

Otevřenost jako norma

Openness as a norm

Radek ČECH | Katedra českého jazyka FF OU

Článek Václava Cvrčka a Jana Chromého vnímám také (možná především) jako apel na institucionální podporu otevřenosti a transparentnosti lingvistiky (a vlastně nejen jí). Názory, které zde autoři prezentují, mi osobně přijdou natolik jasné, že mám pocit, že není moc o čem diskutovat. Vždyť co lze namítat proti tomu, aby byly výsledky práce, která je placena z veřejných peněz, veřejně dostupné? Nebo proti tomu, aby byl výzkum replikovatelný?

Mnozí z nás ale mají neblahou zkušenost jak s jednotlivci, tak s institucemi, kterým jsou principy otevřenosti a transparentnosti cizí – čas a jistě nemalé úsilí, které někteří badatelé strávili s přípravou dat, u nich zřejmě vyvolávají pocit, že jsou v jistém smyslu jejich majiteli a že není možné je jen tak někomu dát k dispozici (např. dokud nejsou dostatečně „vytěžena“ ve vlastních studiích). V těchto případech je to právě institucionální tlak, který může situaci změnit – ať už se jedná o tlak ze strany vedení institucí, nebo ze strany časopisů, které odmítnou publikovat text, u kterého nebudou veřejně dostupná data; viz apel autorů článku na časopisy (Cvrček – Chromý, s. 12–13). V tomto ohledu je článek *Lingvistika jako otevřená a transparentní disciplína* velmi užitečným východiskem, který dává všem, již mají snahu situaci změnit (např. jako členové redakčních rad), ucelenou argumentační oporu.

Za zásadní bod, který by zejména z dlouhodobé perspektivy vedl ke zlepšení současného stavu, považují požadavek autorů na to, aby akademická pracoviště měla „jasná pravidla sdílení dat a dalších materiálů ze závěrečných prací (bakalářských, diplomových i disertačních) svých studentů“ (Cvrček – Chromý, s. 13). Musím se přiznat, že mě doposud nenapadlo uvažovat o této problematice právě z této perspektivy. Jedná se přitom o velmi účinný nástroj, který by nejen dokázal vést studenty k tomu, aby se pro ně transparentnost a vše, co s ní souvisí, staly stejným standardem jako například řádné citování literatury, ale měl by zajisté vliv i na akademiky a jejich vlastní práci.

Autoři článků zmiňují několik příkladů dobré praxe, které jsou inspirací pro to, jak se moderním způsobem s danou problematikou vyrovnat. Je však jasné, že zásadní problém není technický, ale že spočívá spíše v postojích, praxi a celkové atmosféře oboru. Uvažují-li o problematice otevřenosti v lingvistice z většího nadhledu, kladu si otázku, zda by celkové atmosféře nepomohlo, kdyby nejvýznamnější české lingvistické časopisy (alespoň tedy ty, které vydávají veřejné instituce, zejména Akademie věd ČR) nepřešly do režimu open access, v němž by byl veškerý obsah volně přístupný všem čtenářům (nemám přitom na mysli režim open access, v němž jsou požadovány po autorech horentní sumy za publikování článků, což je bohužel u mnohých časopisů běžnou praxí). Ačkoliv nemám žádné informace o tom, kolik časopisy jako *Naše řeč*, *Slovo a slovesnost*, *Acta onomastica* či *Listy filologické* vydělávají prodejem, nepřepokládám, že by šlo o sumy, které hrají v rozpočtu instituce zásadní roli. Mimochodem, současná situace, kdy, pokud vím, neexistuje možnost si předplatit tyto časopisy v elektronické formě přímo u vydavatele, sdílení informací v širším slova smyslu také nepřispívá (uživatelsky velmi nepřívětivý způsob, jak se dostat k jednotlivým článkům za poplatek prostřednictvím Central and Eastern European Online Library, nechávám raději bez komentáře).

Jsem upřímně zvědavý na případnou diskusi o tomto tématu. Pro prosazení idejí otevřenosti a transparentnosti do praxe však budou důležitější konkrétní kroky. Věřím ale tomu, že se nám za pár let bude zdát nechuť sdílet data, prezentovat postupy atd. stejně nepochopitelnou, jako se nám

dnes zdá nepochopitelné, že jsme například kdysi jezdili autem, aniž bychom používali bezpečnostní pásy.

Katedra českého jazyka FF OU
Reální 5, 701 03 Ostrava
radek.cech@osu.cz

Zpřístupňování dat z výzkumu s veřejnou podporou: ne zda, ale jak nejlépe!

Making publicly funded data accessible: not “if”, but “how best”!

Silvie CINKOVÁ | Ústav formální a aplikované lingvistiky MFF UK

S velkým potěšením reaguji na výzvu Václava Cvrčka a Jana Chromého českým lingvistům, aby v největší možné míře sdíleli svá data v repozitářích k tomu určených. Naprosto souhlasím s tím, že výzkum financovaný z veřejných peněz musí zůstat veřejně přístupný. Jsem toho názoru, že výsledek veřejně financovaného výzkumu, na který získá výhradní právo například komerční nakladatel nebo si na něj osobuje výhradní právo řešitel grantu, nebo který ještě několik let po skončení projektu není publikovaný vůbec, neodporuje jenom dobré praxi, ale ani obecněji smyslu veřejné podpory výzkumu a poskytovatel financování by projekt s takovým výsledkem neměl uznat za splněný.

Členové našeho pracoviště (ÚFAL MFF UK) jsou zváni k mnoha zajímavým humanitním projektům. Rádi a často spolupracujeme také neformálně, a to i se studenty všech úrovní studia: většinou pomáháme s automatickým zpracováním přirozeného jazyka nebo s datovou analýzou. Protože jsme však závislí na grantech, a tedy pod neustálým publikačním tlakem, upřednostňujeme spolupráci, která alespoň potenciálně vede ke spoluautorství adekvátní publikace v rozumné době – a tím jsme opět u zmíněné ochoty výzkumníků zveřejnit svůj datový soubor.

LINDAT/CLARIAH-CZ, infrastruktura, kterou spravuje naše pracoviště, bezplatně nabízí celé lingvistické obci své úložné kapacity. Ukládání dat je technicky velice jednoduché – provádí se přes webové rozhraní. Nevyžaduje konverzi dat do složitých formátů jako XML-TEI. Pokud jste svoje data jakkoli vyhodnocovali automaticky (třeba počítali nějaké frekvence), pak je jejich výsledný formát téměř jistě „FAIR“ (Cvrček – Chromý, s. 6). „FAIR“ je dokonce i excelová tabulka uložená jako text (formát „.csv“ nebo „.tsv“) s pojmenovanými sloupci, pokud jejich význam vysvětlíte ve stručné dokumentaci v dalším souboru.

Kolegové Cvrček a Chromý se ve svém textu soustředili zejména na data, která produkují sami badatelé, a jen letmo se dotkli autorsky chráněných textů (např. v korpusech ČNK; Cvrček – Chromý, s. 11). Donedávna byly texty korpusů ČNK přístupné pouze v úryvcích (KWIC) přes webové rozhraní. Odvážným a velmi chvályhodným krokem ÚČNK je zveřejnění licencovaných datových balíčků s promíchaným pořadím vět či odstavců. Pro některé výzkumné úlohy (např. stylometrie) však ani to nestačí a zatím neexistuje způsob, jak autorsky chráněné texty legálně získat a dát k dispozici alespoň recenzentům výsledné publikace.