokacijų ir frazeologizmų – tyrimus. Kaip vienas svarbiausių šio projekto rezultatų bus parengta tekstynų paremta kolokacijų ir frazeologizmų duomenų bazė ir lietuvių kalbos kolokacijų žodynas.

Šio tyrimo objektas apima dvi PJ klases: kolokacijas ir frazeologizmus. Apibrėžiant kolokacijas, teigiama, kad tai įprastas tam tikrų leksemų vartojimas drauge, tiksliau – prasmingi ir gramatiškai mažiausiai dviejų žodžių junginiai, kurių dėmesys yra riboto junglumo (Marcinkevičienė, 2010). Frazeologizmais lietuvių kalboje įprastai laikomi vaizdingi, vientisos, perprasmintos reikšmės, pastovios leksinės ir gramatinės sandaros žodžių junginiai (Jakaitienė, 2010, p.284). Vis dėlto praktiškai dažnai gana keblu atskirti kolokacijas nuo frazeologizmų, nes jiems būdingi visiems PJ bendri požymiai (pastovumas, sustabarėjimas, junginį sudarančių žodžių dažnas ėjimas drauge), be to, ir kolokacijų, ir frazeologizmų pagrindas yra metaforos, metonimijos, palyginimai (Lamiroy, 2016, pp.140–141). Kokie požymiai svarbūs atpažįstant kolokacijas ir frazeologizmus tekste ir kaip tuos požymius nustatyti, toliau diskutuojama šiame straipsnyje.

Straipsnio tikslas – apibrėžti kolokacijas bei frazeologizmus ir aiškiai aprašyti jų atpažinimo kriterijus, kad būtų galima aiškiau atskirti šiuos PJ tipus. Pripažįstama, kad klausimas, kokiais kriterijais remtis siekiant atpažinti skirtingus PJ tipus – vis dar aktualiausia frazeologijos problema (Granger et al., 2008, p.45). Šis klausimas kilo ir vykdant projektą PASTOVU, kai frazeologizmai ir kolokacijos buvo anotuojami tekstynėse rankomis.

PJ atpažinti projekto PASTOVU metu buvo sudarytas internetinės publicistikos tekstynas iš delfi.lt svetainės, kuris apima 2014 m. kovo – 2016 m. lapkričio mėn. laikotarpį ir yra 72 mln. žodžių dydžio. Imti 12 kategorijų tekstai: DELFI veidas, projektai, DELFI mokslas, DELFI auto, sportas, DELFI gyvenimas, DELFI žmonės, DELFI pilietis, verslas ir kt. Tekstynas automatiškai morfologiškai anotuotas.

Šio straipsnio tyrimui buvo pasirinkta dalis minėto tekstyno – 72 000 žodžių. Šiuose tekstuose automatiškai buvo pažymėti *Frazeologijos* žodyne (2015) pateikti frazeologizmai ir įvairaus tipo PJ iš *Lietuvių kalbos daiktavardinių frazių žodyno* (toliau LKDFŽ 2012), iš viso automatiškai pažymėti PJ sudarė 13 proc. Kitas kolokacijas ir frazeologizmus turėjo atpažinti ir pažymėti keturi lingvistai³. Kiekvieną tekstą peržiūrėjo po du lingvistus. Šis lingvistų darbas svarbus tuo, kad rankomis anotuoti duomenys naudojami automatiniam mokymuisi kaip etalonas, t. y. galima palyginti, kurie leksinės traukos įverčiai (angl. *lexical association measures*)⁴ tuose pačiuose tekstuose rado PJ, kokius PJ atpažino, kaip juos sureitingavo. Tačiau šio straipsnio temai lingvistų anotavimo duomenys buvo svarbūs tuo, kad atskleidė problemą, kaip nevienodai atpažįstami ir traktuojami PJ. Ši patirtis vėl paskatino imtis klausimo apie kolokacijų bei frazeologizmų atpažinimo kriterijus taikant tekstynų ir kompiuterinės lingvistikos metodus.

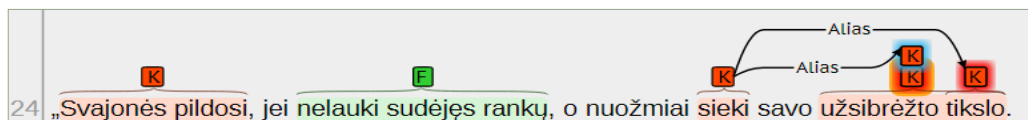
1 Iš dalies šią nišą užpildė 2012 m. sudarytas *Lietuvių kalbos daiktavardinių frazių žodynas*.

2 Projektą finansuoja Lietuvos mokslo taryba (projekto nr. LIP-027/2016).

3 Tai šio straipsnio autorės, dirbančios tekstynų ir kompiuterinės lingvistikos srityje, turinčios PJ analizės patirties.

4 Šiais įverčiais skaičiuojama trauka tarp junginių sudarančių komponentų, vertinant jų junglumo laipsnį. Gauti įverčiai gali būti naudojami galimiems PJ nustatyti, įvertinti bei klasifikuoti. Leksinės traukos įverčiai išbandyti šiame projekte nustatant dvizodžius ir trižodžius PJ (plačiau žr. Mandravickaitė et al., 2016).

Kolokacijos ir frazeologizmai buvo anotuojami rankomis minėtame 72 000 žodžių tekстыne naudojant anotavimo įrankį BRAT⁵. Šis įrankis patogus anotuoti rankomis, o anotavimo pažymos gali būti apdorojamos ir interpretuojamos natūraliosios kalbos analizės metodais. Kadangi įrankio galimybės leidžia pažymėti ne tik šalia esančius (plg. *svajonės pildosi*), bet ir vieną nuo kito nutolusius žodžius, labai patogu žymėti kolokacijas ir frazeologizmus, tarp kurių dėmenų įsiterpia vienas ar keli žodžiai, kurie sujungiami parodant jų sąsają (plg. *sieki ... tikslo*) (žr. 1 pav.).



1 pav.

Kolokacijų ir frazeologizmų žymėjimas anotavimo įrankiu BRAT⁶

Iš viso buvo pažymėti 4 335 PJ: 4 052 kolokacijos ir 283 frazeologizmai. Atkreiptinas dėmesys, kad į šį skaičių įeina visos kaitybinės to paties PJ formos (*siekė tikslo, siekti tikslo* ir pan.), todėl skirtingų kolokacijų ir frazeologizmų yra mažiau – 2 788 (neskiriant kolokacijų ir frazeologizmų, nors akivaizdžiai matyti, kad kolokacijos, o ne frazeologizmai buvo vartoti įvairesnėmis kaitybinėmis formomis). Šie PJ visame analizuotame tekстыne sudaro apie 15 proc.

757 PJ užfiksuoti su įsiterpiančiais žodžiais (jie sudaro apie 17 proc. visų analizuotų PJ). Rasti 1–21 žodžio įsiterpimai, dažniausiai įsiterpia 1–2, rečiau 3–5 žodžiai (žr. 1 lentelę).

Tekstynų lingvistikoje nurodoma, kad galimas atstumas tarp kolokacijos dėmenų yra ± 4 žodžiai (Sinclair, 1991), tačiau šis tyrimas atskleidė, jog to paties junginio dėmenys ne taip jau labai nutolę vienas nuo kito⁸. Dažniausiai įsiterpia vienas ar du žodžiai: ***iškovojo dramatišką pergalę, laiką ramiai leidžia, atveria daugybę galimybių, nesėkmes pirmąją dieną patyrė.***

Apibendrinus rankinį PJ anotavimą, nustatyta, kad sutampa gana nedaug skirtingų anotuoju duomenų. Viena lingvistų grupė analizavo trečdalį tekстыno, kita grupė – du trečdalius. Mažesnę tekстыno dalį anotavusių lingvistų kolokacijų žymėjimas sutampa apie 40 proc.; daugiau tekstų anotavusių lingvistų anotavimo sutapimas dar mažesnis – apie 33 proc. Tai reiškia, kad tik atitinkamai 40 ir 33 proc. tų pačių kolokacijų pažymėta abiejų lingvistų. Galima daryti prielaidą, kad kuo daugiau duomenų, tuo mažesnis anotavimo tikslumas. Bendras abiejų grupių lingvistų kolokacijų žymėjimo sutapimas apie 37 proc.

Manyta, kad lengviau seksis atpažinti frazeologizmus. Vis dėlto net ir šių PJ anotavimo sutapimas gana nedidelis: pirmos grupės lingvistų – apie 41 proc., antros grupės – apie 32 proc.; bendras frazeologizmų anotavimo sutapimas taip pat apie 37 proc.

Manyta, kad lengviau seksis atpažinti frazeologizmus. Vis dėlto net ir šių PJ anotavimo sutapimas gana nedidelis: pirmos grupės lingvistų – apie 41 proc., antros grupės – apie 32 proc.; bendras frazeologizmų anotavimo sutapimas taip pat apie 37 proc.

Kiek žodžių įsiterpia	Kiek procentų sudaro tarp junginių su įsiterpimais
1	67,1
2	19,2
3	5,4
4	3,4
5 ⁷	2,4

1 lentelė

Tarp PJ įsiterpintys žodžiai

5 Šį įrankį sukūrė kelios tyrėjų, besidominčių anotavimu, grupės. Šio įrankio naudojimo galimybės tekстыnams anotuoti gana įvairios: galima anotuoti įvardytąsias esybes, anaforas, žodžių junginius nurodant sintaksinę priklausomybę ar kitokias pažymas, žymėti įvairias metažymas, morfemas, pastraipas ar ilgesnes teksto atkarpas ir pan.

6 Sutrumpinimų paaiškinimai: K – kolokacija; F – frazeologizmas; rodykle su užrašu *Alias* parodoma žodžių sąsaja (nežymint sintaksinio ryšio); tekstas paimtas iš delfi.lt tekстыno žinių rubrikos.

7 6–21 įsiterpintys žodžiai iš viso sudaro tik 2,4 proc., todėl lentelėje nepateikti.

8 Pavyzdys, kur tarp kolokacijos *kelia klausimus* dėmenų įsiterpia 21 žodis (čia ir kitur įsiterpimas pažymėtas laužtiniais skliaustais): *Saugumo Taryboje Lietuva taip pat nuolat **kelia** [atsakomybės už karo nusikaltimus ir nusikaltimus žmoniškumui, neteisėtos prekybos ginklais, tarptautinės humanitarinės teisės pažeidimų, seksualinio smurto, moterų ir vaikų apsaugos konfliktuose] **klausimus***. Šis sakinytis rašytinėje kalboje atrodo natūraliai, taigi net labai ilgi įsiterpimai būdingi vartosenai ir nėra tik visiškos išimty.

