

Dažniausios lietuvių kalbos morfologinio daugiareikšmiškumo rūšys ir jų automatinis vienareikšminimas

Erika Rimkutė, Aušra Grybinaitė

Anotacija. Straipsnyje pristatomas mažai tyrinėtas lietuvių kalbos morfologinis daugiareikšmiškumas ir pirmieji automatinio vienareikšminimo bandymai. Rašoma apie automatinio būdu sulemtą ir morfologiškai anotuotą lietuvių kalbos tekstyną. Ištyrus anotuotą tekstyną, paaiškėjo, kad kalba yra labai daugiareikšmė – apie 50 proc. žodžių ar žodžių formų yra morfologiškai daugiareikšmės. Norint panaudoti morfologiškai anotuoto tekstyno duomenis tolesniems tyrimams (automatinei sintaksinei analizei, kelių kalbų lygiagrečiam nagrinėjimui, automatiniam vertimui) reikia turėti vienareikšmes formas. Kitoms kalboms yra sukurta nemažai specialių vienareikšminimo programų. Lietuvoje ši sritis yra visai nauja ir mažai tirta, todėl tik visai neseniai buvo pradėta išsamiau gilintis į morfologinį daugiareikšmiškumą ir jo ribojimo galimybes. Šis straipsnis – tai lingvistų ir informatikų bendradarbiavimo, ribojant morfologinį daugiareikšmiškumą, rezultatų aptarimas.