

Koncepce rozvoje Ústavu Českého národního korpusu FF UK pro období 1. 2. 2022 – 31. 1. 2024

Předkládaná koncepce navazuje na dvě předchozí z pera kol. Michala Křena, jenž ÚČNK vedl – a mám za to, že zdárně – v letech 2016–21. Hlavní prioritou této koncepce je udržet kontinuitu projektu Český národní korpus („ČNK“), tedy jeho finanční zajištění, ale též institucionální zastřešení v nadcházejících letech. ČNK je bezmála tři dekády budovaná výzkumná infrastruktura, která dnes dobře slouží svým (zdaleka nejen českým) uživatelům a která si v průběhu času vydobyla nemalé renomé doma i v zahraničí. Už jen pohled na četnost a pestrost různých korpusů a na bohatou nabídku nejrůznějších nástrojů a aplikací k vytěžování korpusových dat opravňuje tvůrce k sebevědomému postoji, zvláště srovnáme-li naše výsledky s portfolií obdobných pracovišť v zahraničí.

Role vedoucího ÚČNK – jakožto jedné ze základních součástí FF UK – však zahrnuje především neinfrastrukturní vědu a výzkum spolu s výukou. Tento synergický vztah mezi ÚČNK a infrastrukturou podtrhuje zvláštní status našeho ústavu v rámci FF: kromě vyšší míry finanční „soběstačnosti“ oproti jiným katedrám je to vyšší zastoupení neakademických, tj. vědeckých a technických, pracovníků. (Zvláště bez TP bychom nebyli schopni poskytnout svým uživatelům dostatečnou podporu.) Je rovněž hlavním důvodem pro to, že v případě svého zvolení bych kol. Křena, jako vedoucího infrastrukturního projektu ČNK, rád učinil svým zástupcem.

Aktuálně se ÚČNK podílí na řešení dvou grantů (a v několika dalších je spoluřešitelem¹): Český národní korpus, jeden z Projektů velkých infrastruktur pro VaVal (LM2018137, 2019–2022) a projekt Rozšíření výpočetní kapacity infrastruktury CNC² (operační program Výzkum, vývoj a vzdělávání, CZ.02.1.01/0.0/0.0/18_046/0015829, 2020–2022). Z nich aktuálně plynou téměř tři čtvrtiny rozpočtu ÚČNK, a je tedy klíčové zajistit, aby ústav, chce-li zachovat míru i kvalitu svých aktivit i nadále, získal návazné zdroje financování, včetně těch neinfrastrukturních. To považuji za prioritu příštího ředitele ÚČNK, přičemž si uvědomuji obtížnost tohoto úkolu. Vzhledem k předpokládanému snížení podpory ze strukturálních fondů EU je třeba počítat od začátku r. 2023 i s méně příznivými scénáři a připravit se na ně – a to i navzdory faktu, že v letošním mezinárodním hodnocení infrastruktur jsme byli velmi dobře hodnoceni. Vedení ústavu proto plně podpoří podávání vícero grantů, kvůli kýžené diverzifikaci finančních prostředků, která by v případě výpadku či zásadního omezení primárního finančního zdroje znamenala pro ÚČNK existenční hrozbu.

Jako příklad zde můžu uvést dva projekty, jež vycházejí z mého osobního zájmu (podobnou soukromou iniciativu však vítám i od všech svých kolegů). Jde jednak o (již zdárně započatý) Korpus současné poezie, který představuje celosvětový unikát mezi korpusy svého druhu, jednak o dokončení torza lužickosrbsko-českého slovníku Jiřího Mudry; v tom chci navázat na dřívější, aktuálně ladem ležící lexikografický potenciál ÚČNK. Oba projekty jsou nadústavní, spjaté se spoluprací s ústavu AV ČR (v prvním případě Ústav české literatury, v druhém Slovanský ústav).

¹ Jejich výčet i se stručným popisem viz <https://ucnk.ff.cuni.cz/cs/veda-a-vyzkum/granty/>

² CNC (Czech National Corpus) = anglický akronym pro ČNK.

