

## **Koncepce rozvoje Ústavu Českého národního korpusu FF UK na období 1. února 2016 - 31. ledna 2019**

Hlavním úkolem Ústavu Českého národního korpusu („ÚČNK“) je bezesporu péče o kontinuální a všestranný rozvoj projektu Český národní korpus („ČNK“). Tento projekt je centrálním bodem aktivit ústavu, přesahuje však jeho rámec tím, že jsou do práce na něm zapojena také další pracoviště, zejména kolegové z Ústavu teoretické a počítačové lingvistiky („ÚTKL“); na druhou stranu i aktivity ÚČNK jsou pochopitelně širší a neomezují se pouze na projekt ČNK.

Činnost ÚČNK je možné rozdělit na dvě hlavní složky, servisně-technickou a vědecko-pedagogickou. Náplní první z nich je mapování češtiny prostřednictvím budování jazykových korpusů, při němž jde především o sběr, zpracování a zveřejňování dat, jejich anotaci na různých jazykových úrovních a související servis pro odbornou i laickou veřejnost.

Složka vědecko-pedagogická se zabývá především rozvojem metodologie využívání korpusů pro výzkum (nejenom) češtiny, vlastní vědeckou a publikační činností a v neposlední řadě i výchovou studentů. Poskytuje také zpětnou vazbu pro další vývoj a vylepšování nástrojů pro uživatelskou práci s korpusovými daty, které realizuje složka servisně-technická. Obě složky jsou tedy komplementární, vzájemně se doplňují a obohacují.

Z dosud uvedeného je zřejmé, že ÚČNK je pracoviště, jehož charakter je ve srovnání s ostatními základními součástmi FF UK netypický, a tato netypičnost se pochopitelně projevuje také v předkládané koncepci. Jde především o vysoký podíl vědeckých a technických pracovníků, kteří zajišťují servisně-technickou složku činnosti ÚČNK, a také o způsob financování: existence ÚČNK totiž byla a je v rozhodující míře závislá na externích zdrojích podporujících především další rozvoj projektu ČNK.

Od založení ÚČNK v roce 1994 se pod vedením jeho zakladatele a dlouholetého ředitele prof. Františka Čermáka podařilo ústav vybudovat prakticky z ničeho, neustále ho rozvíjet, rozšiřovat jeho záběr a samozřejmě také získávat na jeho provoz odpovídající prostředky. Toto mnohaleté úsilí bylo završeno získáním projektu *Český národní korpus* v rámci *Velkých infrastruktur pro výzkum, vývoj a inovace* MŠMT (LM2011023, 2012-2016; „infrastruktura“), z něhož v současné době pochází více než 90 % rozpočtu ústavu. V tomto období došlo k řadě významných změn: zavedení nového loga a grafického stylu projektu ČNK, přechod na nové jednotné rozhraní pro práci s korpusy KonText, spuštění portálu korpus.cz společně s řadou nových prvků uživatelské podpory (wiki, poradna, biblio), ustavení Vědecké rady ČNK atd. Na základě širokých diskusí na jaře 2014, do nichž byli zapojeni také kolegové z ÚTKL, pak vznikla *Koncepce rozvoje projektu ČNK pro období do roku 2022* („Dlouhodobá koncepce“). Tyto diskuse už probíhaly v nových prostorách v Panské 7, kde ÚČNK sídlí od podzimu 2013.

Pro blízkou budoucnost je však nejpodstatnější fakt, že během roku 2014 proběhla rozsáhlá evaluace infrastruktur, v níž získal projekt ČNK velmi dobré hodnocení. Na jeho základě byl zařazen na aktualizovanou *Cestovní mapu ČR Velkých infrastruktur pro výzkum, experimentální vývoj a inovace*, čímž splnil nutnou podmínku pro další financování v následujícím období, tj. až do roku 2022. Po dlouhých letech budování projektu a obhajování jeho existence tak doufejme přichází období relativní stability, kdy bychom měli mít dostatek prostředků na další rozvoj projektu i ústavu v oblastech, které považujeme za důležité.

