



## Univerzita Karlova v Praze, Filozofická fakulta Ústav Českého národního korpusu

nám. Jana Palacha 2, 116 36 Praha 1  
tel.: +420 2 21 619 357, ucnk@ff.cuni.cz

*Posudek vedoucího vztahující se k rukopisu doktorské disertace*

Michala Křena

### Diachronní srovnání synchronních korpusů Diachronic comparison of synchronic corpora

Předkládaná práce Michala Křena je výsledkem mnohaleté praktické zkušenosti s užíváním a výstavbou všech verzí *Českého národního korpusu* i zkušenosti teoretické, promítající se do jeho tvorby. Motivován odborně si zde položil otázku, kterou si dosud klade málokdo a v daném rozsahu materiálu kromě něj nikdo, totiž jak a zda vůbec lze srovnávat korpusy jednoho jazyka ve velmi úzkém časovém rozpětí a v souvislosti s tím i druh přínosu k našemu obecnému poznání jazykových změn za tak krátkou dobu. Autor se důkladně poučil ze všech dostupných a aspoň zčásti relevantních zdrojů z jiných jazyků a dospívá k závěru, že je třeba se vydat vlastní cestou, pochopitelně zásadně korpusovou, nazývanou obvykle corpus-driven a založenou tedy pouze na datech v korpusech nalezených.

Po rozsáhlé analýze problematiky reprezentativnosti a povahy použitelných datových vzorků se rozhoduje pro data jednak v korpusech nejbohatší, tj. publicistiku (s vědomím, že nejde o jazyk celý) a jednak pro způsob jejich studia. Volí v důsledku už osvědčené praxe metodu jejich normalizace umožňující přístup ARF na vzorcích ze třech korpusů časově s pětiletým odstupem; vlastním jeho zájmem a cílem se stávají lexémy a jejich kombinace. Zvláště ten druhý cíl je závažný: protože v tomto v zásadě formálním přístupu nelze dojít k systematickým poznatkům pro jazyk zásadním, totiž vývoji a změně významu, dostává se k nim aspoň oklikou studiem kombinací (kolokací) lexémů, resp. tvarů ve zvolených vzorcích.

Práce nabízí i v řadě ohledů vzhled do postupně rozvíjené metodologie přístupu, kdy se ilustrují starší metody (i domácí), kriticky se upozorňuje na jejich nedostatky a navrhuje přístupy vylepšené a modifikované. Týká se to jak srovnávání mezi žánry tak na vzorcích jednotlivých časových údobí uvnitř žánru jediného, vše za použití několika statistických metod, jejich výsledky jsou přehledně pro srovnání tabelizovány s explicitním vytčením, zda jde mezi sledovanými obdobími o pokles v zastoupení formy či její vzestup a tedy vyšší frekvenci. Specificky přitom mj. sleduje systematickosti takového nárůstu či poklesu, což může ukazovat na obecnější tendenci. Rozsah použitých vzorků se omezuje na 50 formálních jednotek. Vhodnost takové metodologie ověřuje do jisté míry i na diachronních korpusech. Obecnou otázkou a problémem však zůstává rozsah dat.

Jakkoliv si je autor plně vědomý úskalí takového dosud nevyzkoušeného přístupu, je to podle mého názoru přístup a pokus zdařilý a ve volbě lexikonu jako objektu ve srovnání s jinými rovinami zkoumání šťastný. Je nepochybné, že časová blízkost dat ztěžuje možný výklad výsledků (který však nebyl cílem jeho práce, je ale podkladem pro další možné studie), kdy se můžou prolínat skutečná jazyková změna s dobovým výkyvem daným společenskou situací, která se nejrychleji odráží právě v lexikonu, specificky v publicistice, k níž se vztahují jeho hlavní a cenné závěry. Představuje zde, na datech synchronních, vlastně metodu srovnání diachronního, což dobře ukazuje na časově provázanou povahu dat i obou oblastí, synchronie a diachronie.

Práce M. Křena, ve které přesvědčivě zvládl novou problematiku a našel k jejímu uchopení adekvátní formální přístup, je za velmi vítaný příspěvek k teorii tvorby korpusů nabízející i řadu zcela praktických výsledků metodologie. Velmi rád ji doporučuji k obhajobě i případné budoucí publikaci.

Prof. PhDr. František Cermák, DrSc, vedoucí práce