

Abstrakt

Práce představuje metodu pro diachronní srovnání synchronních korpusů zachycujících blízké stavy jazyka. Cílem práce je především zhodnotit možnosti a meze detekce vývojových tendencí v jazyce na materiálu synchronních psaných korpusů řady SYN. Metodologicky jde o corpus-driven přístup založený na statistickém vyhodnocení rozdílů mezi normalizovanými průměrnými redukovánými frekvencemi lemmat a lexikálních kombinací. Metoda je aplikována v několika variantách na různě definované subkorpora korpusu SYN a podrobně vyhodnocena.

Provedené srovnání ztěžuje především vliv složení jednotlivých korpusů a provázanost změn v jazyce se změnami společenskými. Protože neumíme spolehlivě odlišit zárodky diachronních posunů od přirozeně existující synchronní variability, je statisticky zjištěná významnost frekvenčních rozdílů jednotlivých výrazů zpětně ověřována na korpusech a interpretace výsledků korigována znalostí jejich přesného složení.

Závěry jsou založeny především na publicistice, která je z psaného jazyka nejvíce otevřená změnám. Změny v jazyce publicistiky lze charakterizovat jako tematický odklon od původní politické a ekonomické orientace směrem k tématům týkajícím se praktického života a využívání volného času spojený se zvyšující se neformálností, která způsobuje posuny ve frekvencích některých slovních druhů, frekvenční nárůst řady lemmat z jádra slovní zásoby, vzrůstající podíl významově oslabených sloves, obměnu některých šablonovitých spojení atd.

K přínosům práce patří také vyhodnocení složení korpusů řady SYN, zvláště reprezentativních korpusů SYN2000, SYN2005 a SYN2010. Výsledkem jsou praktická doporučení ke změnám v konceptu reprezentativnosti, kategorizaci textů a složení korpusových dat, která tvoří cennou zpětnou vazbu pro budování dalších korpusů této řady.