Kaip paaiškėjo, anotavimo nesutapimų atsirado dėl tokių priežasčių:

- 1) vienas iš lingvistų nepažymėjo tam tikro PJ;
- 2) vienas iš lingvistų tam tikrą PJ priskyrė kitam tipui (pvz., vienas tą patį junginį laikė frazeologizmu, o kitas – kolokacija);
- 3) lingvistai pažymėjo skirtingas to paties PJ ribas, todėl nors abu atpažino tą patį PJ, bet dėl skirtingo ilgio anotavimo programa šiuos žymėjimus vertino kaip du skirtingus junginius;
- 4) lingvistai skirtingai interpretavo PJ skyrimo kriterijus.

Žinoma, netikslumų atsirado ir dėl techninių dalykų, pvz., anotuojant su BRAT, tais atvejais, kai kas nors įsiterpia į PJ, tarp vieno ir kito (-ų) dėmens (-ų), turi būti nutempiama rodyklė (žr. jungtį *Alias* tarp *sieki* ir *tikslo* 1 pav.), rodanti, kad tie žodžiai yra susiję. Net ir tai, ar rodyklė nutempiama iš kairės į dešinę ar atvirkščiai, turi įtakos, kad tas pats junginys, nors vienodai anotuotas dviejų lingvistų, automatiškai traktuojamas kaip du skirtingi junginiai.

1–4 punktuose nurodyti anotavimo nesutapimai leido labiau įsigilinti į klausimą, kur lingvistų nuomonės išsiskiria ir kodėl. Anotavimo nesutapimų pavyzdžius aptarsime plačiau.

Neretai vienas iš lingvistų tam tikrą PJ priskirdavo kitam tipui, pvz., kolokacija *nešauna mintis* vieno iš anotuotojų buvo pažymėta kaip frazeologizmas, nes neatpažinta, kad tik vienas šio PJ dėmuo yra perkeltinės reikšmės (*nešauna* – „neateina“, o *mintis* metaforiškai siejama su strėle). Metaforinės kolokacijos (pvz., *karštis atslūgs*) gali būti painiojamos su frazeologizmais, nes ir vieni, ir kiti dažnai susiformavę metaforiniu reikšmės perkėlimo pagrindu. Renkantis tarp kolokacijos ir frazeologizmo, svarbu nustatyti, ar viso, ar tik dalies junginio reikšmė perkeltinė. Pavyzdžiui, junginys *ašarų nubraukėme* buvo pažymėtas kaip kolokacija, bet laikytinas frazeologizmu, nes čia viso junginio reikšmė perkeltinė (frazeologizmas, perteikiantis reikšmę „verkti“).

Iš anotavimo paaiškėjo, kad PJ ribų klausimas – vienas iš sudėtingiausių. Kitaip tariant, dažnai kyla klausimas, kuriuos žodžius dar reikėtų skirti prie PJ, o kurie jau nepriklauso junginiui. Nesant aiškių kriterijų, lingvistų interpretacijos kartais išsiskiria. Tokių kolokacijų buvo su nekaitomais žodžiais (*dideliu užsidegimu* – *su dideliu užsidegimu*, *peršasi išvada* – *peršasi išvada, kad...*), taip pat su kaitomais, dažniausiai veiksmažodžiais (*laisvės atėmimo baudmė* – *skirta laisvės atėmimo baudmė; tikrąjį veidą* – *parodė tikrąjį veidą*). Pavyzdžiui, *parodė tikrąjį veidą* atveju, abu anotuotojai atpažino šį PJ, bet nevienodai nustatė jo ribas. Be to, vienas anotuotojas šį junginį pažymėjo kaip kolokaciją, o kitas – kaip frazeologizmą. Vadinas, netiksliai nustatytos ribos gali suklaidinti nustatant PJ tipą.

Junginio ribų klausimas aktualus ne tik kolokacijoms, bet ir frazeologizmas. Pavyzdžiui, E. Jakaitienė (2010, p.301), aptardama lyginamuosius frazeologizmus, rašo, kad vienų kalbininkų nuomone, frazeologizmu laikytina visa lyginamoji konstrukcija su vienu ar keliais žodžiais, kurie dažniausiai nusako lyginimo aspektą, pvz.: *lekia kaip akis išdeges*, *juodas kaip bato aulas*, *prarijo kaip šuo muilą* (lyginimo aspektą nusakantys žodžiai pabraukti). Kiti kalbininkai, kaip teigia E. Jakaitienė, mano, kad tokio tipo pasakymuose frazeologizmai yra tik žodžiai, kuriantys vaizdą: *kaip akis išdeges*, *kaip bato aulas*, *kaip šuo muilą*, o ankstesniuose pavyzdžiuose pabraukti žodžiai sudaro šių frazeologizmų būtinąją (frazeologinę) aplinką. Vadinas, net teoriškai ne visada sutariama dėl PJ ribų.

PJ sąrašuose matyti daug junginių, kuriuos vienas anotuotojas pažymėjo, o kitas – nepažymėjo. Diskutuojant paaiškėjo, kad dažnai tokiais atvejais lingvistai skirtingai interpretavo iš anksto apibrėžtus PJ skyrimo kriterijus. Kaip jau minėta, nemažai painiavos kelia perkeltinės reikšmės nustatymas visam junginiui arba tik vienai jo daliai, tačiau ne mažiau keblus klausimas yra junginio pastovumo įvertinimas, ypač kolokacijų atveju, kurių pastovumą le-

mia vartojimo įprastumas (konvencionalumas). Vertindami vartojimo įprastumą, anotuotojai susiduria su subjektyvumu: kas vienam įprasta, kitam – nelabai įprasta; kaip patikrinti, ar konkrečių dviejų ar trijų žodžių junginys yra įprastas. Diskutuodami lingvistai minėjo kriterijus „didelis / mažas kintamumas“, asociacijos stiprumas („pasakius vieną kolokacijos žodį kitas ateina į galvą savaime“), ir šie aspektai buvo svarbūs apibrėžiant kriterijus, kurie padėtų įvertinti kolokacijų vartojimo pastovumą (žr. skyrių *Kolokacijų skyrimo kriterijai*).

Vertinant kolokacijų pastovumą, nesutapimų atsirado ir dėl to, kad vieni lingvistai labiau rėmėsi statistiniu žodžių sukibimo vertinimu. Kaip minėta, tekstuose, kuriuos anotavo lingvistai, automatiškai buvo pažymėti ne tik *Frazeologijos* žodyne pateikti frazeologizmai, bet ir įvairaus tipo PJ iš LKDFŽ, tarp kurių buvo daug PJ, kurie statistiškai įvertinti kaip pastovieji, pvz., *ankstyvas rytas, atsiskaitė litais, atsargiai elgtis*. Tačiau dėl šių kolokacijų anotuotojų nuomonės išsiskyrė, nes tarp tokių junginių daug trivialiųjų kolokacijų, kurių junglumas motyvuotas. Dauguma tyrėjų tikrosiomis laiko tik arbitralumu pasižyminčias kolokacijas (plg. Marcinkevičienė, 2010, p.91) – nemotyvuoto junglumo junginius, pvz., *laikas eina*, bet ne **laikas žingsniuoja*. Riba tarp arbitraliųjų ir trivaliųjų kolokacijų neaiški, kai kurie tyrėjai kolokacijomis laiko ir vienas, ir kitas, bet, kaip teigia R. Marcinkevičienė (2010, p.78), „jei spęstume pagal esamus kitų kalbų kolokacijų žodynus [...], demarkacinė linija, skirianti tikrąsias ir trivaliasias kolokacijas, arba, kitaip pasakius, leksinius ir semantinius apribojimus, daugeliu atveju sutampa su abstrakčiųjų ir konkrečiųjų daiktavardžių junglumu ir vartosena. Kolokacijos, kurių pagrindą sudaro abstraktieji daiktavardžiai, gali būti laikomos tikrosiomis, nemotyvuotomis, o pastovūs junginiai su konkrečiais daiktavardžiais paprastai remiasi nekalbine tikrove ir yra trivalūs“. Šis kolokacijų požymių aspektas privertė lingvistus dar kartą tikslinti kolokacijų sampratą ir atsargiau remtis statistiniais duomenimis.

Atsižvelgiant į aprašytas problemas, darytina išvada, kad tiek kolokacijų, tiek frazeologizmų atpažinimas yra sudėtingas procesas ir iš esmės nesiskiria savo sudėtingumu. Nustatyti PJ sunku ne tik automatiškai, bet ir anotuojant tekstus rankomis, todėl reikalingi aiškesni kriterijai, kuriais remiantis pirmiausia lingvistai galėtų lengviau nustatyti PJ, lengviau atskirti kolokacijas ir frazeologizmus. Apie šiuos kriterijus diskutuojama 3 ir 4 skyriuose *Kolokacijų skyrimo kriterijai, Frazeologizmų samprata ir požymiai*.

Kolokacijų samprata

Nors kolokacijos tiriamos jau seniai ir susilaukė nemažai dėmesio ir lietuvių lingvistikoje, vis dėlto aiškaus sutartinai priimto kolokacijų apibrėžimo nėra, o skirtingi mokslininkai jas suvokia ir apibrėžia gana nevienodai (apžvalgą anglų kalba žr., pvz., Nesselhauf, 2004; apžvalgą lietuvių kalba žr. Marcinkevičienė, 2010). Nepaisant to, jog *kolokacijos* terminas ir samprata egzistuoja jau bent nuo šeštojo praėjusiojo amžiaus dešimtmečio Firtho darbų (Firth, 1957), vis dar nėra bendro sutarimo, kaip tiksliausia apibrėžti, kas laikytina kolokacijomis, kaip atpažinti kolokacijas, kaip tinkamai įvertinti kolokacijų traukos stiprumą (Gries, 2013).

Apibendrinant dažniausiai pateikiami du požiūriai į kolokacijas – **dažnumu paremtas ir frazeologinis** (Nesselhauf, 2004, taip pat žr. Granger et al., 2008).

1 Remiantis pirmuoju, **dažnumo, požiūriu** (angl. *frequency-based approach*), kolokacijomis laikomi bet kurie žodžiai, vartojami drauge konkrečiame kontekste (pvz., Sinclair, 1991). Dažniausiai pagal šį požiūrį leidžiama, kad tarp kolokacijos dėmenų įsiterptų iki keturių kitų žodžių. Toks statistinis požiūris į kolokacijas yra grįstas šių junginių vartojimo dažnumu bei ribotu leksiniu junglumu ir paprastai naudojamas tekstynų lingvistikoje. Kolokacija apibrėžiama taip: tai žodžiai, kurių tikimybė būti pavartotiems drauge didesnė už atsitiktinę. Žiūrint į kolokacijas kaip į statistiškai įvertinamą reiškinį, galima nustatyti aiškius bei objektyvius kriterijus ir, jais remiantis, automatinis kolokacijų atpažinimas nebekelia problemų.

