

Posudek disertační práce

Michal Křen: *Diachronní srovnání synchronních korpusů*

Předložená disertační práce je první studií v české korpusové lingvistice, která se podrobně zabývá vzájemným porovnáním českých korpusů na statistickém základě s jasnou představou směru interpretace. K porovnávání korpusů vede přirozený vývoj daný růstem jejich počtu a existují proto jak domácí, tak zahraniční úvahy a studie na toto téma. Autor prokazuje jejich znalost, nabízí jejich kritické posouzení a osvědčuje schopnost vybírat z možných řešení ta, která jsou aplikovatelná na české korpusy řady SYN. Umí shromáždit a zhodnotit podstatné postřehy a budovat na nich i vedle nich vlastní myšlenkový základ. Jeho cílem přesto není pouhé, byť pracovně neobyčejně náročné srovnávání korpusů, ale odhalování možností a mezi detekce odlišností, zejména s přihlédnutím k různým technickým i věcným podmínkám vzniku korpusů a vzhledem k vývoji jejich textové základny. Autor si je vědom obtížnosti rozlišování mezi oscilacemi hodnot a zárodky diachronních posunů (mathesiovským kolísáním a měnlivostí) a hledá pro svůj pohled co nejcitlivější, ale zároveň co nejobecnější metodu. To vše vyžaduje vedle znalosti a schopnosti orientovat se v oboru i v dosud ne zcela ustáleném popisu nových postřehů také schopnost nadhledu, kontroly a tříbení poznatků „za pochodu“.

Autor nezapomíná na provázanost změn v korpusech se změnami společenskými a je si vědom všech zvláštností ve struktuře a reprezentativnosti českých synchronních korpusů. Počítá s polemickými otázkami a ihned na ně odpovídá. Věcně, kriticky i s pochopením analyzuje rozdíly v budování korpusů, má k němu oprávněné výhrady, ale nedá se jimi strhnout, umí s danostmi pracovat. Zvolí-li směr (lexikon), nepouští ze zřetele ani jeho průvodní okolí (kolokace, gramatické kategorie). Zvolí-li metodu nebo metody, zná a ihned vysvětlí její nebo jejich genezi s přednostmi i slabinami. Konstrukci své metody vede srozumitelně v detailech i v delších obloucích výkladu. Postup hodnocení autor vede důsledně s vědomím významu dat, nikdy slepě aritmeticky, a to až do úrovně pravděpodobnostního chápání nulových hypotéz. To samo je sice vedlejší, ale důležitý klad celé práce. Zjevné anomálie proto autor dovede věcně analyzovat a nalézat pro ně vysvětlení. Pro spolehlivost budoucí interpretace je to přístup klíčový. Široká analýza, jasně formulované úvahy podložené sondami a detailní znalostí stavby užitých korpusů vedou autora k myšlence postupovat po subkorpusech a z jejich možností k zaměření na publicistické texty. Některé dílčí postupy pro svou práci již vyřešil a publikoval dříve.

Za opěrnou statistickou charakteristiku této práce autor volí normalizovanou ARF (průměrnou relativní frekvenci), významnou, mezinárodně obhájenou a mnohonásobně osvědčenou specialitu svého pracoviště. Nezůstává však na ní, bere v úvahu i další možnosti, porovnává je a interpretačně testuje jejich vhodnost. To mu

umožňuje nikoli obejít, jak skromně píše, ale překlenout problém reprezentativnosti a vyváženosti korpusů vzniklých v různých časových úsecích. Nikoli obecně novým, ale v kontextu korpusových studií průlomovým je autorovo rozhodnutí pracovat nad statistickou významností rozdílů mezi výchozími údaji, tedy nad pravděpodobnostními charakteristikami statistických dat. Po soustředění (autor píše o redukci záměru) na 50 lemmat a 50 kolokací a po shromáždění enormního množství numerických dat autor zhušťuje výsledky do šestirozměrných binárních vektorů a s nimi pak dál objevně pracuje. Prokazuje tak i vzácnou schopnost vytvářet abstrakci nad abstrakcí, v jejím světě uvažovat a srozumitelně interpretovat poznatky. Autorův návrh šestic (obecně n-tic) „indikátorových dvojic“ je třeba považovat za velmi významný metodologický návrh, který by se mezi jinými, nápadnějšími klady celé práce neměl ztratit.

Vzhledem k časové sevřenosti rozestupu užitých korpusů (2000-2005-2010) jsou výsledkem sérií sledování – v autorově formulaci – spíše změny diskurzu, než změny v jazyce. Příklady ilustrovaných výsledků jsou velmi zajímavé, dobře (i efektně) interpretovatelné a v textu práce se nabízejí desítky dalších. Za důležité a přínosné je třeba považovat také zaměření na kombinace lexémů jako možné zárodky frazémů, byť další práce s nimi nemůže postrádat intuitivní korekci, což ovšem autor ví. Základ tu však je a zdá se originální. Možnosti výsledků zvolených variant vlastní metody jsou dobře patrné v grafech, v jejich výpovědní hodnotě, v interpretovatelnosti a inspirativnosti. Jasná shoda racionálního pohledu s interpretovanou intuicí zároveň velmi cenným způsobem zpětně prokazuje oprávněnost autorova přístupu, prokazuje jeho spolehlivost a podtrhuje jeho nepřehlédnutelnou heuristickou hodnotu. Autor ukazuje i cesty a směry, kterými by mohlo jeho téma pokračovat.

Předložená disertační práce tedy není jen důkazem autorovy schopnosti orientovat se v literatuře daného oboru a ve zpracovávání úctyhodně velkého množství statistických údajů, ale zároveň a možná především je originálním metodologickým artefaktem. Její opření o aktuální světový kontext studií v oboru korpusové lingvistiky a přitom vlastní cesta, pečlivost a nadhled při hledání a nalézání vlastních návrhů jsou objevné. Prokazatelně vystupují nad požadavky dobré disertační práce. Předloženou disertační práci *Diachronní srovnání synchronních korpusů* proto s uspokojením hodnotím jako výbornou.

Závěr:

Na základě předložené práce *Diachronní srovnání synchronních korpusů* doporučuji bez výhrad přiznat uchazeči Mgr. Michalu Křenovi titul Ph.D.

RNDr. Jan Králík, CSc.,
Ústav pro jazyk český AV ČR, v. v. i.