

Lietuvių kalbos arbitraliosios kolokacijos ir joms skirti ištekliai

Erika Rimkutė, Jolanta Kovalevskaitė

[Loïc Boizou, Ieva Bumbulienė, Jurgita
Vaičėnienė]



Projektas „Arbitraliosios lietuvių kalbos kolokacijos: atpažinimas, aprašymas, vartojimas“ (ARKA); finansuoja Lietuvos mokslo taryba (Nr. S-LIP-20-18)

<https://arka.pastovu.vdu.lt/>



VYTAUTO
DIDŽIOJO
UNIVERSITETAS
M C M X X I I

Ankstesni darbai

- 2016–2018 m. vykdytas projektas „Lietuvių kalbos pastoviųjų žodžių junginių automatinis atpažinimas (PASTOVU)“. Projektą finansavo Lietuvos mokslo taryba (projekto nr. LIP-027/2016).
- Tikslas – sukurti lietuvių kalbos pastoviųjų žodžių junginių tyrimo metodiką ir atlikti išsamią lietuvių kalbos pastoviųjų žodžių junginių analizę, sukurti arba pritaikyti reikalingus įrankius bei išteklius.

<https://pastovu.vdu.lt/>

- PASTOVU vykdytojų grupė: Erika Rimkutė, Agnė Bielinskienė, Loïc Boizou, Ieva Bumbulienė, Jolanta Kovalevskaitė, Tomas Krilavičius, Justina Mandravickaitė, Laura Vilkaitė.

Pastovu projekto rezultatai

- Parengta „Lietuvių kalbos pastoviųjų žodžių junginių bazė“: <https://resursai.pastovu.vdu.lt/paieska/paprastoji>;
- Lietuvių kalbos pastoviųjų žodžių junginių atpažintuvas *Colloc*: <https://resursai.pastovu.vdu.lt/atpazintuvas>;
- „Lietuvių kalbos kolokacijų žodynas“: <https://pastovu.vdu.lt/zodynas/>;
- DELFI.lt tekstynas: <https://klc.vdu.lt/pastovuSearch.html>.
- Analizuotos kolokacijos ir kiti pastovieji žodžių junginiai su 97 dažniausiai į šiuos junginius įeinančiais daiktavardžiais (pvz.: *aplinka, darbas, dėmesys, pagalba, moteris, vaikas, sprendimas*).
- Apie 12 tūkst. kolokacijų.

ARKOS projektas

- Vykiant PASTOVU projektą, dėl apsiribojimo, t. y. kolokacijos tik su 97 daiktavardžiais, pastebėta, kad aprėpta nedidelė dalis duomenų, taigi:
 - 1) trūksta įvairovės,
 - 2) dalis kolokacijų, nors gramatiškai ir leksiškai yra taisyklingos, bet iš esmės sudaromos reguliariai, todėl tiek gimtakalbiams, tiek besimokantiems lietuvių kalbos nekelia didelių problemų, pvz., *graži moteris, saulėta diena, šešeri metai*.

Kolokacijų tipai

1) motyvuotosios (trivialiosios) kolokacijos (žr. R. Marcinkevičienė *Lietuvių kalbos kolokacijos*. Kaunas: Vytauto Didžiojo universiteto leidykla, 2010): jų junglumas paaškinamas semantiniiais apribojimais (pvz.: *graži diena, naujas tyrimas*);

2) **arbitraliosios** (tikrosios) kolokacijos (toliau AK) yra nemotyvuotos, joms būdingas ribotas leksinis junglumas: pagal panašią reikšmę teoriškai galima rinktis daugiau negu vieną žodį, bet linkstama vartoti būtent kurį nors vieną, pavyzdžiui:

- *platus akiratis* – **didelis akiratis*,
- *stipri sveikata* – **galinga sveikata*,
- *galingas automobilis* – **stiprus automobilis*.