Přes tento optimistický výhled je potřeba mít neustále na paměti, že budoucnost je vždy více či méně nejistá. Nejistý je další vývoj situace ve vysokém školství a ve financování vědy, v postavení humanitních oborů atd. Nejasná je však také situace ve financování ÚČNK: ve chvíli, kdy vzniká tato koncepce, stále ještě nevíme, jaké prostředky bude mít infrastruktura ČNK v dalších letech k dispozici. Jasně v tuto chvíli není ani co se týče dalších předpokládaných zdrojů financí (zejména OP VVV a Prvok, neznáme ani výsledky projektů podaných do NAKI). Protože by nebylo účelné předkládat již nyní konkrétní varianty dalšího vývoje, drží se tato koncepce spíše obecné roviny a nastavení takových mechanismů, které umožní na případné změny pružně reagovat.

### **Personální oblast**

Jak již bylo řečeno, pracovní kolektiv ÚČNK tvoří z velké části neakademičtí, vědečtí a techničtí pracovníci. Zároveň jde o kolektiv převážně mladý, což je vzhledem k pracovnímu elánu a chuti inovovat nesporná výhoda; za další přednost považují různorodé profesní zaměření jednotlivých členů ústavu. Na druhou stranu ale citelně chybějí pracovníci s vyššími akademickými tituly, členy ústavu budou v příštím roce pouze jeden profesor (v důchodovém věku) a jeden docent. I pokud vezmeme v úvahu specifický charakter pracoviště, situace v tomto ohledu rozhodně není uspokojivá.

Budu proto všestranně podporovat osobní a kvalifikační růst akademických a vědeckých pracovníků, zejména ve směru k dalším habilitacím. Je to nutné nejenom kvůli akreditacím, v nejbližší době můžeme očekávat také nový kariérní řád a zvýšený tlak na kariérní postup vůbec. Měli bychom se na to připravovat hlavně v nadcházejícím období, kdy by měla být příhodná situace (konkrétnější návrhy uvádím dále).

Činnost projektu ČNK je rozdělena do šesti sekcí: sekce lingvistické, sekce počítačnické (technické), sekce mluvených korpusů, sekce diachronních korpusů, sekce paralelních korpusů a sekce lingvistické analýzy a anotace. Toto rozdělení se podle mého soudu osvědčilo, sekce považuji za funkční způsob rozdělení činností v rámci projektu. Přesto myslím, že řada činností probíhající napříč sekcemi je koordinována na vyšší úrovni, než by bylo nutné. Vidím také jisté rezervy v informovanosti zejména řadových pracovníků o tom, co konkrétního se v jiných sekcích děje. Další motivací, která mě vede k vylepšením navrhovaným dále, je větší zapojování jednotlivých pracovníků do úkolů,

díky nimž budou moci pronikat do způsobu práce v ostatních sekcích a získávat tak organizační zkušenosti. Organizaci práce chci proto vylepšit v následujících bodech:

- Opětovné využití interního semináře mj. jako fóra, na kterém se budeme navzájem informovat o tom, co je nového a co se v jednotlivých sekcích chystá.
- Využití již existující platformy schůzek vedoucích sekcí také pro praktickou provozní agendu, větší koordinaci činnosti sekcí, průběžnou kontrolu výstupů a rozhodování o prioritách dalšího směřování projektu. Bude to znamenat častější a delší schůzky, věřím však, že půjde o dobře investovaný čas a úsilí.
- Vedoucí sekcí by se podle mého názoru měli ve svých funkcích (tam, kde je to možné) střídat s řadovými pracovníky. Nejde o projev nedůvěry ke stávajícím vedoucím, naopak; právě proto, že vykonávají náročnou a zodpovědnou funkci, měli by mít možnost vzít si „oddechový čas“ a věnovat se více vlastnímu výzkumu a kariéernímu růstu. Nenavrhuji konkrétní mechanismus, chci si ale na toto téma s jednotlivými vedoucími promluvit a zkusit najít cestu, jak to v jednotlivých případech udělat.
- V případě konkrétních cílů či výstupů vyžadujících rozsáhlejší činnost jdoucí napříč sekcemi chci do budoucna využívat samostatné pracovní skupiny. Vznikat budou pouze v případě potřeby pod vedením koordinátora, jehož úkolem bude (po dohodě s vedoucími dotčených sekcí) zajistit, že práce bude odvedena kvalitně a včas; nedílnou součástí koordinace bude hlídání návaznosti jednotlivých činností překračujících hranice sekcí. Koordinátoři tak budou přirozeným způsobem získávat zkušenosti s vedením, zároveň tak uděláme další krok k výchově nových vedoucích sekcí a také potenciálních nástupců ve vedení projektu.