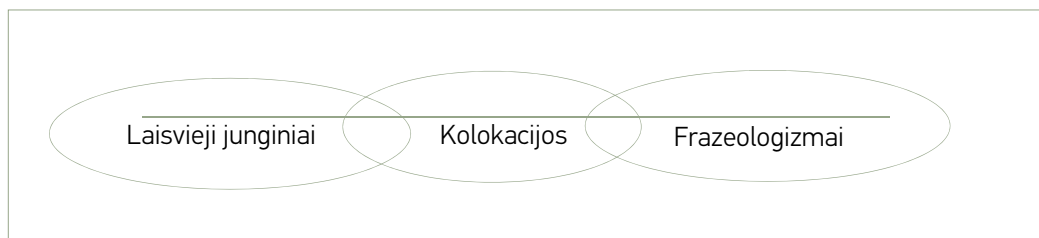
Kolokacijų samprata ir požymiai

Dažniausiai šiuo metodu taikomas tam tikro minimalaus dažnumo tekstyne kriterijus drauge su vienu iš statistinių įverčių traukai tarp žodžių apskaičiuoti (pvz., MI – *mutual information score* (Church ir Hanks, 1990), *t-score*, *z-score*, *log-likelihood* ir t. t.). Tokiu požiūriu be J. Sinclairio ir G. Kjellmero, M. Stubbs vadovaujasi, pavyzdžiui, D. Biberis, S. Johanssonas, G. Leechas, S. Conradas ir E. Fineganas (1999), D. Shinas ir P. Nationas (2008). Nors šis požiūris leidžia lengvai atskirti kolokacijas nuo laisvųjų junginių, tačiau, remiantis tokiu požiūriu, kolokacijos labai varijuoja pagal tai, kiek išlaikomas jų prasminių sandų savarankiškumas: išskiriami tiek visiškai tiesioginę reikšmę turintys junginiai (pvz., angliškai *last night*), tiek junginiai su perkeltinę reikšmę turinčiais elementais (pvz., angliškai *broke the silence*). Be to, jau anksčiau minėto J. Sinclairio krypties tyrėjai (pvz., G. Kjellmer) atkreipė dėmesį, kad reikia nepamiršti ir gramatinio taisyklingumo, nes tarp automatiškai atpažintų statistiškai reikšmingų kolokacijų (sutapčių (angl. *co-occurrence*)) pasitaiko daug nepilnų junginių.

2 Remiantis antruoju, **frazeologiniu, požiūriu** (angl. *phraseological approach*), atpažįstant kolokacijas taikomas semantinis kriterijus, o kolokacija apibrėžiama kaip žodžiai, kurių reikšmė nėra visiškai semantiškai skaidoma (dažniausiai kuris nors vienas žodis išlaiko savo tiesioginę reikšmę, o kito reikšmei būdingas perkėlimas), o junginio sandara apribota (angl. *restricted*) – t. y. žodžiai junginyje negali būti lengvai keičiami panašios reikšmės žodžiais ar sinonimais. Tokios kolokacijos yra įprastos kalbinėje bendruomenėje, konvencionalios. Šio požiūrio atstovai (A. Cowie, I. Mel'čuk, F. J. Hausmann) dažnai suvokia kolokacijas kaip junginius, esančius tam tikro kontinuumo tarp laisvųjų junginių (angl. *free combinations*), kurių reikšmė visiškai daloma, ir frazeologizmų (angl. *idioms*), kurių reikšmė visiškai perkeltinė, viduryje (žr. 2 pav.) (pvz., Howarth (1998), taip pat Mel'čuk (1998).

2 pav.

Kolokacijos samprata
PJ kontinuume



Kolokacija dažnai suvokiama kaip PJ, kuriame dėmenų svoris yra nevienodas, tačiau tą asimetriją (Barnbrook et al. Terminas (2013) abiejų požiūrių atstovai aiškina skirtingai: I. Mel'čukui ir F. J. Hausmannui ši vieno dėmens priklausomybė nuo kito pagrįsta semantika (pagrindiniam dėmeniui parenkamas priklausomas dėmuo, derantis pagal reikšmę), o, pavyzdžiui, J. Sinclairio supratimu, asimetriją lemia nevienodas kiekvieno kolokacijos dėmens dažnis (žodis, kurio dažnis didesnis, su antruoju žodžiu bus susijęs stipriau ir jam ta kolokacija bus reikšmingesnė) (Nesselhauf, 2004, pp.16–19). Pagal F. J. Hausmanną, tipišką kolokacijos sudarytos iš dviejų žodžių, kur vienas yra pagrindinis („bazė“), semantiškai savarankiškas ir pasirenkamas kaip atraminis; jis lemia kito, priklausomojo, žodžio pasirinkimą reikiama kolokacijai sudaryti, pvz.: kolokacijoje *meet the deadline* pagrindinis yra daiktavardis (*deadline*), kuriam kaip tinkamas, konvencionalus kolokatas parenkamas veiksmažodis *meet*, o ne, pvz., *fulfill*. Pasirinkimo galimybės priklauso nuo pagrindinio žodžio junglumo apimtys, kuri svarbi vertinant, ar junginys laikytinas kolokacija (žr. 3.2 poskyrį apie vartojimo pastovumo kriterijus).

Vadovaujantis frazeologiniu požiūriu, pastebėtina, kad riba tarp kolokacijų ir frazeologizmų ar kolokacijų ir laisvųjų žodžių junginių nėra iki galo aiški ir tarp šių junginių galima brėžti sa-

votišką kontinuumą (žr. 2 pav.). Tai viena iš priežasčių, kodėl frazeologinis požiūris susilaukia kritikos (pvz., Granger et al., 2008). Tačiau A. Wray (2012) pabrėžia, kad vis dėlto toks PJ kaip kontinuumo suvokimas ganėtina supaprastintas, nes galėtume brėžti daug tokių kontinuumo tiesių: pagal PJ dažnumą (kolokacijos būtų gerokai dažnesnės už frazeologizmus), pagal PJ sustabarėjimo lygį, pagal reikšmės perkėlimo lygį ir t. t. Taigi tiksliai nustatyti ribą tarp skirtingų PJ klasių sudėtinga.

Lyginant dažnumu pagrįstą ir frazeologinį požiūrį į kolokacijas, teigiama, kad antrasis yra fragmentiškas ir neapima visų galimų kolokacijų, o pirmasis suteikia daug „žalios“ medžiagos, bet ji ne visa yra kokybiška (Siepmann, 2005, p.411). Vis dėlto tikslumo dėlei svarbu akcentuoti, kad junginiai, atpažįstami kaip kolokacijos vadovaujantis dažnumo ir reikšmės kriterijais, nebūtinai yra skirtingi. Dalis kolokacijų, kurios turi aiškų semantinį elementą, kartu gali būti dažnos kalboje, o jų dėmenys – statistškai susiję. Kitaip tariant, remiantis pirmuoju požiūriu atpažįstamos kolokacijos ar bent jų dalis būtų atpažįstamos ir remiantis antruoju požiūriu. Taigi skirtingi požiūriai į kolokacijas nebūtinai yra visiškai nesuderinami.

Dažniausiai sutariama, kad dalis sunkumų apibrėžiant ir tiriant kolokacijas kyla dėl to, kad kolokacijos tėra tik vartojimo tendencijos, o ne visiškai pastoviai vartojami žodžių junginiai (Brown, 2014; Wray, 2002). R. Marcinkevičienės (2010, p.91) teigimu, „leksiniai apribojimai, kurie ir laikytini tikrųjų kolokacijų junglumo pagrindu, yra ne griežtų taisyklių, o tendencijų reikalas“. Anot A. Wray (2002), kolokacijos yra įprastai vartojamos žodžių poros, tačiau tai nereiškia, kad tie žodžiai išimtinai turi būti vartojami tik kartu. Pavyzdžiui, lietuviškai *kantrybė* gali *trūkti* ar *plyšti*, *kainos* gali *kilti*, *didėti* ar *augti*, vadinasi, šiuo atveju ir daiktavardis *kantrybė*, ir daiktavardis *kaina* su savo junglumo partneriais sudaro kelias lietuvių kalboje įprastas kolokacijas.

Kita problema atpažįstant kolokacijas yra ta, kad įvairios sandaros žodžių junginiai gali būti laikomi kolokacijomis. Paprastai skiriamos leksinės kolokacijos (sudarytos bent iš dviejų prasminių žodžių, pvz., *daryti įtaką*, *kelti klausimą*) ir gramatinės kolokacijos (kolokacijos su funkciniais žodžiais, pvz., *jaudintis dėl*, *kol kas*, *bet kuris*) (tokių skirtį taiko, pavyzdžiui, Durrant (2009). Jei tiriamos ir gramatinės kolokacijos, tai, žinoma, gerokai praplečia tyrimo objektą.

Dėl kolokacijos sudarančių žodžių skaičiaus taip pat yra įvairių nuomonių (Siepmann, 2005; Granger et al., 2008), bet dažniausiai kolokacijomis laikomi dvižodžiai junginiai, nors tai retai aptariama eksplisitiškai (Nesselhauf, 2004, p.19).

Šiame straipsnyje ir projekte tiriamos tik leksinės kolokacijos, t. y. žodžių junginiai, sudaryti bent iš dviejų prasminių žodžių. Kolokacijos samprata apima tokius leksinės kolokacijos požymius: kolokacija – bent dviejų leksemų PJ, sudarantis vieną reikšmės vienetą, įprastai gramatiškai taisyklingą, kuris dėl leksinių apribojimų yra reguliariai pasikartojantis tam tikros diskurso bendruomenės sakininėje ir rašytinėje komunikacijoje.

Be tekstyno analizės nustatyti kokio nors junginio dažnumą sudėtinga – iš tyrimų matyti, kad gimtakalbiai gali tiksliai nusakyti, kurie junginiai kalboje labai dažni arba labai reti, tačiau dėl vidutinio dažnumo junginių tiksliai nuspręsti negali (Siyanova-Chanturia ir Spina, 2015). Projekto PASTOVU metu anotuojant PJ rankomis, taip pat susidurta su problema, kaip įvertinti, ar kolokacija įprasta, konvencionali. Kaip tik dėl to atsirado daug anotavimo nesutapimų tarp lingvistų, kai lingvistai vertino kolokacijos konvencionalumą kalboje remdamiesi tik savo kalbos jausmu. Vėliau nuspręsta, kad anotuojant PJ rankomis bus daugiau remiamasi semantiniu (ne visiškai skaidomos reikšmės) požymiu. Dažnumo kriterijus formaliai taikomas nebuvo, nors ir galima būtų teigti, kad nusistovėjusios vartosenos kolokacijos iš esmės ir yra dažnai vartojamos kolokacijos.

Frazeologinis kolokacijos apibrėžimas šiam tyrimui pasirinktas dėl kelių priežasčių. Žodžių junginiai, turintys ir savarankišką, neskaidomą reikšmės elementą, yra svarbesni daugelyje kalbotyros sričių: vertimuose (nes nebūtinai pažodžiui sutampa abiejose kalbose), mokant(is) kitos kalbos (nes mokiniai dažnai klysta juos vartodami, labiau vargsta juos mokydami), kompiuterinės lingvistikos taikymuose. O junginiai, kurie pasižymi dažna vartosena, bet jų sandai išlaiko tiesioginę reikšmę, problemų dažniausiai nekelia. Be to, būtent reikšmės komponentą turinčios kolokacijos sunkiausiai atpažįstamos automatiškai, o dažnumu grįstos kolokacijos gali būti patikimai ir gana lengvai nustatomos be papildomų tyrimų – tiesiog apibrėžus aiškius kriterijus.

