

Podezřelá kvantitativní evidence

Barbora Chybová — Filozofická fakulta Univerzity Karlovy, Praha —
Ústav pro jazyk český AV ČR

Za publikací *Slovo a text v historickém kontextu — perspektivy historickosémantické analýzy jazyka* stojí autorský kolektiv z Ostravské univerzity v Ostravě v čele s vedoucím Katedry českého jazyka Jaroslavem Davidem, dále nakladatelství Host, a v neposlední řadě Grantová agentura České republiky, která v letech 2011–2013 finančně podporovala projekt s názvem *Historická sémantika*. Řešitel mj. přislíbil, že „výstupem projektu bude monografie *Kapitoly z historické sémantiky*, v naší lingvistice první pokus o moderní komplexní pohled na tuto část diachronní lexikologie, včetně postižení nejnovějších trendů.“¹ Reálným výsledkem projektu je však „výstup“, k jehož charakteristice lze stěží použít slova „monografie“, „historická sémantika“, „komplexní pohled“ či „diachronní lexikologie“.

Svazek je rozdělen do deseti vnitřně disparátních kapitol, s nejasným vztahem k údajné monografičnosti publikace. I přes hierarchické (až čtyřstupňové) členění práce není možno vysledovat logiku zvolené struktury, vládne jí prostá adice a nahodilost témat. Tak v rámci úvodní kapitoly *Historická sémantika — mezi historií a lingvistikou* (Jaroslav David) čteme logicky nemotivovatelnou podkapitolu o jazyce totality. Druhá, stěžejní kapitola publikace s názvem *Kvantitativnělingvistické metody a jejich využití v historické sémantice*, skýtá — opět bez vnitřního opodstatnění — sondy do lexika novoročních projevů československých a českých prezidentů (Radek Čech) a do publicistiky Karla Čapka, Jaroslava Durycha a Ladislava Jehličky (Jana Davidová Glogarová, Radek Čech, Jaroslav David). V ostatních kapitolách jsou postupně pojednány staročeské satiry, nejfrekventovanější propria z různých korpusů, pojmenování feudálních sídel (*hrad, zámek a tvrz*) od 19. do 20. století, pojmenování osob pracujících v průmyslu a zemědělství, slova *měšťák, partaj, strana a demokracie*, prvomájová hesla (vše Jaroslav David), politická témata ostravského regionálního týdeníku v roce 1900 (Hana Šústková) a zločinecký slang (Lucie Radková). Pouhý výčet témat ukazuje jejich vzájemnou mimoběžnost, kterou lze jen obtížně ospravedlnit „rozdílným badatelským zaměřením jednotlivých členů autorského kolektivu“ (s. 12) či veskrze (ač ne důsledně) kvantitativně lingvistickým přístupem k materiálu. Právě tematický rozvrh publikace ukazuje na její klíčové problémy, ohlašující se konkrétněji a problematičtěji i v ostatních rovinách textu: na otázku výběru a souvislostí. Pro ilustraci je třeba stručně reprodukovat alespoň některé ze sond, které autoři podnikají (prozatím jen s nejnutnějším komentářem):

Rozbor prezidentských novoročních projevů (z let 1949–2013) spočívá ve stanovení nejfrekventovanějších, tzv. tematických lemmat (takových, u nichž se pořadí slova ve frekvenčním výčtu rovná jeho frekvenci nebo je nižší než frekvence; podrobněji s. 36) a vypočtení tzv. tematické koncentrace textu. Výchozí hypotéza zní toliko, že

1 Viz informace o projektu, včetně kladného hodnotitelského posudku, vyhledatelné zde: <http://www.isvav.cz/prepareProjectForm.do>.



prezidentské projevy totalitního období budou tematicky sevřenější než projevy prezidentů demokratických. U jednotlivých projevů bylo zjištěno 0–4 tematických slov, mezi nimi převažují lemmata *rok* a *nový* daná žánrem. Dále je vypočtena průměrná hodnota tematické koncentrace všech projevů každého z prezidentů a tyto hodnoty jsou vzájemně porovnávány — výsledkem je konstatování buď signifikantního, nebo nesignifikantního rozdílu mezi tematickými koncentracemi projevů jednotlivých prezidentů a je vyčíslena míra podobnosti ve volbě tematických slov. Hypotéza je nakonec pomocí tabulek, grafů a výpočtů (a navzdory celkově nízkým počtům výskytů tematických slov) potvrzena, o lexiku projevů a sémantice vytčených lexémů, o autorství či vztahu projevů k ideologii se však nic (nového) nedozvídáme.