Tyto organizační změny nejsou nijak zásadní, věřím však, že se díky nim podaří zvýšit zainteresovanost na chodu projektu i ústavu, vědomí sounáležitosti a pracovní motivaci, a zároveň udržet dobré mezilidské vztahy.

### **Pedagogická činnost**

ÚČNK ve spolupráci s ÚTKL realizuje doktorský studijní program „matematická lingvistika“, který je akreditován do 1. 12. 2017. Již brzy bude tedy potřeba začít pracovat na akreditaci nové, jejíž přípravu chci koordinovat ve spolupráci s členy oborové rady. Už teď je zřejmé, že reakreditace bude spojena s výraznými změnami, za důležité považuji především zvýšení důrazu na schopnost studentů prakticky pracovat s jazykovými daty a větší zapojení doktorandů do chodu pracoviště. Měli bychom také zvážit akreditaci oboru v angličtině.

Kromě doktorského studia nabízí ÚČNK úvodní semináře korpusové lingvistiky pro Bc. a Mgr. studenty některých filologických oborů FF, a také několik seminářů specializačních. Tato praxe se v posledních letech rozvinula a považuji ji za dobrou, výrazný další nárůst počtu vyučovaných předmětů by už ale myslím nebyl účelný. Iniciativu jednotlivců ovšem nijak omezovat nechci, výuka je koneckonců nutným předpokladem habilitace.

Vzhledem k již zmíněné specifčnosti projektu ČNK je podíl výuky jednotlivých pracovníků ÚČNK až na výjimky relativně malý. Vzhledem k našemu působení na fakultě je však výuka potřeba pro kariérní postup, který je po zisku titulu Ph.D. vázán i na pedagogickou činnost včetně vedení závěrečných prací. Nemáme ale jiné vlastní studenty než doktorandy, které pracovníci bez habilitace mohou vést jen v omezené míře, a tak vzniká začarovaný kruh. Protože akreditaci vlastního magisterského nebo bakalářského oboru nepovažují za realistickou, je další možností zapojovat se kromě výuky pro jiné obory také do vedení Bc. a Mgr. prací studentů těchto oborů. Je to sice doslova práce navíc, může ale být vykompenzována zapojením našich doktorandů do plnění úkolů souvisejících s projektem.

Poslední bod se týká workshopů pro odbornou i laickou veřejnost, které mívají velmi dobrý ohlas a které se prozatím konají pouze nepravidelně. Chci se zasadit o jejich pravidelné konání (nejlépe 2x ročně, na jaře a na podzim), protože jsou velice oblíbené mezi VŠ i SŠ pedagogy, z pohledu infrastruktury jde o servisní činnost a je to pro ČNK vhodný způsob propagace.

### **Vědecká činnost**

Vědeckou činnost ÚČNK můžeme rozdělit na dvě oblasti: infrastrukturní a neinfrastrukturní. Infrastrukturní vědecká činnost je taková, která vede k rozvoji a vylepšování služeb poskytovaných infrastrukturou ČNK a jako taková může být hrazena z jejího rozpočtu. Naopak vědecká činnost, která zdrojů a služeb infrastruktury pouze využívá, z něho být hrazena nemůže.