Kolokacijų skyrimo kriterijai

Nors kolokacijos samprata projekte buvo pasirinkta prieš rankomis anotuojant tekstyną, tačiau, kaip jau minėta, iš anotavimo rezultatų paaiškėjo, kad ji tebekelia problemų. Atsižvelgiant į anotavimo nesutapimų analizę, buvo ieškoma, kokiais tikslesniais kriterijais būtų galima įvertinti kolokacijoms ir frazeologizmams bendrus požymius. Buvo stengtasi aprašyti apibrėžtus ir patikrinamus kriterijus bei testus, kurie galėtų būti taikomi atpažįstant kolokacijas ne tik šiame, bet ir kituose tyrimuose. 2 lentelėje parodoma, kaip šie požymiai, kriterijai ir testai taikomi kolokacijoms⁹.

Dažniausiai kolokacijos yra skiriamos pagal kriterijus, kurie susiję su vartojimo pastovumo ir semantinio nedalomumo požymiais. **Vartojimo pastovumas** iš principo labai susijęs su kolokacijos dažnumu kalboje, tačiau, kaip jau minėta, be duomenų apie dažnumą tiriamajame tekстыne sunku įvertinti kiekvienos anotuojamos kolokacijos dažnumą. O vartojimo pastovumas gali būti įvertintas remiantis riboto leksinio junglumo kriterijumi, pritaikant pakeičiamumo panašios reikšmės žodžiu testą (plg. Nesselhauf, 2005): veiksmazodis arba daiktavardis yra ribotos reikšmės (autorė vartoja terminą *restricted sense*), vadinasi, sudaro kolokaciją, jeigu 1) ta reikšme gali būti vartojamas tik su tam tikrais žodžiais iš tos pačios reikšminės grupės arba 2) ta reikšme gali būti vartojamas ne su visais galimais žodžiais, kurie semantiškai ir sintaksiškai yra įmanomi (plačiau Nesselhauf, 2005, p.225). Pavyzdžiui, *read* reikšme „skaityti“ laikomas žodžiu, neturinčiu vartojimo apribojimų, nes laisvai gali būti derinamas su visais perskaitomais dalykais (*read a book, newspaper...*), ir taip sudaro laisvuosius junginius, o veiksmazodžio *deal* junglumas junginyje *deal a number* yra ribotas, nes ši *deal* reikšmė realizuojama tik su daiktavardžiu *number*, vadinasi, *deal a number* yra dideliu vartojimo pastovumu pasižyminti kolokacija (atitinka 1 kriterijų 2 lentelėje). Tarkime, anglų kalbos veiksmazodis *perform* gali būti vartojamas su panašių reikšmių daiktavardžiais *task, operation, duty, action*, bet anglų kalboje neįprasta, netinka sakyti: **perform a survey*. Vadinasi, nors veiksmazodis turi daugiau junglumo galimybių, bet iš panašios leksinės grupės žodžių jis įprastai vartojamas ne su visais (2 kriterijus).

Pakeičiamumo testu tikrinamas vartojimo pastovumas, kuris įprastai nustatomas statistiniais ar tikimybiniais metodais vertinant kolokacijų dėmenų traukos stiprumą. Svarbu pabrėžti, kad ne visoms kolokacijoms būdingas vienodas vartojimo pastovumas. Pastoviausiose kolokacijose vieno iš dėmenų leksinio junglumas yra labai stipriai ribotas ir tas žodis dažniausiai vartojamas tik su vienu ar keliais kitais žodžiais, pavyzdžiui, *bevizis režimas* (žodis *bevizis* dažniausiai vartojamas būtent su žodžiu *režimas*). Kitais atvejais žodis gali turėti daug sinonimiškų kolokacijų ir šitas riboto leksinio junglumo kriterijus bus mažiau išreikštas, pavyzdžiui, *kainos kyla / didėja / išaugo*. Taigi kolokacijos pastovumas susijęs su

⁹ Atskirti PJ požymius ir atpažinimo kriterijus jau siūlyta aprašant kitus PJ – morfologines samplaikas (žr. J. Kovalevskaitės disertacijos (2012) 4.1 poskyrį *Samplaikų požymiai ir jų raiška*).

2 lentelė

Kolokacijų skyrimo kriterijai

Požymiai	Kriterijai	Testai	Pavyzdžiai
VARTOJIMO PASTOVUMAS	Ribotas (leksinis) junglumas	Vieno kurio nors iš kolokacijos dėmenų pakeičiamumas (t. y. negalėjimas pakeisti) sinonimu ar panašios reikšmės žodžiu.	<i>pastaruuju metu – *praėjusiu metu;</i> <i>darė stebuklus – *gamino stebuklus;</i> <i>išvadas padarė – *išvadas sukūrė;</i> <i>laisvų darbo vietų – *tuščių darbo vietų.</i>
SEMANTINIS NEDALOMUMAS	Perprasminimas (perkeltinė reikšmė)	Vienas iš kolokacijos dėmenų vartojimas perkeltine reikšme (perkėlimas gali būti grįstas metonimija, metafora ar konceptualiąja metafora).	<i>(kažkelintoji) planetos raketė;</i> <i>apklausos rodo;</i> <i>stiprus vėjas;</i> į prisiminus panyra; įtemptas grafikas.
	Reikšmės vientisumas	Bendra junginio reikšmė. Kriterijus dažniausiai būdingas išblukusios reikšmės veiksmažodžių konstrukcijoms (junginiams, kuriuose semantinį svorį turi daiktavardis, o veiksmažodis yra praradęs reikšmę, pvz., <i>vykdyti, atlikti, (pa-/su-)daryti, turėti</i> , ir sudėtiniams pavadinimams.	Išblukusios reikšmės veiksmažodžių konstrukcijos: <i>turėjo galimybę – galėjo;</i> <i>priimti sprendimą – nuspręsti.</i> Sudėtiniai pavadinimai: <i>sveikatos apsaugos sistema;</i> <i>bevielis internetas;</i> <i>ligonių kasa.</i>
SUSTABARĖJIMAS	Morfologinis ir sintaksinis stabilumas 1. Didžiausias sustabarėjimo laipsnis (visi testai neigiami), kolokacijoms nebūdingas. 2. Mažesnis sustabarėjimo laipsnis (bent vienas iš testų teigiamas). 3. Mažiausias sustabarėjimo laipsnis (keli arba visi testai teigiami).	<ul style="list-style-type: none"> _ Gramatinių formų kaita; _ sintaksinės struktūros kaita; _ dėmenų sukeičiamumas; _ įsiterpimas, praleidimas. 	Mažesnio sustabarėjimo laipsnio junginiai: <ul style="list-style-type: none"> _ sudėtiniai pavadinimai (<i>vai- kų darželis; bevizis režimas</i>); _ daiktavardiniai junginiai (<i>gyva muzika; gyvenimo būdas</i>). Mažiausio sustabarėjimo laipsnio: <ul style="list-style-type: none"> _ veiksmažodiniai junginiai (<i>iškėlė klausimą; išvadas padarė</i>).

vieno iš žodžių junglumo apimtimi (kontekstinės paradigmos pločiu), o ta paradigma gali būti platesnė, siauresnė arba visai uždara, jeigu tam tikras kolokacijos žodis vartojamas tik vienoje kolokacijoje¹⁰.

Semantiniame nedalomumui įvertinti gali būti vartojami perprasminimo ir reikšmės vientisumo kriterijai. Semantinis nedalomumas bus nevienodas skirtingoms kolokacijoms. Išblukusios reikšmės veiksmažodžių konstrukcijos dažnai turės visiškai nedalomą reikšmę: *priimti sprendimą = nuspręsti*. Kituose junginiuose vienas iš dėmenų gali būti perprasminintas, o kitas vartojamas visiškai aiškia pirmąja reikšme (pvz., *iškelti bylą: iškelti* – perprasminintas, *byla* – tiesioginė reikšmė), tačiau junginio reikšmė vis tiek sudaroma iš dviejų žodžių reikšmių. Kai kuriais atvejais kolokacijos dėmenys įprastai vartojami kartu ir perprasminimą

10 Žinoma, norint vertinti kolokacijos pastovumą pakeičiamumo testu, svarbu apsvarstyti, kaip tikrinti tam tikro žodžio junglumo galimybes: rentis gimtakalbio kompetencija, teksto duomenimis, žodynuose fiksuojamomis konkretais žodžio kolokacijomis, informacija apie žodžių junglumą ir pan.

gali būti sunku suvokti. Pavyzdžiui, junginyje *paskelbė karą – skelbti* vartojamas tarsi pirmąją savo reikšmę „viešai garsinti, skleisti žinias“ (DLKŽ¹¹), tačiau ši reikšmė vis dėlto šiek tiek pakitusi, nes apie karą ne tik viešai pranešama, bet jis ir pradedamas. Kolokacijomis su vienu perprasmintu dėmeniu dažnai yra perteikiama metafora, plg. *viltis rusena, nerimas auga, puolė į neviltį, didelė laimė, mažas nesusipratimas, užkrečiamas juokas* (pavyzdžiai iš Marcinkevičienė, 2010).

Trečiasis PJ bendras požymis – **sustabarėjimas** – kolokacijoms neaktualiausias, ypač jei kolokacijos – daiktavardžio ir veiksmažodžio junginiai. Tarp tokių junginių dažni įsiterpimai (*iškėlė svarbų klausimą*), jų žodžiai gali būti keičiami vietomis (*pirmieji klausimą iškėlė darb-daviai*), junginys vartojamas įvairiomis formomis (*iškėlė klausimą / kelia klausimą / kels klausimą / kelti klausimą*). Kolokacijų formos gali varijuoti: kolokacijos žodžių tvarka gali kisti, tarp kolokacijos dėmenų gali įsiterpti kitų žodžių. Pavyzdžiui, jei skiriame kolokaciją *sėdėjo prie vairo*, ta pati kolokacija bus išlaikyta ir junginyje *prie kurio vairo sėdėjo*. Lygiai taip pat kolokacija *pateisinti viltis* ir jos variantas *pateisino visas 1993 m. į jį sudėtas viltis*.

Nors tarp kolokacijų yra ir tokių, kurios vartojamos tik viena gramatine forma (pvz., *burtų keliu, išimties tvarka*), vis dėlto tokie atvejai netipiški, todėl galima sakyti, kad didelė dalis kolokacijų bus mažesnio arba mažiausio sustabarėjimo laipsnio. Iš LKDFŽ dvižodžių junginių gramatikos tyrimų (Boizou et al., 2015) matyti, kad prijungiamojo ryšio požymio raiškos junginiai yra dažniausi, sudaryti daugiausia iš dviejų daiktavardžių, rečiau – iš būdvardžio ir daiktavardžio arba veiksmažodžio ir daiktavardžio. Tiriant dvižodžius junginius pagal dėmenų kintamumą, nustatyta, kad yra beveik apylygiai tokių junginių, kurių abu dėmenys keičiasi (sujungiamojo ryšio ir prijungiamojo ryšio požymio), ir tokių, kurių pirmas dėmuo yra pastovus (prijungiamojo ryšio požymio ir prijungiamojo ryšio aplinkybiniai junginiai).