Analýza tematických charakteristik publicistických politicky orientovaných textů Karla Čapka, Jaroslava Durycha a Ladislava Jehličky vychází ze souboru vybraných, časově nesouběžných textů (77 Durychových z let 1937–1940, 95 Jehličkových z let 1936–1942 a 95 Čapkových z let 1936–1938), které byly automaticky lemmatizovány. Výsledkem je opět pouze konstatování (ne)signifikantních rozdílů mezi autory. Podobně jako v předcházející kapitole autoři setrvávají i v závěrech u prosté evidence a programově (v rámci „objektivity“) odmítají interpretaci získaného materiálu: „Zůstává otázkou, čím je tato skutečnost [tj. rozdílnosti či podobnosti v tematické koncentraci textů Čapka, Durycha a Jehličky, doplnila B. Ch.] způsobena. Volbou témat? Autorským stylem? Ideovou podobností? Vzhledem k tomu, že nemáme žádnou teorii textu, která by nabízela obecné vysvětlení takovýchto vztahů, musíme tyto otázky prozatím nechat bez odpovědí“ (s. 83). K nenulovým výsledkům by autoři snad dospěli, kdyby podrobili lexikální analýze alespoň slova *člověk*, *dítě* a *víra*, která stanovili jako společná tematická lemmata autorů; polysémii, popř. kolokačním a textovým souvislostem těchto lemmat však pozornost nevěnují.

Kapitola teoreticky pojednávající o problematice klíčových slov odbočuje k analýze staročeských satir. Stanovení tematických slov, jejich tabulkové přehledy a vypsání několika častých kolokací nenesou o mnoho větší informaci, než jakou získáme ze samotných názvů zkoumaných textů. Ve *Sváru vody s vínem* byla vytčena lemmata *voda*, *víno* a *odpovedati* (slovo vyplývající z dialogické povahy textu). *Desatero kázanie božie* kromě lemmat *bôh* a *kázanie* vyneslo sloveso *dáti* (natolik polysémní, že bez podrobného sémantického rozboru nicneříkající), substantivum *sbožie* (vyskytující se téměř výhradně v sedmém přikázání „Nekradte cuzého sbožie“) a adjektivum *zlý* (uplatňující se především v enumeracích, tedy v souvislosti s užitím rétorické figury, s. 94). Text *Podkoní a žák* nevykázal žádné tematické lemma, je proto hodnocen — podobně jako prezidentské projevy Václava Havla a publicistika Karla Čapka — jako „tematicky roztržštěný“, *Desatero* oproti tomu jako „tematicky nejvyhraněnější“ (s. 98).

Srovnání kolokací lemmatu *demokracie* v publicistických textech před rokem a po roce 1989 zakládá na neurčitá, materiálem neopodstatněná konstatování typu „výskyt výrazu *demokracie* bez určujícího adjektiva [...] je výrazem praktické realizace demokratických principů v politickém životě“ apod. (s. 180). Vytěžený materiál zůstává netknutý sémantickou analýzou či jen prostou selekcí. Tak se vedle lemmatu *demokracie* objevuje s nejvyšší frekvencí adjektivum *sociální*, aniž by bylo vyloučeno pro výskyt v ustáleném spojení — v pojmenování politické strany; stejně tak není vyřazeno proprium *Miloš* (zaujímající 12. místo ve frekvenčním žebříčku), vztahující se také výhradně k realii politické strany, nikoli tedy k té *demokracii*, o níž Jaroslavu

Davidovi patrně jde, ale do jejíž sémantické analýzy se nepouští. Ve výčtu kolokací se neměla objevit ani zkratka *soc.* (27. místo), nejde pochopitelně o samostatné lemma, nýbrž opět o adjektivum *sociální*.

Tolik ukázky „analýz“. Autoři publikace *Slovo a text v historickém kontextu* si nicméně nezaslouží — také vzhledem k jejich možnému pokračujícímu působení v oboru — jen pohoršlivě tázavé „proč“ a „k čemu“, nýbrž podrobný rozbor rozporuplného snažení.