Personální oblast

Jsem přesvědčen, že se plně osvědčilo rozdělení ÚČNK do jednotlivých sekcí (lingvistická, technická, mluvená, diachronní, paralelní a anotační), jejichž vedoucí jsou mezičlánkem mezi vedením ústavu a řadovými pracovníky. Stejně tak podporuji pravidelné setkávání sekčních vedoucích s vedením (nejméně jednou za čtvrt roku), tyto porady jsou dobré nejen ke koordinaci aktivit jednotlivých úseků (včetně mezisekční spolupráce), ale i k celkovému směřování ústavu. Uvítal bych větší informovanost řadových zaměstnanců o tom, co se děje v jiných sekcích; vhodnou platformou se mi pro to jeví náš interní seminář – na něm by měla každá sekce periodicky představovat témata, jimž se aktuálně věnuje.

Dlouhodobé manko vidím v nízkém počtu pracovníků s vyšším akademickým titulem, byť i zde se situace poslední dobou obrací k lepšímu. Aktuálně působí v ústavu 1 profesor (v důchodovém věku), 1 docent a 12 doktorů (z toho 2 na rodičovské dovolené). Jeden pracovník úspěšně zahájil jmenovací řízení; habilitační přihlášku by měli v dohledné době podat 3 kolegové.

Hodlám – ve shodě s připravovanými kariéerními předpisy – všestranně podporovat snahu kolegů habilitovat se (např. formou pracovních úlev, sabbaticku apod., příp. též finančně), a pokud mi to pracovní povinnosti dovolí, sám bych rád šel příkladem a do konce r. 2023 zahájil habilitační řízení. Mimoto chci motivovat ke zdárnému dokončení disertační práce i doktorandy, jež chci i více zapojit do samotného chodu ústavu tak, aby se cítili součástí našeho týmu.

V neposlední řadě hodlám podporovat spolupráci s jinými spřízněnými pracovišti, resp. navazovat nové kontakty tam, kde ještě nejsou. Zvláště pro získávání mezinárodních grantů je však důležité neomezit se na český prostor, ale rozvíjet kontakty a spolupráci též se zahraničím. To už se venkoncem daří díky projektu CLS INFRA, jehož nedílnou součástí jsou mj. stáže zahraničních kolegů.³ Doufám, že náš ústav pro ně bude atraktivní destinací i nadále; vidím jej jako otevřené a vstřícné pracoviště, ochotné podělit se o svá data i know-how a spolupracovat s kolegy, kteří o to projeví zájem.

Pedagogická činnost

Situace oproti minulým letům zůstává stejná: ÚČNK nemá vlastní pregraduální program (a tudíž ani žádné kmenové studenty), významně se však podílí na sylabech zvláště filologických oborů, mj. nově akreditovaného programu Empirická a komparativní lingvistika. Kurz Úvodu práce do jazykových korpusů („ÚPJK“) je součástí společného základu pro většinu filologií, je vypisován každý semestr a aktuálně (LS 2021/22) jej navštěvuje přes 120 studentů. Ostatní kurzy – v každém semestru jich je vysáno asi 10 – jsou volitelné, ale mezi studenty je o ně dlouhodobě zájem a také evaluace mluví vesměs v náš prospěch. V budoucnu nelze očekávat žádné zásadní změny, s výjimkou občasného střídání témat podle aktuálního pedagogova zájmu (co největší variabilita je jistě hodna podpory); naše síly bych proto napnul do vylepšování výše zmíněného ÚPJK a také do rozšiřování výukových materiálů na korpusovém portálu (zejm. v rámci korpusové wiki) a na našem YouTube kanálu (viz též níže).

Doktorské studium v oboru Korpusová a teoretická lingvistika prošlo v nedávné době obsahovými revizemi a získalo novou akreditaci. Omezený počet doktorandů – řádově jednotky (aktuálně 10 studentů, z toho 2 mají studium přerušeno a 2 by jej měli ukončit v tomto akademickém roce) – nechápu jako nedostatek, byť i zde by se lepší propagací a včasným podchycením potenciálních zájemců o studium snad dalo leccos zlepšit. V centru doktorandovy pozornosti nyní stojí disertace, která má vznikat nikoliv nárazově, ale kontinuálně a pokrok na ní je pravidelně kontrolován.