Apibendrinant svarbu pabrėžti, kad žodžių junginiai, kurie yra semantiškai dalomi ir nevartojami pastoviai, nelaikytini kolokacijomis. Tačiau junginiai, laikomi kolokacijomis, gali varijuoti pagal semantinį nedalomumą ar pastovų vartojimą. Kadangi kolokacijos atpažįstamos pirmiausia iš pastovios vartosenos ir pagal semantinio nedalomumo požymį, net tais atvejais, kai tarp junginio dėmenų įsiterpia kitų žodžių, savo semantiką ir pastovumą jis išlaiko. Jei anotuojama rankomis, tokie įsiterpimai taikant atpažinimo kriterijus ir nustatant kolokacijas problemų nekelia, bet, siekiant automatiškai aptikti kolokacijas su įsiterpimais, kyla sunkumų.

Lietuvių kalbotyroje įsigalėjusi tradicinė frazeologijos samprata apima keletą požymių, be kurių žodžių junginys negali vadintis frazeologizmu. R. Marcinkevičienė (2010, p.117), apibendrinama keletu tyrėjų išvadas, skiria šiuos frazeologizmų požymius: raiškos atžvilgiu frazeologizmai suprantami kaip sustabarėję, stabilūs, pastovios leksinės sudėties ir gramatinės struktūros junginiai, atkuriami iš atminties, imami gatavi, o ne kuriami kiekvieną kartą iš naujo, todėl vartojami kaip nusistovėję, įprastiniai kalbos vienetai; semantiniu aspektu tai – vientisi, apibendrintos reikšmės, nusakomi vienu žodžiu; vartosenos atžvilgiu – turintys apibrėžtą leksinę aplinką, vadinamąją frazeologinę distribuciją, ir stilistiniu aspektu – kaip vaizdingi junginiai¹².

Labai svarbus frazeologizmų požymis – reikšmės vientisumas arba semantinis nedalomu-

11 Naudotasi elektroniniu *Dabartinės lietuvių kalbos žodyno* variantu (DLKŽ).

12 Dalis *Frazeologijos žodyne* (2015) nurodomų frazeologizmų šiame tyrime frazeologizmais nelaikomi, nes priskiriami morfologinėms samplaikoms (pvz., *kažin kas, kas kita, kur kas*, pvz.: *Manai, tavo kepurė didelė, tai tu kažin kas* (reikšmė „didelis, svarbus dalykas“). *Dabar bus visai kas kita* (reikšmė „visai kitaip“). *Aš už jį kur kas didesnis* (reikšmė „žymiai, daug“). Šios samplaikos daugelio tyrėjų darbuose suprantamos kaip frazeologizmai tada, jei turi perkeltinę reikšmę, tačiau tokių atvejų analizuojamame tekste beveik nepasitaikė.

mas. Tai reiškia, kad frazeologizmo reikšmė nėra sudaryta iš skirtingų jį sudarančių žodžių reikšmių, o yra atskira (nekompoziciška (Lamiroy, 2016, p.136) ir perteikiama viena sąvoka, be to, dažniausiai gali būti apibūdinama vienu žodžiu (Jakaitienė, 2010, p.280; Paulauskas, 2003, p.3). Semantiškai frazeologizmas yra perprasmintas junginys ir vartojamas perkeltine reikšme (pvz., *traukti per dantį* – šaipytis, *varnas šaudyti* – žiopsoti). Todėl, matyt, dėl perkeltinės reikšmės išryškėja posakio vaizdingumas, kuris taip pat laikomas vienu iš frazeologizmo požymių. Vis dėlto, kaip mini ir R. Marcinkevičienė (2010, p.117), tai labiau subjektyvus požymis, kurį sunku patikrinti objektyviais metodais.

Kolokacijos gali šiek tiek varijuoti leksiškai, bet pabrėžiama, kad frazeologizmams leksinė variacija nebūdinga. Tai turėtų būti visiškai riboto junglumo junginiai, tiksliai atkuriami iš atminties. Daugelis tokie ir yra. Vis dėlto analizuojant tekstyną ir remiantis kitais tyrimais (Butkutė, 2010), matyti, kad neretai frazeologizmai į tekstą įtraukiami modifikuoti, nors ir išlieka atpažįstami kaip tokie, pvz., *prieina [savotišką] kryžkelę*.

Reikšmės vientisumą lemia tai, kad pats frazeologizmas yra sustabarėjęs morfologiškai ir sintaksiškai. Sustabarėjimas laikomas vienu iš pagrindinių frazeologizmų požymių ir padeda identifikuoti frazeologizmus tekste. Įprasta, kad frazeologizmai turi morfolominių ir sintaksiinių apribojimų, jų formos mažai varijuoja skaičiaus, asmens, laiko, struktūros požiūriu. Kita vertus, šiems junginiams neretai būdinga gramatinė kaita. Kadangi lietuvių kalba fleksinė, tai kinta veiksmažodžių galūnės, t. y. jie kaitomi, gali kisti dėmenų tvarka, galiausiai įsiterpti kitų žodžių ir pan. Galima teigti, kad struktūriškai dauguma frazeologizmų atrodo kaip laisvieji junginiai, tik skiriasi savo semantika arba kontekstine aplinka.

Leksinės ir gramatinės variacijos rodo, kad frazeologiją kaip reiškinį gana sudėtinga apibrėžti, nes patys frazeologizmai pagal jiems būdingus požymius gali būti kelių lygmenų, ne visada prototipiški, juo labiau, kad ir laisvieji žodžių junginiai gali turėti junglumo ir gramatinių apribojimų, todėl anksčiau minėtos ypatybės nėra išskirtinai būdingos tik frazeologizmams (plg. Lamiroy, 2016, p.137). Kita problema, kaip parodė tyrimas, ta, kad frazeologizmus nelengva atskirti nuo kolokacijų, ypač kai vienas kolokacijos dėmuo yra perprasmintas (žr. 2 lent.) arba junginys yra kolokacijų ir frazeologizmų paribyje (žr. 2 pav.).

Frazeologizmai priklauso sunkiau automatiškai atpažįstamų PJ grupei (Ramisch, 2014). Kai kuriuos galima atpažinti, jeigu jie pažeidžia įprastas gramatikos taisykles arba yra neįprasti leksikos požiūriu (*nešti kudašius, po paraliais*). Statistiniais metodais frazeologizmų taip pat atskirti beveik neįmanoma, juo labiau, kad jie nėra labai dažni tekste. Taigi atsižvelgiant į šiuos ir minėtus sunkumus, su kuriais susidūrė PJ anotavę lingvistai, galima išskirti keletą pagrindinių požymių bei kriterijų, kuriais vadovaujantis tyrėjui lengviau atpažinti frazeologizmus (žr. 3 lent.).

Remiantis trimis pagrindiniais požymiais: vartojimo pastovumu, semantiniu nedalomumu ir sustabarėjimu, frazeologizmams atpažinti galima taikyti objektyvius kriterijus ir tam tikrus testus.

Laikoma, kad, pavyzdžiui, **vartojimo pastovumą** įrodo riboto junglumo kriterijus, jį galima patikrinti atlikus pakeičiamumo sinonimu testą: jei testas neigiamas, t. y. junginio dėmuo negali būti pakeičiamas sinonimu ar artimos reikšmės žodžiu, vadinasi, pagal šį požymį junginį galima laikyti frazeologizmu. Žinoma, svarbu nepamiršti, kad frazeologizmams taip pat gali būti būdingas leksinis varijavimas: *iki grabo / kapo lentos; imti vairą / vadžias į rankas; nešti kailį / kudašius / muilą*.

Be to, natūraliosios kalbos analizės srityje dažnai akcentuojamas PJ, ypač frazeologizmams, būdingas nereguliarumas, kai, pavyzdžiui, vienas iš frazeologizmo dėmenų yra morfologiš-

3 lentelė
Frazeologizmų
skyrimo kriterijai

Požymiai	Kriterijai	Testai	Pavyzdžiai
VARTOJIMO PASTOVUMAS	Ribotas (leksinis) junglumas	Pakeičiamumas sinonimu ar panašios reikšmės žodžiu (t. y. negalima pakeisti).	<i>Pūsti miglą – *pūsti rūką; varnas šaudyti – *šarkas šaudyti.</i>
SEMANTINIS NEDALOMUMAS	Perprasminimas (perkeltinė reikšmė) 1.1. Grynosios idiomos 1.2. Perprasmintos idiomos	Visas junginys perkeltinės reikšmės.	1.1. <i>per dantį traukė</i> – šaipėsi, išjuokė; 1.2. <i>Vynioja aplink pirštą</i> – apgaulinėja; <i>juodas darbas</i> – nusikaltimas, nuostolinga veikla
	Reikšmės vientisumas	Bendra junginio reikšmė.	<i>įvartyti į kampą</i> – prisipirti
	Frazeologinė aplinka	Kontekstas, kuris rodo perkeltinę junginio reikšmę.	<i>Nusiimti kaukes, plg.: Gal tai vieta, kur jaučiamės saugūs, nu- siimame kaukes? Pailsę karnavalo dalyviai nusiima kaukes ir imasi kasdienių darbų</i>
SUSTABARĖJIMAS	Morfologinis ir sintaksinis stabilumas: 2.1. Didžiausio sustabarėjimo laipsnio (visi testai neigiami) 2.2. Mažesnio sustabarėjimo laipsnio (bent vienas iš testų teigiamas) 2.3. Mažiausio sustabarėjimo laipsnio (keli arba visi testai teigiami).	– Gramatinių formų kaita; – sintaksinės struktūros kaita; – dėmenų sukeičiamu- mas; – įsiterpimas, praleidimas.	2.1. <i>toli gražu – *toliausiai gražu; gerai arba nieko;</i> 2.2. <i>šauna į galvą, į galvą šauna</i> 2.3. <i>Pastatyti [ji] ant kojų; prieina [savotišką] kryžkele; iškopė į [populiarumo] aukštumas; daro savo [...]; kovoti iki paskutinio [...].</i>

kai nekaitomas, plg. *turėti omenyje*, po šimts (angl. *cranberry words*) (Trawiński et al., 2008). Svarbu tai, kad, taikant frazeologizmų kriterijus ir testus, išryškėja skirtingi kiekvieno požymio laipsniai. Jie ryškiausiai atsiskleidžia nustatant semantinį frazeologizmo nedalomumą ir sustabarėjimą.

Semantinio nedalomumo lygmenys skiriami atsižvelgiant į perprasminimo laipsnį. Pagal tai skiriamos:

1 grynosios idiomos (angl. *pure idioms* (Cowie, 1998) arba lietuvių kalbotyroje įprastesnis *nemotyvuotųjų frazeologizmų* terminas, kai frazeologizmo reikšmė išryškėja iš viso junginio, kuris yra perprasmintas, o ne iš atskirų jo dėmenų. Junginiai semantiškai neįprasti, ypač negimtakalbiui, pvz.: *varnas šaudyti* – žiopoti, *atrišo liežuvį* – daug kalbėjo, *per dantį traukė* – šaipėsi, išjuokė.

