

Abstrakt

Pád je v češtině základním prostředkem morfologické roviny, jímž substantiva vyjadřují svou funkci ve větě. Cílem této práce je popsat z frekvenčního hlediska vztah mezi formou a funkcí substantiv, přesněji řečeno, jak často se prosté a předložkové pády substantiv používají k realizaci syntaktických funkcí ve větě. Práce je založena na rozsáhlém korpusu synchronní psané češtiny SYN2005.

Abychom získali údaje o frekvencích syntaktických funkcí substantiv ve vztahu k jejich pádům, opatřili jsme korpus SYN2005 závislostním syntaktickým značkováním, jehož formát jsme převzali z analytické roviny Pražského závislostního korpusu. Syntaktickou anotaci jsme uskutečnili pomocí stochastického MST parseru. Spolehlivost syntaktické anotace však nebyla dostatečně vysoká, vytvořili jsme proto automatický opravný modul, který vyhledává chyby syntaktické anotace ve výstupu stochastického parseru a na základě lingvistických pravidel tyto chyby opravuje. Implementovali jsme 26 různých pravidel, počet chyb anotace se však podařilo snížit jen o 6–8 %. Opravný modul je však možné dále rozvíjet. Lze jím korigovat výstup kteréhokoli závislostního parseru natrénovaného na datech Pražského závislostního korpusu.

Syntakticky anotovaný korpus SYN2005 jsme využili jako základ výzkumu frekvence syntaktických funkcí substantiv se zřetelem k jejich pádu. Manuálními sondami do korpusu jsme ověřili a korigovali získané statistické údaje. Podrobně jsme analyzovali relevantní kombinace syntaktické funkce a prostého nebo předložkového pádu. Frekvence syntaktických funkcí a jejich realizací pády substantiv jsme představili v přehledných tabulkách, a to pro celý korpus SYN2005 i pro jeho tři subkorporusy: beletrie, odborné literatury a publicistiky. Tyto subkorporusy jsme porovnali jak podle celkové frekvence syntaktických funkcí substantiv, tak podle použití některých kombinací syntaktické funkce a pádu.