Vedle studentů cílíme rovněž na laickou veřejnost, které už mnoho let nabízíme pravidelné (2× ročně konané) workshopy. Rád bych do budoucna změnil jejich koncepci a profiloval je s ohledem na

³ <https://clsinfatna.sciencescall.org/>

pokročilé uživatele, kterým už není třeba vysvětlovat, co je to korpus a k čemu slouží, ale lze s nimi probírat složitější problematiku (CQL dotazy, syntax, specializované výzkumné otázky). Začátečníky by měl do dané oblasti uvést (kromě wiki) náš YouTube kanál, který před posledním workshopem doznal zásadních změn a který hodlám nadále plnit dalšími videotutoriály i záznamy přednášek a workshopů.

Vědecká činnost

ÚČNK dlouhodobě vykazuje solidní objem vědecké i publikační činnosti (vč. renomovaných zahraničních zdrojů). Z tohoto standardu není pochopitelně žádoucí slevit, už proto, že příznivě ovlivňuje naše šance získávat další finanční zdroje. Po vědecké stránce je ÚČNK myslím vyprofilován dostatečně, proto chci toto směřování podporovat i nadále. Díky svému infrastrukturnímu pověření je ÚČNK původcem celé řady různých nástrojů a rozhraní, ty je potřeba vytvářet s ohledem na cílového uživatele, tedy aby byly pokud možno rychlé, intuitivní a uživatelsky vstřícné; samozřejmostí je solidní dokumentace (tuto roli plní naše wiki) a uživatelská podpora. Rovněž neinfrastrukturní výzkum hraje důležitou roli: sami nejlépe znající své vlastní nástroje měli bychom ukazovat ostatním vědcům, jak správně s nimi pracovat, inspirovat je metodologicky, prošlapávat dosud neobjevené cesty.

Propagace

Záměrně zařazuji explicitně tento bod, protože v této oblasti cítím jisté manko, jakkoliv se ÚČNK o propagaci upřímně snaží; to ostatně dosvědčuje i interní sdílený dokument, kde jsou všechny propagační aktivity zaznamenávány. (Čítá desítky položek, jen v letošním roce se jedná o sérii sloupků pro časopisy *Vesmír* a *Tvar* a popularizačních článků pro portál reflex.cz, rozhovor pro ČRo2, vystoupení v rámci Noci vědců aj.)

Měli bychom využít každou možnost, abychom společnosti nabízeli tolik potřebnou alternativu k všelijakým preskriptivním názorům, pěstovali v lidech sebevědomí (a sebe-odpovědnost) rodilého mluvčího a upozorňovali je na výhody (ale též limity) empirického přístupu. Vedle osvědčených forem (rozhovory pro média, články a jazykové sloupky) bych rád ovlivňoval veřejné mínění ještě dalšími způsoby (blog, podcast, desková hra aj.). V úvahu připadá také kolektivní popularizační kniha, která by uvedla čtenáře do světa jazykových korpusů a korpusové lingvistiky; koneckonců takový počin jsme si v malém vyzkoušeli díky seriálu článků pro *Lidové noviny*.

Dlouhodobě sleduji také fenomény crowdsourcingu a gamifikace, který může být nám vědcům prospěšný⁴, a chci ho zohlednit i v budoucím směřování ústavu.

Resumé

Shrnuji zde priority svého případného působení ve funkci ředitele ÚČNK (byť se z větší části překrývají s dalším rozvojem infrastruktury, a tudíž vyžadují shodu v širším realizačním týmu):

- rozšiřování dat psaných a především mluvených (cílem magická hranice 10M) a jejich vzájemná konfrontace (téma diglosie);
- rozvoj specializovaných korpusů (poezie, internetový jazyk – se zaměřením na podchycení neologie);
- dokončení monitorovacího korpusu češtiny, propojení dat diachronních se synchronními;
- podpora uživatelských korpusů;
- další úrovně anotace (Universal Dependencies – komparatistický potenciál), vylepšování těch stávajících (zapojení neuronových sítí);

⁴ Vycházím z vlastní zkušenosti lexikografa, kde – jak jsem pevně přesvědčen – laická veřejnost může vykonat mnoho dobrého a prospěšného – a grátis!

- podpora rozhraní KonText jakožto „rodinného stříbra“ (exportní potenciál), taktéž WaG s jeho flexibilitou (odvozená rozhraní typu Slovo v poezii);
- obohacování našeho YouTube kanálu o nové videotutoriály, další rozvoj wiki;
- spolupráce se spřízněnými pracovišti v Česku i zahraničí;
- posílení PR ústavu navenek.

V Praze dne 29. 11. 2021
Mgr. Michal Škrabal, Ph.D.