2 perprasmintos idiomos (angl. *figurative idioms* (Cowie, 1998)), dar vadinamos *motyvuotaisiais frazeologizmais*¹³, kurių dėmenys neparadę semantinio savarankiškumo, bet

13 E. Jakaitienė (2010) motyvuotuosius skirsto į: 1) tropinius frazeologizmus: a) metoniminius: šaukštą padėti, peržengti slenkstį, gauti į skudurus; b) metaforinius: *ranka numoti, suimti už gerklės, tupėti po sparneliu*; 2) lyginamuosius frazeologizmus (turi palyginimo formą): *sėdėjo lyg ant adatų, išbalo kaip drobė, išėjo kaip stovi*.

visas junginys yra perprasmingas ir turi atskirą reikšmę, pvz.: *vynioja aplink pirštą* – apgaudinėja, *vaikščioja peilio ašmenimis* – rizikuoja, *suvesti sąskaitas* – atsilyginti už ką, keršyti, *įvaryti į kampą* – prispirti, šalta širdis – nejautrumas, *duoti grąžos* – atsilyginti už pikta, *stoti prie vairo* – perimti veiklą, užkrauti ant pečių – perduoti atsakomybę. Dažniausiai visas junginys yra perkeltinės reikšmės. Nors ir labai problemiškas dalykas, tačiau atskiri frazeologizmo dėmenys gali būti suvokiami kaip turintys perkeltines reikšmes: pvz., Gibbs (1993, pp.62–63) analizuoja anglišką pavyzdį *spill the beans* (išduoti paslaptį), kuriame *spill* atitinka *išduoti*, o *beans* – *paslaptį*. Tokie pavyzdžiai įrodo, kad ir frazeologizmuose atsispindi, kaip konceptualiuoju lygmeniu suvokiama nekalbinė tikrovė pasitelkiant metaforas.

Sustabarėjimo laipsniai išryškėja atsižvelgiant į sustabarėjimo stiprumą. Jis gali būti tikrinamas taikant kelis testus: ar kinta gramatinės formos; ar kinta sintaksinė struktūra; ar dėmenys gali būti sukeičiami vietomis; ar būdingas įsiterpimas ir praleidimas (žr. 3 lent.). Pagal testų atsakymus TAIP / NE (žr. 3 lent. pavyzdžius), galima spręsti apie junginio sustabarėjimo laipsnį. Taigi frazeologizmai skiriami pagal tris sustabarėjimo laipsnius:

1 didžiausio sustabarėjimo laipsnio (visi testai neigiami): frazeologizmas vartojamas tik tam tikra būdinga forma, kuri nevarijuoja, pvz., *po paraliais, eik po velnių, viso labo, toli gražu, gerai arba nieko, va bank*; gali būti tautologija, pvz.: *nori nenori* (daugiau tokių vienafornų frazeologizmų ir kitų PJ pavyzdžių galima rasti Rimkutė et al., 2015);

2 mažesnio sustabarėjimo laipsnio (bent vienas iš testų teigiamas): pvz., gali kisti vieno kurio nors žodžio forma ir žodžių tvarka, pvz., *šauna į galvą, į galvą šovė*, bet frazeologizmai gana tiksliai atkuriami iš atminties, pvz.: *ieškoti laimės – laimės ieškoti, kibti į gerklę – įsikibti į gerklę*;

3 mažiausio sustabarėjimo laipsnio – keli arba visi testai teigiami. Tai reiškia, kad pačiam frazeologizmui būdingas didelis gramatinio variantiškumo (Schmitt, 2005) laipsnis – visų ar kelių junginio žodžių gramatinė forma kinta, keičiasi junginio žodžių tvarka, į junginį gali įsitempti kitų žodžių ir pan. Šiuo atveju frazeologizmai gali būti perfrazuojami, papildomi arba modifikuojami¹⁴, bet frazeologizmo pagrindas išlaikomas. Dėl šios priežasties frazeologizmus sunkiau identifikuoti tekste tiek rankomis, tiek ir automatiškai. Vienais atvejais frazeologizmas atkuriamas ne visos struktūros, pvz., *užpilti druskos* (praleista *ant žaizdos*); *kovosiu iki paskutinio* (praleista *kraujo lašo*). Tačiau daug dažniau būdingas priešingas reiškinys – kai įterpiama daugiau žodžių: *tvirtai ant (abiejų) kojų stovinčio; iškopė į (populiarumo) aukštumas; pakilo (vienu) laipteliu aukščiau; laikyti vairą (savo) (rankose)* ir pan.

Kitų kalbų tyrimai įrodo (Abeillé, 1995; Moon, 1998a; Moon, 1998b; Fellbaum, 2007), kad frazeologizmai toli gražu nėra visiškai sustabarėję PJ. Frazeologizmų variantiškumas nėra retas reiškinys ir tarp lietuvių kalbos frazeologizmų: remiantis G. Virbickienės tyrimu (2017), net pusė iš tirtų 427 frazeologizmų pasižymi gramatiniu variantiškumu, iš jų daugiau negu pusė pasižymėjo formos kaita, o trečdaliui buvo būdinga ir formos, ir žodžių tvarkos kaita. Variantiškiems frazeologizmams mažiau būdingas kitų žodžių įsiterpimas ir kintama žodžių tvarka. Dėl tokio gramatinio variavimo išryškėja jau anksčiau minėtas probleminis aspektas – frazeologizmų ribų klausimas, dėl kurio aptikti frazeologizmus sunkiau, nes iki galo neaišku, kokios struktūros jie turėtų būti, pvz., *(nelaukia, nesėdi) sudėjęs rankų; (paliko) likimo valiai; (...) vienose rankose*.

Anotuojant ir atpažįstant tekste frazeologizmus, taip pat labai svarbus kontekstas. Vieni frazeologizmai aiškūs ir be konteksto (pvz., *atrišo liežuvį*), tačiau daugelis jų tampa frazeologizmais tik tam tikrame kontekste, tik jiems būdingoje frazeologinėje aplinkoje, nes be konteksto jie gali būti palaikyti tiesioginės reikšmės junginiais, pvz., *Daryti [drąsų] posūkį*

¹⁴ Frazeologizmų modifikavimo būdus savo disertacijoje yra plačiai analizavusi L. Butkutė (2010).

politikoje (plg. *Daryti staigų posūkį į dešinę*). Ekipos vairą perima treneris Tony Pulisas (plg. *Kitas vairuotojas perėmė vairą likus valandai iki avarijos*). Frazeologizmo kontekstas paprastai signalizuoja apie perkeltinę junginio reikšmę.

Kolokacijų ir frazeologizmų bendrumai bei skirtumai

Kolokacijų ir frazeologizmų bendrumai bei skirtumai

Šiame poskyryje apibendriname kolokacijų ir frazeologizmų bendrumus bei skirtumus, parodydami jų santykį su laisvaisiais junginiais ir įvertindami trijų 2 ir 3 lentelėse analizuotų požymių būdingumą.

Pasak F. J. Hausmanno (cit. iš Nesselhauf, 2004, p.16), atskirti kolokacijas nuo laisvųjų junginių galima pagal kolokacijoms būdingą ribotą leksinį junglumą (angl. *restricted combinability*); atskirti kolokacijas nuo frazeologizmų – pagal tai, kad kolokacijos yra semantiškai dalomos (angl. *transparent*), o frazeologizmai – ne. Vieno iš įtakingiausių frazeologinio požiūrio atstovų A. Cowie (1998) pateikiamoje PJ skalėje pagal semantinio dalomumo ir vartojimo pastovumo laipsnį išdėstomos abi PJ grupės – kolokacijos (angl. *restricted collocations*) ir du frazeologizmų tipai, besiskiriantys perprasminimo laipsniu: motyvuotieji frazeologizmai (angl. *figurative idioms*) ir nemotyvuotieji frazeologizmai (angl. *pure idioms*) (cit. iš Granger et al., 2008, p.36). Kolokacijos labiausiai priartėja prie laisvųjų junginių (žr. 3 pav., plg. ir 2 pav.):

3 pav.

Laisvųjų junginių,
kolokacijų ir
frazeologizmų sąsajos

laisvieji junginiai || kolokacijos → motyvuotieji frazeologizmai → nemotyvuotieji frazeologizmai

Nors A. Cowie (1998) klasifikacijoje tik frazeologizmų klasė yra skaidoma į tipus, tačiau kituose frazeologinio požiūrio atstovų darbuose, pvz., I. Mel'čuko, kolokacijos skirstomos į 4 tipus (cit. iš Nesselhauf, 2004, p.12): 1) kolokacijos su išblukusios reikšmės veiksmažodžiu (*take a step*); 2) kolokacijos, kurių vieno dėmens perkeltinė reikšmė realizuojama tik konkrečioje kolokacijoje (*black coffee*, *black* reikšmė „be cukraus“); 3) kolokacijos, kurių ta pati perkeltinė reikšmė realizuojama ir kitose kolokacijose (*strong coffee*, *strong argument*); 4) kolokacijos, kurių vieno dėmens semantiniai požymiai pasikartoja ir kito dėmens reikšmėje (*the horse neighs* – arklio žvengimas). Vadinasi, nevienodą semantinio dalomumo laipsnį atspindinčiose klasifikacijose patvirtinama, kad ir kolokacijų, ir frazeologizmų klasės nėra vienalytės.

Remdamasi savo ir kitų mokslininkų tyrimais, B. Lamiroy (2016, p.139) akcentuoja, kad PJ sustabarėjimas bei nedalomumas skirtingose PJ klasėse yra nevienodas, ir tai viena iš priežasčių, kodėl juos sunku vienareikšmiškai apibrėžti. Ir kolokacijų, ir frazeologizmų klasėse B. Lamiroy (2016) skiria tipiškiausius, ne tokius tipiškus ir netipiškiausius vienetus: pavyzdžiui, kolokacijų klasėje tipiškiausiais laikomi tokie junginiai, kurių abu žodžiai perteikia savo reikšmes, o netipiškiausiomis kolokacijomis laikomos kolokacijos su perkeltinės reikšmės sandu (Lamiroy, 2016, p.142), nes tokios kolokacijos priartėja prie frazeologizmų¹⁵. Ne visai tipiškos yra ir kolokacijos su išblukusios reikšmės veiksmažodžiu (pvz., *atlikti tyrimą – tirti*), nes tokių kolokacijų reikšmė – vientisa, kaip būdinga frazeologizmams. Tipiškiausiams

15 Vis dėlto diskutuotina, ar kolokacijos su perkeltinės reikšmės sandu turėtų būti laikomos netipiškomis, nes 1) tokios kolokacijos yra metaforinės ir sudaro didelę kolokacijų dalį, be to, 2) tokiose kolokacijose abu žodžiai perteikia savo reikšmes, todėl esame linkusios priskirti tokias kolokacijas tipiškoms, o netipiškomis laikyti tik kolokacijas su išblukusios reikšmės veiksmažodžiu.

frazeologizmams būdinga nedaloma reikšmė, vadinasi, motyvuotieji frazeologizmai laikytini ne tokiais tipiškais nei nemotyvuotieji.