Za ústřední metodologické problémy práce považují: 1. výběr zkoumaného materiálu (autoři pracují s příliš malými soubory dat, jež jsou pro kvantitativní analýzu nevhodné, nereprezentativní); 2. text není uchopen důsledně jako prvek určitého stylu a žánru (chybí reflexe závislosti sémantiky slov a smyslu textových celků na žánru, navíc při nízkých počtech absolutních výskytů může výsledek radikálně vychýlit např. ironie, metafora, citace); 3. ignorování polysémie zkoumaných lexikálních jednotek, popř. nerefektování změny významu v ustálených spojeních; 4. zájem o kolokaci omezenou pouze na bezprostředně sousedící lexikální jednotky namísto porozumění kontextu (zcela obejít se bez něj nelze ani při kvantitativní analýze); 5. vyloučení koreferenčních vztahů; 6. izolovanost slova od vztahů synonymie, (ko)-hyponymie a hyperonymie, tedy absence pohledu na slovo jako součást systému; 7. programové odmítání řádné interpretace nashromážděného materiálu, setrvávání u prosté evidence (zcela chybí reflexe výpovědní hodnoty materiálu); 8. neschopnost rozlišovat mezi znakem, jeho mentálním (pojmovým) zpracováním a skutečností.

Právě nezpůsobnost orientovat se v procesu semiózy, neschopnost diferencovat realitu jazykovou od společenské/politické/sociální a celkově neurčitá představa o tom, jakých vztahů mohou nabývat jazyk a skutečnost a z jakých pozic to posuzujeme (ovšemže ne zvnějšku), rozsévají v publikaci řadu nepřesností a způsobují celkově chaotický výklad. Viz zmatečné výroky „adjektivum *soukromý* [...] je reakcí na předcházející období socialistického zemědělství a nahrazuje adjektivum *družstevní*“ (s. 163); „Analýza ukázala [...] sémantické procesy, které využíval komunistický newspeak při přehodnocování významu tematizovaných lexikálních jednotek“ (s. 182); „Ve srovnání s předchozím obdobím vykazují sledované výrazy po roce 1989 sémantickou deideologizaci (*měšťák, partaj, strana, demokracie*)“ (s. 182). Poslední z citátů bych zároveň vytkla jako charakteristický pro celkově černobílé pojetí jak proměn lexikálního systému, tak vlastního vztahování se k tomuto systému i k teoretickému diskurzu. Kvantitativní metody práce jsou inscenovány v rámci dialektického zápasu „objektivit“ (resp. „intersubjektivit“) se „subjektivitou“ a přístupu „nového“, „moderního“ vymezujícího se proti těm „tradičním“. Výsledkem je absurdní konstrukt nového, nepředpojatého badatele a naivní pokus situovat se kamsi do nulového bodu, vně zkoumaného materiálu.² A protože se během analýz stává tento požadavek obtížně realizovatelným, pomáhají si autoři, kde mohou; nejčastěji sahají k tvrzením „tradiční“ (tím rozumí ne-quantitativní) literární historie (od Hrabáka, přes Putnu až po Papouška) a historické sémantiky a používají je jako podklad pro svá bádání: nikoli pro komparaci se svými výsledky, nýbrž jako prosté východisko potřebné k formulování hypotézy — a jako argument pro případnou interpretaci a potvrzení

2 Srov. Michal L. Hořejší: Unikavost objektivit. *Host* 30, 2014, č. 4, 17. 4., s. 98 <<http://www.osu.cz/dokumenty/monitoringmedii/2297.pdf>> [28. 4. 2015].

hypotézy. „Inspirace“ tak proudí opačným směrem, než autoři deklarují (s. 62). Tento v sobě zacyklený postup uplatňují zcela nereflktovaně a konformně, bez kritického posouzení konkrétních pramenů. Nacházejí se tak v nebezpečné blízkosti svého tématu, totiž manipulujícího a mýtotvorného jazyka propagandy (jenž je ostatně na některých místech textu reprodukován bez vyznačení citace či badatelova odstupu).

