

ARBITRALIŲJŲ LIETUVIŲ KALBOS KOLOKACIJŲ NUSTATYMAS

Santrauka

Šiame straipsnyje sprendžiama arbitraliųjų lietuvių kalbos kolokacijų atpažinimo problema. Tam skirtas projektas „Arbitraliosios lietuvių kalbos kolokacijos: atpažinimas, aprašymas, vartojimas (ARKA)“. Arbitraliosiomis kolokacijomis laikomos kolokacijos, kurių žodžių junglumo pagrindas yra ne semantiniai, o leksiniai ribojimai. Tokios kolokacijos tiesiogiai neatspindi nekalbinės tikrovės, o ryšys tarp jų žodžių – arbitralus: nors pagal panašią reikšmę teoriškai galima rinktis daugiau negu vieną junglumo partnerį, bet linkstama vartoti kurį nors vieną, pvz., *suprasti užuominą*, bet ne **suvokti užuominą*.

Tyrimo šaltinis yra Lietuvių kalbos pastoviųjų žodžių junginių duomenų bazė, kurioje pateikiama apie 12 tūkst. kolokacijų; ištyrus kolokacijas su 97 daiktavardžiais, arbitraliosiomis priskirta apie 2400 kolokacijų, kurių struktūra yra būdvardis / būdvardiškai vartojamas dalyvis + daiktavardis, veiksmazodis + daiktavardis, daiktavardis + daiktavardis.

Straipsnyje aprašyti du pagal minėtą projektą išbandyti arbitraliųjų kolokacijų nustatymo metodai: neautomatinis lingvistinės analizės ir automatizuotas kompiuterinės lingvistikos metodas. Nustatant AK neautomatiškai, taikyti du pagrindiniai kriterijai: perkeltinės reikšmės kriterijus ir riboto leksinio junglumo kriterijus, kuriam įvertinti taikyti veiksmazodžio ar būdvardžio sinonimų pakeičiamumo ir (arba) daiktavardžių leksinių semantinių grupių testai. Vienas iš AK nustatymo kriterijų – riboto leksinio junglumo – buvo tikrinamas ir automatizuotai, vektorių metodu nustačius pakeičiamumo testui reikalingus sinonimus. Automatizuoto metodo taikymą sudarė šie etapai: 1) automatinis, kai vektorių metodu buvo surenkamos vektorių eilės su panašų kontekstą turinčiais žodžiais; 2) neautomatinis – vektorių eilių tvarkymas paliekant tik sinoniminiais ryšiais susijusius žodžius; 3) iš dalies automatinis etapas, kai kolokacijos buvo

JOLANTA KOVALEVSKAITĖ, ERIKA RIMKUTĖ,
JURGITA VAIČENONIENĖ

automatiškai susiejamos su konkrečiomis sinonimų eilėmis ir po peržiūros lingvistų patvirtinamos arba nepatvirtinamos kaip AK.

Neautomatiniu metodu nustatyta apie pusė AK, automatizuotu metodu – apie trečdalis. Neautomatinei analizei taikytas ne vienas, o du atrankos kriterijai, todėl neautomatiškai nustatytų duomenų daug daugiau negu nustatytų automatizuotai. Nors automatizuotu metodu galima sumažinti subjektyvumą vertinant duomenis ir supaprastinti arbitralumo vertinimo procedūrą, tačiau pastebėta ir tai, kad arbitralumo pripažinimą lemia vektorių eilėje esantys sinonimai, taigi dalis AK gali likti neatpažintos. Vektorių metodu surinktus sinonimus reikia atskirti nuo kitais semantiniais ryšiais susijusių žodžių, tam reikalinga lingvisto patikra. Tačiau, nepaisant šio ribotumo, vektorių metodas yra tinkamas dideliame duomenų kiekiui apdoroti, o rankiniu įvertinti gausybės žodžių konteksto panašumą būtų sunku.

Abiem metodais atpažintos kolokacijos sudaro tik penktadalį visų nustatytų AK, todėl yra pagrindo manyti, kad siekiant atpažinti AK svarbu taikyti tiek automatizuoto, tiek ir neautomatinio metodo teikiamas galimybes.

ESMINIAI ŽODŽIAI: arbitraliosios kolokacijos, lietuvių kalba, ribotas leksinis junglumas, vektorių metodas, automatizuotas nustatymas, lingvistinis neautomatinis nustatymas.

JOLANTA KOVALEVSKAITĖ, ERIKA RIMKUTĖ,
JURGITA VAIČENONIENĖ

Vytauto Didžiojo universitetas
Kristijono Donelaičio g. 58, 44248 Kaunas

jolanta.kovalevskaite@vdu.lt

erika.rimkute@vdu.lt

jurgita.vaicenonienė@vdu.lt