Galima pritarti kitų autorių pastebėjimams, kad kolokacijos ir frazeologizmai susiję atvirkštinio ryšiu (plg. Lamiroy, 2016, p.143): tipišku frazeologizmų reikšmė yra viso junginio reikšmė (junginys semantiškai nedalomas), o tipišku kolokacijų reikšmė susideda iš atskirų kolokaciją sudarančių žodžių reikšmių (junginys semantiškai dalomas). Dėl šio aspekto kolokacijos dažnai suvokiamos kaip esančios pusiaukelėje tarp laisvųjų žodžių junginių ir frazeologizmų.

Riba tarp frazeologizmų ir kolokacijų dažnai ne iki galo aiški būtent dėl to, kad kartais sunku atskirti frazeologizmus nuo kolokacijų, turinčių perkeltinės reikšmės sandą (Lamiroy, 2016, p.143), metaforas perteikiančių kolokacijų (pvz., *perkratyti atmintį*). Vis dėlto vienas iš aiškiausių skiriamųjų kriterijų tas, kad kolokacijos sudarytos iš sandų, kurių bent vienas turi išlaikyti savo prasmę savarankiškumą, o frazeologizmų atveju įvyksta viso junginio perprasminimas, visas junginys įgyja naują reikšmę ir dėl to frazeologizmo reikšmė tampa visiškai neskaidoma. Kaip aiškina B. Lamiroy (2016, p.143), prancūzų kalbos kolokacijos *peur bleue* (klaiki baimė) būdvardis *bleue* vis dėlto nurodo baimės pobūdį, o frazeologizmo *cordons bleus* (puikus virėjas) dėmenys nesusiję jokių semantinių ryšių nei su kaspinu (*cordons*), nei su kuo nors mėlynu (*bleus*). Vadinasi, akcentas, kad kolokacijos reikšmė yra sudaryta iš visų ją sudarančių žodžių reikšmių (Granger et al., 2008; Bergenholtz et al., 2013), yra esminis.

Analizuojant PJ vartojimo pastovumą ir sustabarėjimą, aiškėja, kad šių dviejų požymių atžvilgiu atvirkštinis ryšys tarp kolokacijų ir frazeologizmų klasių nėra toks ryškus kaip semantinio nedalomumo. Kolokacijų ir frazeologizmų klasė nėra vienalytė sustabarėjimo bei vartojimo pastovumo požiūriu, o sustabarėjimas, kaip matyti taikant testus, gali būti kelių lygmenų. Vis dėlto teiginys, kad frazeologizmai pasižymi didesniu sustabarėjimu nei kolokacijos, ne visai atspindi vartosenos duomenis, nes tyrimai rodo (Virbickienė, 2017), kad beveik pusei frazeologizmų būdingas gramatinis variantiškumas. Kita vertus, šie duomenys nepaneigia to, kad, priklausomai nuo kolokacijos struktūros (pvz., su kilmininku ar be jo), kolokacijos daugiau kaitomos – jų gramatinė forma labiau varijuoja. Dėl formos kintamumo, ypač galimybės įsiterpti, keisti dėmenis vietomis kolokacijos panašesnės į laisvuosius junginius.

Tipiškomis kolokacijoms galima priskirti kolokacijas su perkeltinės reikšmės dėmeniu (*stiprus vėjas*), sudėtinius pavadinimus (*vaikų darželis*), o netipiškoms kolokacijoms – junginius su išblukusios reikšmės veiksmažodžiu (*atlikti tyrimą*). Tipiškais frazeologizmais galima laikyti nemotyvuotuosius frazeologizmus (*per dantį traukia*), o motyvuotieji (*nusiimti kaukes*) būtų ne tokie tipiški. Kaip matyti iš 4 lentelės, netipiškos kolokacijos pagal semantinio nedalomumo laipsnį priartėja prie tipišku frazeologizmų, o netipiški frazeologizmai, kadangi jų semantinio nedalomumo laipsnis mažesnis, atitolsta nuo prototipiškų frazeologizmų.

Pagal kitus du požymius – vartojimo pastovumą ir sustabarėjimą – tipiškos ir netipiškos kolokacijų bei frazeologizmų grupės atsiskiria ne taip aiškiai: taip yra dėl to, kad šių ypatybių

	Vartojimo pastovumas	Semantinis nedalomumas	Sustabarėjimas
Tipiškos kolokacijos			
Netipiškos kolokacijos			
Tipiški frazeologizmai			
Netipiški frazeologizmai			

4 lentelė

(Ne)tipišku kolokacijų ir frazeologizmų požymiai¹⁶

16 Šviesiai pilku fonu pažymėtas mažesnis tam tikro požymio laipsnis, o tamsiai pilku – didesnis.

raiška toje pačioje klasėje gali būti skirtingo laipsnio, atsižvelgiant į konkretaus junginio leksinę ir gramatinę sandarą. Vienais atvejais tipiškų kolokacijų vartojimo pastovumo požymis gali būti ryškesnis, kitais – ne toks ryškus, pvz., jeigu vieno iš dėmenų junglumas ribotas, tada ši kolokacija „stipresnė“, jeigu vieno iš dėmenų junglumo apimtis didesnė, tada kolokacija „silpnesnė“.

Palyginti su kolokacijomis, didesnis vartojimo pastovumo laipsnis būdingesnis ir tipiškiems, ir netipiškiems frazeologizmams, o sustabarėjimo laipsnis, kaip ir kolokacijų atveju, labiau priklauso nuo junginio sandaros. Kitaip sakant, vienas tipiškas frazeologizmas gali pasižymėti didesniu sustabarėjimu negu kitas tipiškas frazeologizmas. Kolokacijų sustabarėjimo požymio intensyvumas taip pat susijęs su kolokacijų struktūra: netipiškų kolokacijų su išblukusios reikšmės veiksmažodžiu antrasis dėmuo dėl valdymo ryšio išlieka stabilus (plg. *daryti įtaką, atlikti tyrimą*), tačiau dėmenys keičiami vietomis, galimi įsiterpimai, todėl sustabarėjimas – mažesnis. Tarp tipiškų kolokacijų daug PJ, kur abu dėmenys kaitomi (pvz., būdvardžio ir daiktavardžio junginiai), tačiau nemažai ir sudėtinių pavadinimų, o jų pirmasis dėmuo dažnai yra pažymins, reiškiamas daiktavardžio kilmininku (plg. *vaikų darželis, ligo-nių kasa*), todėl, atspindint abiejų šių grupių ypatybes, nurodyta, kad tipiškų kolokacijų sustabarėjimas mažesnis. Ši pastaba veda prie tokios minties: nors sustabarėjimas laikytinas abiem grupėms būdingu požymiu, nes įrodo, kad ir kolokacijos, ir frazeologizmai – daugiau ar mažiau sustabarėję PJ, bet siekiant atskirti vieną grupę nuo kitos sustabarėjimas nėra toks esminis požymis kaip vartojimo pastovumas ir semantinis nedalomumas. Taigi semantinis nedalomumas yra esminis požymis kolokacijoms nuo frazeologizmų atskirti, o vartojamo pastovumas svarbesnis kolokacijoms nuo laisvųjų žodžių junginių atskirti.

Išvados

Šiame straipsnyje analizuoti kolokacijų ir frazeologizmų nustatymo kriterijai, remtasi trimis abiem grupėms bendrais požymiais: vartojimo pastovumu, semantiniu nedalomumu ir sustabarėjimu. Visų šių požymių laipsnis gali būti nustatomas pagal kriterijus ir testus. Taikant testus, galima aiškiau identifikuoti esminius kolokacijų bei frazeologizmų skirtumus ir objektyviau nuspręsti, į kurią grupę skirti kurį pastovųjį junginį, be to, galima matyti, kiek tam tikras požymis yra būdingas kolokacijų klasei ir frazeologizmų klasei priskiriamams junginiams.

Ir kolokacijų, ir frazeologizmų klasės nėra vienalytės: abiem atvejais yra tipiškų ir ne tokių tipiškų junginių. Kolokacijomis laikytini dvižodžiai PJ, kuriuose vieno iš žodžių junglumas yra ribotas; įprastai šie PJ yra semantiškai dalomi (arba abu žodžiai vartojami tiesioginėmis reikšmėmis, arba vienas iš žodžių perkeltinės reikšmės), nors yra ir nedalomų (kolokacijos su išblukusios reikšmės veiksmažodžiais). Frazeologizmų, kurie įprastai yra ilgesni nei dvižodžiai junginiai, viso junginio reikšmė yra perkeltinė, tačiau semantinio nedalomumo laipsnis gali būti mažesnis (nemotyvuotųjų) ir didesnis (motyvuotųjų). Ir kolokacijų, ir frazeologizmų sustabarėjimo laipsnis priklauso nuo junginio sandaros.

Iš trijų analizuotų požymių kaip esminius galima būtų išskirti vartojimo pastovumą ir semantinį nedalomumą. Semantinis nedalomumas yra esminis požymis kolokacijoms nuo frazeologizmų atskirti, o vartojamo pastovumas svarbesnis kolokacijoms nuo laisvųjų žodžių junginių atskirti. Kriterijai ir testai, taikomi esminiams požymiams įvertinti, gali padėti apsispręsti dėl to, ar PJ laikytinas kolokacija ar frazeologizmu. Sustabarėjimą tikrinantys testai daugiau parodo kolokacijos ar frazeologizmo formos lankstumą, tačiau nelaiiktini būtinai siekiant atskirti frazeologizmus nuo kolokacijų.

Straipsnyje aprašomi kriterijai lingvistams, tačiau būtų svarbu automatizuoti kriterijais pagrįstų testų taikymą: tai padėtų gauti tikslesnius automatinio PJ atpažinimo rezultatus, t. y.

tiksliu atskirti kolokacijas nuo frazeologizmų. Remiantis kitoms kalboms aprašytu įdirbiu (Baldwin et al., 2010), aktualu toliau eksperimentuoti su lietuvių kalbos kolokacijų ir frazeologizmų atpažinimo kriterijais: kaip automatizuotai patikrinti žodžių junglumo apimtį, kaip įvertinti, kiek tam tikram PJ yra būdingas semantinis nedalomumas. Tikėtina, kad tie patys kriterijai galėtų būti vienokiais testais tikrinami žmogaus, kitokiais – automatiškai, taip vienas kitą papildydami ir sumažindami subjektyvumo faktorių.

