ERIKA RIMKUTĖ, JOLANTA KOVALEVSKAITĖ & JURGITA VAIČENONIENĖ
Vytauto Didžiojo universitetas, Lietuva

Lietuvių kalbos arbitraliosios kolokacijos ir jų atpažinimas

SANTRAUKA
Šio pranešimo tikslas – pristatyti vykdomą Lietuvos mokslo tarybos finansuojamą projektą „Arbitraliosios lietuvių kalbos kolokacijos: atpažinimas, aprašymas, vartojimas (ARKA)" (Nr. S- LIP-20-18; http://arka.mwe.lt/), kuriame analizuojamos lietuvių kalbos arbitraliosios kolokacijos (toliau AK). Priešingai nei motyvuotosios kolokacijos (žr. Marcinkevičienė, 2010), kurių junglumas paaiškinamas semantiniais apribojimais (pvz., *graži diena, naujas tyrimas),* AK yra nemotyvuotos: pagal panašią reikšmę teoriškai galima rinktis daugiau negu vieną žodį, bet linkstama vartoti būtent kurį nors vieną, pavyzdžiui, *platus akiratis,* bet ne *didelis akiratis, stipri sveikata*, bet ne *galinga sveikata.* Pirmajame projekto etape atliktas ankstesniame projekte (žr. http://mwe.lt/) sukauptų būdvardis+daiktavardis kolokacijų tyrimas. Palyginus visas su tam tikru būdvardžiu sudarytas kolokacijas, nustatyta, kad dalies AK būdvardžiai pavartoti netiesiogine reikšme: atsižvelgiant į daiktavardį, įvertinta, kuriose kolokacijose būdvardis vartojamas perkeltine reikšme. Iš maždaug 8000 būdvardis+daiktavardis kolokacijų pagal netiesioginės reikšmės kriterijų atpažinta apie 800 AK, pvz.: *nepatogi vieta, skaidrūs rinkimai.* Tarp būdvardis+daiktavardis kolokacijų yra tokių, kuriose nėra perkeltinės reikšmės žodžių, bet kurios, tikėtina, laikytinos AK, pavyzdžiui, *egzistencinis klausimas, principinis sprendimas, priimtina kaina*. Todėl antroje tyrimo dalyje testuojami papildomi AK atpažinimo kriterijai, vienas jų – kolokacijų dėmenų pakeičiamumas (Nesselhauf, 2003). Jeigu kolokacijos būdvardiškasis dėmuo nėra pakeičiamas kitu artimos reikšmės sinonimu, yra pagrindo tokią kolokaciją laikyti AK. Artimiems sinonimams nustatyti naudojamas GloVe žodžių vektorių (angl. *embeddings*) modelis (Pennington et al., 2014). Daroma prielaida, kad kolokaciją galima laikyti AK, jeigu ji užfiksuota tik su vienu būdvardžiu iš galimų artimų sinonimų eilės. Jeigu pakeičiamumo testavimui trūksta duomenų arba jie nėra tinkami, tikrinamas trečiasis kriterijus – vertinama, kokia yra trauka tarp būdvardžio ir daiktavardžio. Šis rodiklis (*collocational factor*, Handl, 2008), apskaičiuojamas naudojant abiejų kolokacijos žodžių pavartojimo dažnį tekstyne ir kolokacijos dažnumą.

PAGRINDINĖS SĄVOKOS
Arbitraliosios kolokacijos; lietuvių kalba; atpažinimas; vektoriai; metaforos.

LITERATŪRA
Handl, S. (2008). Essential Collocations for Learners of English: The Role of Collocational Direction and Weight. In F. Meunier & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 43–660).

Marcinkevičienė, R. (2010). *Lietuvių kalbos kolokacijos*. Mokslo monografija. Vytauto Didžiojo universitetas. https://talpykla.elaba.lt/elaba-fedora/objects/elaba:4603559/datastreams/MAIN/content

Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, *24*(2), 223–242.

Pennington, J., Socher, R., & Manning, Ch. (2014). GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). *Association for Computational Linguistics*, 1532–1543. https://www.aclweb.org/anthology/D14-1162