V oblasti infrastrukturní vědecké činnosti chci dbát na sledování vývoje v oblasti technických standardů a nástrojů tak, aby zpracování dat probíhalo co nejefektivněji. Všechny nástroje vyvinuté v rámci ČNK, které mají potenciál pro využití i mimo něj, by měly být koncipovány obecně tak, aby mohly být zveřejněny a používány co nejširší komunitou; to může přinést projektu mj. nové mezinárodní spolupráce.

V oblasti neinfrastrukturní přicházejí v úvahu zejména hlavní směry zmíněné v Dlouhodobé koncepci: moderní diachronie, mediální a diskurzní analýza, syntagmatika, jazyková variabilita, kontrastivní a translátologický výzkum. Budu proto podporovat snahu o získání dalších externích finančních zdrojů, které budou pro jejich realizaci potřebné.

Kvalitní vědecká činnost v obou těchto oblastech je důležitá pro osobní rozvoj každého pracovníka i pro hodnocení pracoviště, je nezbytná také pro získávání dalších zdrojů externího financování. Ačkoli se může zdát, že ÚČNK není díky infrastruktuře na hodnocení vědy tolik závislý, jde pouze o přechodný stav, který je ovšem potřeba využít.

Pracovní vytížení akademických a vědeckých pracovníků by tedy mělo poskytovat dostatek času na vlastní výzkum. Především těm, kteří se věnují sběru a zpracování dat, by měla pomoci preference variability a kvality dat před kvantitou, která je již na úrovni celého projektu uplatňována. Právě vytěžování dat spojené s jejich detailní znalostí je dobrým předpokladem pro kvalitní výzkum na těchto datech založený. Chci také apelovat

na publikace v časopisech zařazených do mezinárodních databází – ty budou i nadále přinejmenším preferovány, ať už se metodika hodnocení vědy bude vyvíjet jakkoliv.

S vědeckou činností souvisí také ediční politika ČNK. Ta prošla v letošním roce reformou, která nastavila pravidla pro publikování titulů s finanční spoluúčastí projektu. Tuto změnu podporuji a nehodlám v tomto ohledu nic podstatného měnit.

### **Další oblasti rozvoje a výhled do budoucna**

Vzhledem k tomu, že rozhodující část financí na provoz ÚČNK pochází v současné době z infrastruktury, jsou priority jeho činnosti v dalších letech do značné míry dány právě tímto projektem. Konkrétně jsou pro výhled do budoucna důležité především dva dokumenty: závazný oficiální návrh projektu infrastruktury a konkrétnější interní Dlouhodobá koncepce, z níž návrh projektu vycházel. Ta sice závazná není, prošla ale rozsáhlou vnitřní oponenturou a proto ji považuji za důležité vymezení dalšího směřování projektu, a tedy i celého ústavu. To ovšem na druhou stranu neznamená, že by bylo možné se Dlouhodobé koncepce beze změny držet; bude jistě potřeba reagovat na aktuální situaci a přizpůsobovat jí v dokumentu obsažené priority, případná aktualizace však bude předmětem jednání vedoucích sekcí. Osobně vidím pro nejbližší období jako důležité hlavně následující body:

- důkladné zvážení priorit dalšího rozvoje KonTextu na stávajícím backendu, kde bude potřeba vybírat z řady možností (grafická tvorba dotazů, pokročilé kolokační moduly, vícerozměrná frekvenční distribuce s vizualizací, podpora mobilních zařízení atd.);
- rozšíření spektra sběru dat na oblasti dosud nepokryté, především specifický internetový jazyk, formálnější jazyk mluvený a postupné vyplňování mezery mezi psaným korpusem synchronním a diachronním;
- projekt využití korpusů při výuce češtiny na středních a základních školách, který považuji za potřebný také pro zvýšení povědomí o ČNK mimo akademické prostředí.

Na obecné rovině chci usilovat o budování „značky“ ČNK jako univerzálního zdroje dat, nástrojů, metodologie a podpory pro empirickou práci s jazykem ze strany odborné i laické veřejnosti, jako solidního a spolehlivého partnera, bez něhož si lingvisticky zaměřený výzkum nebo výuku bude možné jen stěží představit.

V Praze dne 30. listopadu 2015

Mgr. Michal Křen, Ph.D.