

# EXKURZE DO ÚSTAVU ČESKÉHO NÁRODNÍHO KORPUSU FF UK V PRAZE

**Datum konání:** 3. 3. 2023

**Místo konání:** Ústav Českého národního korpusu Filozofické fakulty Univerzity Karlovy v Praze

**Počet účastníků:** 12

**Zpracoval:** Krňávek David

Dne 3. března 2023 proběhla exkurze do **Ústavu Českého národního korpusu Filozofické fakulty Univerzity Karlovy v Praze** za doprovodu Mgr. Hany Žižkové, Ph.D. Ústav Českého národního korpusu sídlí v Kounickém paláci v Panské ulici na Novém Městě v Praze. Za vstupními dveřmi do prostor ústavu čekalo brněnské studenty pracovní prostředí s velmi uvolněnou atmosférou. Pracovníci ústavu studentům na počátku exkurze nabídli kávu nebo čaj dle výběru, došlo také na zákusky a zdravou stravu v podobě mrkve.

Historií a současností ÚČNK nás provedla PhDr. Mgr. Marie Kopřivová, Ph.D., ze sekce mluvených korpusů. V úvodu stručně popsala historii ústavu, která se začíná psát roku 1994. Ústav založil prof. PhDr. František Čermák, DrSc., o němž padla zmínka hned v úvodu exkurze. Počátky ÚČNK byly velmi skromné, první korpus měl přibližně 35 000 slov a byl obsluhován pracovníky ústavu prostřednictvím příkazového řádku. Konkordanci ústav poskytl na vyžádání. V současnosti čítá rodina korpusů ÚČNK mnoho korpusů různých druhů (psané, mluvené, synchronní, diachronní atd.), z nichž největší, korpus *SYN verze II* skládající se ze synchronních korpusů od roku 2000 (*SYN2000*) do roku 2020 (*SYN2020*) a dalších publicistických textů vydaných po roce 2020, má 5 miliard slov. Kromě korpusů češtiny vznikly v ÚČNK také korpusy obsahující texty v jiných jazycích, např. paralelní korpusy *InterCorp* obsahující shodné texty v překladech do různých jazyků nebo korpus *Žalm 77*, zaměřený zejména na rumunský překlad žalmu 77.

Dnes je vygenerování konkordance libovolného slova v libovolném korpusu ÚČNK otázkou několika kliknutí v korpusovém manažeru *KonText*. Ten je plně dostupný studentům a zaměstnancům mnoha univerzit včetně Masarykovy univerzity a část jeho funkcí je přístupná i uživatelům mimo akademické prostředí. Kromě korpusového manažeru *KonText*, který je základním nástrojem pro práci s jazykovými daty v korpusech ÚČNK, je součástí projektu Českého národního korpusu mnoho dalších nástrojů, z nichž lze jmenovat např. databázi překladových ekvivalentů *Treq*, nástroj *SyD* pro zkoumání proměn variant, nástroj *K words* pro analýzu textů na základě srovnání s obecným územím, mapovou aplikaci *Mapka* znázorňující geografické rozložení nářečí nebo aplikaci *Slovo v kostce* sloužící k vytvoření přehledu využití zadaného slova.

ÚČNK se dříve věnoval výhradně tradičním, ručně budovaným korpusům, které byly referenční (jejich obsah se od zveřejnění neměnil). Od roku 2020 však ústav zveřejňuje také webový monitorovací korpus ONLINE, jehož obsah, zařazovaný do korpusu bez zásahu lidí, se neustále mění. Zdroje dat tohoto korpusu má na starost externí firma, která také zajišťuje legálnost využití textů stažených z internetu.

Exkurze do ÚČNK FF UK v Praze předčila má očekávání, a to jak nečekaně uvolněnou a neformální atmosférou pracoviště, tak množstvím zajímavých projektů, jimiž se ústav zabývá a o

nichž jsem před exkurzí nevěděl. Exkurzi bych doporučil i dalším ročníkům, zejména pak studentům zajímavícím se o korpusovou lingvistiku a novinky v tomto odvětví.

The screenshot shows the KonText application interface. At the top, there's a search bar and navigation options like 'Dotaz', 'Korpusy', 'Uložit', 'Konkordance', 'Filtr', 'Frekvence', 'Kolokace', 'Zobrazení', and 'Nápověda'. Below that, it displays search statistics: 'Korpus: syn2020 | Dotaz: byt (4 335 340 výskytů) ~ Podrobnosti'. The main area shows a list of search results for 'Americké psycho'. One result is expanded, showing a text snippet with highlighted words and their frequency in the corpus. The snippet discusses the character of the 'Americké psycho' and their behavior in a taxi.

Uživatelské rozhraní aplikace KonText, zdroj: <https://www.korpus.cz/kontext/>



Vstup z nádvoří Kounického paláce



Vchodové dveře do pracoviště ÚČNK s nápusem: „Zde se buduje Český národní korpus“



*PhDr. Mgr. Marie Kopřivová, Ph.D., hovoří o projektech ÚČNK; foto: Katarína Petreková*