AK nustatymas: neautomatinis

Taikyti du kriterijai:

- **riboto leksinio junglumo kriterijus**, kuriam įvertinti taikyti šie testai:
 - a) sinonimų pakeičiamumo (*nepaklusti įstatymui – *nusižengti įstatymui, įdėtas darbas – *įkištas / įkeltas darbas*) ir (arba)
 - b) leksinių semantinių grupių (*AK statyti automobilį, nes prišvartuoti laivą, nutupdyti lėktuvą; ne AK nuskaičiuoti mokesčius*);
- **perkeltinės reikšmės kriterijus** (*sugriauti gyvenimą, slidus dalykas, švarus darbas, kamuojantis klausimas, ryškus įvykis*).

Daugiausia AK nustatyta neautomatiniu metodu.

AK nustatymas: automatizuotas (1)

- **Vektorių metodu** buvo nustatyti pakeičiamumo testui reikalingi sinonimai.
 - 1 žingsnis: imama vektorių eilė su dviem sinonimų poromis a) *gauti* ir *įgyti*; b) *duoti* ir *teikti*;
 - 2 žingsnis: pagal šią vektorių eilę tikrinama, su keliais iš šių sinonimų daiktavardis *apdovanojimas* sudaro kolokacijas; yra dvi kolokacijos: *gauti apdovanojimą* (sinonimas iš poros *gauti* ir *įgyti*; nėra **įgyti apdovanojimą*) ir *teikti apdovanojimą* (sinonimas iš poros *duoti* ir *teikti*, nėra **duoti apdovanojimą*);
 - 3 žingsnis: priimamas sprendimas, kad *gauti apdovanojimą* ir *teikti apdovanojimą* laikyti AK.

AK nustatymas: automatizuotas (2)

1 žingsnis: imama vektorių eilė su dviem sinonimais *įkurti* ir *įsteigti*;

2 žingsnis: pagal šią vektorių eilę tikrinama, su keliais iš šių sinonimų daiktavardis *asociacija* sudaro kolokacijas; yra dvi kolokacijos: *įkurti asociaciją* ir *įsteigti asociaciją*;

3 žingsnis: priimamas sprendimas, kad *įkurti asociaciją* ir *įsteigti asociaciją* nelaikyti AK.

Publikacijos apie AK nustatymą

- Kovalevskaitė J., Rimkutė E., Vaičėnienė J. 2021: [Arbitraliųjų lietuvių kalbos kolokacijų nustatymas](#). *Bendrinė kalba* 94, 1–37. doi.org/10.35321/bkalba.2021.94.06.
- Kovalevskaitė J., Rimkutė E., Vaičėnienė J. 2021: [Automatizuotas arbitraliųjų kolokacijų atpažinimas: būdvardžių ir daiktavardžių kolokacijos](#). *Studies about Languages / Kalbų studijos* 39, 71–84.

Lietuvių kalbos pastoviųjų žodžių junginių duomenų bazė

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

- Pateikiama apie 14,5 tūkst. dvižodžių (12702) ir trižodžių (1755) kolokacijų;
- Iš jų 4525 AK (jos pavartotos 35 169 kaitybinėmis formomis).
- Struktūra:
 - būdvardis + daiktavardis: *didelis dėmesys, sunkus darbas, brangus žmogus, didžioji pertrauka;*
 - veiksmažodis + daiktavardis (+ kiti žodžiai): *atkreipti dėmesį (į), užkirsti kelią, kilti klausimas, neatmesti galimybės;*
 - daiktavardis + daiktavardis: *pramogų pasaulis, problemos sprendimas, rankų darbo, saugumo užtikrinimas.*

Tolesni darbai

- Baigti tvarkyti AK duomenų bazėje (bus apie 18 tūkst. kolokacijų, iš jų apie 8–9 tūkst. AK).
- Parengti mokomąją priemonę apie kolokacijas, skirtą šioms pagrindinėms tikslinėms grupėms: besimokantiesiems lietuvių kalbos (ir kaip gimtosios, ir kaip svetimšios); besimokantiems vertėjams, redaktoriams.
- Parengti „Lietuvių kalbos arbitraliųjų kolokacijų žodyną“.