Literatūra

1. Abeillé, A., 1995. The Flexibility of French Idioms: A Representation with Lexicalized Tree Adjoining Grammar. *Idioms: Structural and Psychological Perspectives* (eds.) M. Everaert, E. van der Linden, A. Schenk, R. Schreuder, Hillsdale, NJ: Lawrence Erlbaum Associates, pp.15–42.
2. Baldwin, T., Kim, S. N., 2010. Multiword Expressions. *Handbook of Natural Language Processing* (eds.) N. Indurkha and F. J. Damerau, CRC Press, Boca Raton, USA, pp.267–292.
3. Barnbrook, G., Mason, O., Krishnamurthy, R., 2013. Collocation: Applications and Implications, Palgrave: Macmillan. <https://doi.org/10.1057/9781137297242>
4. Bergenholtz, H., Gouws, R., 2013. A Lexicographical Perspective on the Classification of MWE. *International Journal of Lexicography*, vol. 27, no. 1, Oxford University Press, pp.1–24.
5. Biber, D., Johansson, S., Leech, G. et al., 1999. *Longman Grammar of Spoken and Written English*. Harlow, England: Longman.
6. Boizou, L., Kovalevskaitė, J., Rimkutė E., 2015. Lietuvių kalbos dvižodžių junginių morfologinių ir sintaksinių ypatybių sąsajos. *Darbai ir dienos* 64, pp.115–131.
7. Brown, J., 2014. Knowledge of Collocations. *Dimensions of Vocabulary Knowledge*. (eds.) J. Milton & T. Fitzpatrick. Basingstoke: Palgrave Macmillan, pp.123–139. https://doi.org/10.1007/978-1-137-36831-7_9
8. Butkutė, L., 2010. Okazionaliųjų frazeologizmų stilistinės išgalės publicistiniame stiliuje. *Daktaro disertacija*. Vilnius: Vilniaus universitetas.
9. Church, K. W., & Hanks, P., 1990. Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, no. 16 (1), pp.22–29.
10. Cowie, A. P., 1998. *Phraseology: Theory, Analysis, and Applications*. Oxford University Press.
11. Durrant, P., 2009. Investigating the Viability of a Collocation List for Students of English for Academic Purposes. *English for Specific Purposes*, no. 28 (3), pp.157–169.
12. Ermanytė, I., Kažukauskaitė, O., Naktinienė, G. et al. (reng.), 2015. *Frazeologijos žodynas*. Vilnius: Lietuvių kalbos institutas. Prieiga internete: <http://lkiis.lki.lt/> [žiūrėta 2017-09-23].
13. Fellbaum, Ch. (ed.), 2007. *Collocations and Idioms: Corpus-Based Linguistic and Lexicographic Studies*. Continuum Press, Birmingham, UK.
14. Firth, J. R., 1957. *A Synopsis of Linguistic Theory 1930–1955*. Selected Papers of Firth 1952–1959 (ed.) F. Palmer. London: Longman, pp.168–205.
15. Gibbs, R. W., 1993. Why Idioms are not Dead Metaphors. *Idioms: Processing, Structure, and Interpretation* (eds.) C. Cacciari, P. Tabossi. Hillsdale, NJ: Lawrence Erlbaum Associates, pp.57–77.
16. Granger, S., & Paquot, M., 2008. *Disentangling the Phraseological Web*. *Phraseology: An Interdisciplinary Perspective* (eds.) S. Granger & F. Meunier. Amsterdam; Philadelphia: John Benjamins Publishing, pp.27–46. <https://doi.org/10.1075/z.139.07gra>
17. Gries, S. T., 2013. 50-something Years of Work on Collocations: What is or should be next... *International Journal of Corpus Linguistics*, no. 18 (1), John Benjamins Publishing Company, pp.137–166.
18. Howarth, P., 1998. *The Phraseology of Learner's Academic Writing*. *Phraseology: Theory, Analysis, and Applications* (ed.) A. P. Cowie. Oxford University Press, pp.161–186.
19. Jakaitienė, E., 2010. *Leksikologija*. Vilnius: Vilniaus universiteto leidykla.

20. Kovalevskaitė, J. 2012. Lietuvių kalbos samplaikos. Daktaro disertacija. Kaunas: Vytauto Didžiojo universitetas.
21. Lamiroy, B. 2016. For a Typology of Phraseological Expressions: How to tell an Idiom from a Collocation? Defining Collocation for Lexicographic Purposes. From Linguistic Theory to Lexicographic Practice (eds.) A. Orlandi, L. Giacomini. *Linguistic Insights. Studies in Language and Communication*, vol. 219, Peter Lang, pp.123–150.
22. Mandravickaitė, J., Rimkutė, E., Krilavičius, T., 2016. Hybrid Approach for Automatic Identification of Multi-Word Expressions in Lithuanian, Proceedings of the Seventh International Conference Baltic HLT 2016. Amsterdam, Berlin, Tokyo, Washington, DC: IOS Press, pp.153–159. Prieiga internete: http://mwe.lt/wp-content/uploads/2016/09/Hybrid_Approach.pdf [žiūrėta liepos 9, 2017].
23. Marcinkevičienė, R. 2001. Tradicinė frazeologija ir kiti sustabarėję kalbos vienetai. *Lituanistica*, 2001, 4 (48), Lietuvos mokslų akademijos leidykla, pp.81–98.
24. Marcinkevičienė, R. 2010. Lietuvių kalbos kolokacijos. Mokslo monografija. Kaunas: Vytauto Didžiojo universitetas. Prieiga internete: http://vddb.laba.lt/fedora/get/LT-eLABa-0001:B.03-2010-ISBN_978-9955-12-656-0/DS.001.0.01.BOOK [žiūrėta birželio 23, 2017].
25. Mel'čuk, I., 1998. Collocations and Lexical Functions. *Phraseology: Theory, Analysis, and Applications* (ed.) A. P. Cowie. Oxford University Press, pp.23–54.
26. Moon, R. 1998a. Frequencies and Forms of Phrasal Lexemes in English. *Phraseology: Theory, Analysis, and Applications* (ed.) A. P. Cowie. New York: Oxford University Press, pp.79–100.
27. Moon, R. 1998b. *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford: Oxford University Press.
28. Nesselhauf, N. 2004. What are Collocations? Phraseological Units: Basic Concepts and Their Application (eds.) D. J. Allerton, N. Nesselhauf, P. Skandera. Schwabe Verlag Basel, pp.1–21.
29. Nesselhauf, N., 2005. The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, no. 24 (2), pp.223–242.
30. Paulauskas, J., 2003. Lietuvių kalbos frazeologijos žodynas. Kaunas: Šviesa.
31. Ramisch, C., 2014. *Multiword Expressions Acquisition*. Springer.
32. Rimkutė, E., Bielinskienė, A., Kovalevskaitė, J. (sud.), 2012. Lietuvių kalbos daiktavardinių frazių žodynas. Kaunas: Vytauto Didžiojo universitetas. Prieiga internete: http://donelaitis.vdu.lt/lkk/pdf/daikt_fr.pdf [žiūrėta birželio 20, 2017].
33. Rimkutė, E., Kovalevskaitė, J. 2015: Lietuvių kalbos daiktavardinių frazių žodyno vienaformiai pastovieji junginiai. *Darbai ir dienos* 64, pp.153–167.
34. Schmitt, N., 2005. Formulaic Language: Fixed and Varied, *ELIA: Estudios de Lingüística Inglesa Aplicada*, no. 6, pp.13–39.
35. Shin, D., & Nation, P., 2008. Beyond Single Words: the Most Frequent Collocations in Spoken English. *ELT Journal*, no. 62 (4), pp.339–348.
36. Siepman, D., 2005. Collocation, Colligation and Encoding Dictionaries. Part 1. *International Journal of Lexicography*, vol. 18, no. 4, pp.410–443. <https://doi.org/10.1093/ijl/eci042>
37. Sinclair, J., 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
38. Siyanova-Chanturia, A., & Spina, S., 2015. Investigation of Native Speaker and Second Language Learner Intuition of Collocation Frequency. *Language Learning*, no. 65 (3), pp.533–562.
39. Trawiński, B., Sailer, M., Soehn, J.-Ph. et al., 2008. Cranberry Expressions in English and German. Proceedings of the LREC Workshop Towards a Shared Task of Multiword Expressions, pp.35–38.
40. Virbickienė, G., 2017. Lietuvių kalbos frazeologizmų gramatinis variantiškumas: tekstynu paremtas tyrimas. *Bakalauro darbas*. Kaunas: Vytauto Didžiojo universitetas.
41. Wray, A., 2002. *Formulaic Language and the Lexicon*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511519772>
42. Wray, A., 2012. What do We (Think We) Know About Formulaic Language? An Evaluation of the Current State of Play. *Annual Review of Applied Linguistics*, no. 32, Cambridge University Press, pp.231–254.

Šaltiniai

1. Delfi.lt tekstynas.

Agnė Bielinskienė, Jolanta Kovalevskaitė, Erika Rimkutė, Laura Vilkaitė. Towards the Criteria for Identification of Idioms and Collocations

Summary

The paper analyses two types of multi-word expressions: collocations and idioms. It provides definitions of these two types of lexical items and reviews their understanding in works of Lithuanian and foreign scholars. A problem of identifying multi-word expressions arose when carrying out a research project *Automatic Identification of Lithuanian Multi-word Expressions (PASTOVU)*. When the analysis of the corpus compiled from delfi.lt news portal texts started, it turned out that there is a need for more precise criteria to identify collocations and idioms in Lithuanian data. Therefore, this paper aims at describing clear criteria that would allow linguists to identify these particular multi-word expressions in texts as objectively as possible. It seems that collocations and idioms have three main features: conventionality, semantic non-transparency, and fixed form. All these features and their strength can be evaluated applying various tests designed based on identification criteria. The paper explains the proposed tests and gives examples for each one of them in order to practically reveal the similarities and differences between collocations and idioms. From the data described, it becomes clear that the classes of collocations and idioms are not homogeneous: both of them have typical and less typical cases. We suggest conventionality and semantic non-transparency as two main features, which could help to separate collocations from idioms by applying particular criteria and tests.

Agnė Bielinskienė

Dr., Vytauto Didžiojo universitetas

Tyrimų sritys

Tekstynų lingvistika, pastovieji junginiai, automatinė kalbos analizė.

Adresas

V. Putvinskio g. 23-216, Kaunas

El. paštas:

agne.bielinskiene@vdu.lt

Jolanta Kovalevskaitė

Dr., Vytauto Didžiojo universitetas

Tyrimų sritys

Tekstynų lingvistika, pastovieji junginiai, terminologija ir leksikografija.

Adresas

V. Putvinskio g. 23-216, Kaunas

El. paštas:

jolanta.kovalevskaite@vdu.lt

Erika Rimkutė

Doc. dr., Vytauto Didžiojo universitetas

Tyrimų sritys

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

Adresas

V. Putvinskio g. 23-216, Kaunas

El. paštas:

erika.rimkute@vdu.lt

Laura Vilkaitė

Dr., Vilniaus universitetas

Tyrimų sritys

Antrosios kalbos leksikos įsisavinimas, pastovieji junginiai, psicholingvistika.

Adresas

Universiteto g. 5, LT-01513 Vilnius

El. paštas:

laura.vilkaite@ff.vu.lt

Apie autorens