Stejně jako argumentační postup i stylistická stránka vykazuje jistý ráz nepoctivosti. Persvazivní rétorické figury provázejí čtenáře celým textem, pro případ, že by se ocitl na pochybách, že tato „práce [...] odvážně překračující hranice oboru“ (s. 12) představuje „radikální badatelský posun“ (s. 34), protože „mnohé z [...] metod jsou relativně nové a v jazykovědné, či dokonce literárněvědné a literárněhistorické analýze textu snad i do jisté míry průkopnické“ (s. 12). Konkrétní kontext citovaných sebehodnotících výroků není snad podstatný, podstatné je, že výroky tohoto typu přísluší pronášet jiné straně než autorské. Leitmotivem práce je pokus situovat své bádání někam mezi ostatní humanitní vědy; toto místo je však tím neurčitější, čím více disciplín je bez bližšího určení zapojeno: na prvních místech lingvistika, historie a literární věda, dále filozofie, sociální a politické vědy ad. Také pojetí subdisciplín hlavních oborů lze označit za přinejmenším vágní, například: „Jde nám o hledání nových cest nejen v historické sémantice, ale také v textologii“ (s. 12) — ačkoli textologii na žádném místě práce pozornost věnována není; pokud jí ovšem autoři nemyslí prostou práci s — textovým — materiálem korpusů. Charakteristické je taktéž definování cílů zkoumání per negationem namísto jejich pozitivního vytčení. A tendenci (nezáměrnou?) k rozostření vlastní badatelské pozice dovršují četné protimluvy, jako např. tento ze dvou sousedních stran: „Ve snaze jednoznačně odkázat na předchozí tradici českého historikosémantického bádání [...] jsme zvolili titul [...]“ (s. 13) a „Bezprostředně ani nenavazujeme na práce českých lingvistů radících se k okruhu historické sémantiky, i když jsme si vědomi tradice i současnosti [...]“ (s. 12). Publikaci provází paradox formálně přehledné argumentace (mechanicky se opakují formule „naším cílem není/je...“ a „analýza ukázala, potvrdila...“ apod.) a nelogických, vzájemně protimluvných, z materiálu neprokazatelných tvrzení. Ta zjevně vycházejí i z potřeby odpovědět kritickým odsudkům (zaznamenaným dle úvodního přiznání řešitele již během publikování dílčích výstupů projektu), poněkud jim vyjít vstříc přiznáním vlastních pochybení a zároveň se pokusit svou práci obhájit pro případ, že její smysluplnost není obsažena implicity.

Otázky nad formálními náležitostmi práce a nad (ne)vstřícností v přístupu autorů (resp. redaktorky) k recipientovi lze pojednat ve světle výše řečeného již stručně. Na jedné straně je text přetížěn redundantními informacemi, na straně druhé pocítujeme nedostatek potřebných informací k tomu, abychom mohli text plynule sledovat, popř. ověřovat. Některá místa textu (zvl. závěry kapitol a podkapitol a závěr konečný, s. 230–234) doslovně opakují místa jiná. Cizojazyčné zdroje jsou citovány výhradně v originálních jazycích (anglicky, německy a francouzsky),³ jazykové kódy střídány i v rámci jediné věty. Vlastní teoretický text doprovází desítky mott a citátů z primární literatury, jejichž přítomnost je nefunkční, zbytečná. Chybí vysvětlivky k některým tabulkovým hodnotám nebo jsou umístěny na jiných místech, než je třeba pro

3 Přítom řešitel projektu děkuje svým kolegům „za překlady cizojazyčných textů“ (s. 9), patrně má ale na mysli překlady (závěrečného českého shrnutí) do cizích jazyků.

porozumění výkladu. Text je přetížen opakovaným odkazováním ke stejným zdrojům (např. s. 67), které však nejsou běžně přístupné nebo přinejmenším dohledatelné.⁴ Autoři pracují s vlastní, odborné veřejnosti nepřístupnou, textovou bankou nazvanou *Historická sémantika* (shrnující — poněkud nesouvisle se svým názvem — novoroční projevy prezidentů, výběrovou publicistiku spisovatelů Čapka, Durycha a Jehličky, vybrané úvodníky *Rudého práva* a vybraná hesla k oslavě májových dnů). Část textové banky je otištěna jako příloha knihy (35 stran) — nemůže-li však být prohledavatelná, plní funkci nadbytečné ilustrace.

Co platného z publikace vytěžit? Zvláště odkaz na zakladatelskou studii Josefa Macka *Historická sémantika*,⁵ dodnes aktuální práci, podnětnou erudovaným mezinoborovým přístupem, metodologickým rozvrhem i jednotlivými mikroanalýzami. Josef Macek mj. zmiňuje možné využití kvantitativních metod pro historickosémantické bádání. A právě z jeho studie by měli autoři publikace *Slovo a text v historickém kontextu* čerpat důsledněji: zvláště a bezpodmínečně v diferencování mezi historickou realití, jejím označením a pojmovým zpracováním, a dále například v uvažování o moderních (politických) mýtech na základě vztahu významů, pojmů a stereotypů. Protože v mýtotvorných stereotypch sami autoři pohřchu uvízli.

AD:

Jaroslav David, Radek Čech, Jana Davidová Glogarová, Lucie Radková a Hana Šústková: *Slovo a text v historickém kontextu — perspektivy historickosémantické analýzy jazyka*. Host a Filozofická fakulta Ostravské univerzity v Ostravě, Brno a Ostrava 2013. 323 s.

4 Srov. např. položku pramenného soupisu: Arbes, Jakub (2010). Jakub Arbes [textový korpus, dostupný na CD]. Praha.

5 Zničenou ve stránkové korektuře v roce 1970 a vydanou až v roce 1991 (*Český časopis historický* 89, 1991, č. 1, s. 1–30).