

# POČÁTKY JAZYKOVÝCH KORPUSŮ U NÁS

**Datum konání:** 17. března 2017

**Místo konání:** Filozofická fakulta Masarykovy univerzity, učebna M11

**Název přednášky:** Počátky jazykových korpusů u nás

**Přednášející:** Prof. PhDr. František Čermák, DrSc.

**Počet účastníků:** 35

**Zpracovali:** Jakub Lédl, Marie Kaňovská

## Prof. PhDr. František Čermák, DrSc.

Profesor František Čermák absolvoval obory česká, anglická a holandská filologie na Filozofické fakultě Univerzity Karlovy. Kromě své akademicko-pedagogické činnosti (v roce 1994 jmenován profesorem českého jazyka) spoluzakládal a následně řídil Ústav Českého národního korpusu, který se pod jeho vedením rozrostl v provozovatele jednoho z největších a nejznámějších projektů v korpusové lingvistice na světě. Profesor Čermák je taktéž členem mnoha odborných společností, a to jak domácích, tak zahraničních. Odborně se věnuje zejména lexikologii a lexikografii, frazeologii a idiomatice a obecné a korpusové jazykovědě. Mezi jeho publikace spadá, kromě mnoha monografií, překladů apod., rovněž čtyřsvazkový *Slovník české frazeologie a idiomatiky* (FF UK, 2017).

## Obsah

V pátek 17. března 2017 jsme se sešli na přednášce profesora Františka Čermáka na téma „Počátky jazykových korpusů u nás“. Na úvod byly představeny motivace, které ke snaze, jejímž dnešním výsledkem je projekt Českého národního korpusu, vůbec vedly. Hlavní motivací bylo úsilí o rozvoj české lingvistiky jako takové, zejména pak potřeba spolehlivých podkladů pro nový český slovník, jelikož předchozí činnost v tomto směru byla mimo jiné politicky ovlivněná. Prvním krokem bylo založení neoficiální skupiny nazvané *Počítačový fond češtiny*, jejíž členové se snažili budování korpusu finančně zajistit (Schmiedtová, 1999). Zmíněna byla též rozporuplná role tehdejšího ředitele Ústavu pro jazyk český AV ČR Františka Daneše, který daný projekt odmítal povolit.

Zásadní zlom nastal v roce 1994, kdy tehdejší děkan FF UK docent František Vrhel vydal dekret zakládající Ústav Českého národního korpusu, jehož ředitelem se stal právě profesor Čermák. Přes tento pozitivní vývoj bylo zajišťování zdrojů pro projekt velmi obtížné. Až posléze začaly projekt podporovat i externí skupiny, mezi prvními bylo nakladatelství Lidových novin.

Obr. 1. Druhá část přednášky – stav před vznikem ČNK



Po této úvodní části se přednášející věnoval otázce stavu korpusů v zahraničí a u nás před vznikem Českého národního korpusu (dále ČNK). Jmenoval zahraniční předchůdce elektronických korpusů, například práci dánského jazykovědce Otta Jespersena nebo díla jako *Oxfordský slovník* a zejména *Slovník nizozemského jazyka*. Též zdůraznil význam tzv. Brownova korpusu, který byl podkladem pro prestižní slovník *American Heritage Dictionary* a na kterém spolupracoval i český lingvista Henry Kučera. Z domácích projektů byla zmíněna například kartotéka excerpčních lístků sestavená počátkem 20. století Kanceláří jazyka českého a dále práce Marie Těšitelové *Frekvence slov a slovních druhů*.

Další část přednášky byla věnována praktickým otázkám spojeným se sestavováním korpusů. Po organizační stránce bylo nesnadným úkolem zkoordinovat odborníky

z různých disciplín – lingvisty, informatiky, matematiky a další. Z hlediska technického řešení lemmatizace zdrojových textů, zásadní součástí výstavby jazykového korpusu, představovala náročnou výzvu např. proměnnost zdrojového materiálu, která přinášela nové jazykové jevy, nebo nutnost tzv. funkční lemmatizace – analýzy víceslovných lemat, kterými jsou například v češtině četná slovesa pouze zvrtná. Zajímavou otázkou je rovněž možnost praktického zachycení jazykové reality, která je proměnná, ale její zásadní proporce zůstávají neměnné.

V závěrečné části byly nastíněny možné cíle budoucího rozvoje projektu ČNK. Má-li korpus věrně odrážet existující jazykovou realitu, mělo by jít z velké části o korpus mluvený, jehož sestavení a technická realizace je nicméně mnohem obtížnější – z tohoto důvodu je většina současných českých korpusů textových. Současně by měl takový korpus zachycovat český jazyk z různých hledisek, tedy i v podobě hovorové, z hlediska diachronního, dialektologického atd. Zajímavá je rovněž snaha ČNK o zapojení korpusů do veřejného života, zejména do programů vzdělávacích institucí. Přednáška byla ukončena zdůrazněním důležitosti projektu ČNK a korpusové jazykovědy obecně jakožto prostředku k nejspolehlivějšímu popisu češtiny, který, pokud bude zachován, poslouží i jako významný odkaz budoucím generacím.

Po skončení přednášky následovala část věnovaná dotazům, během níž profesor Čermák rozvedl například spolupráci se zahraničními institucemi a zpětnou vazbu, kterou tyto instituce projektu poskytly, nebo historický vývoj a současnou situaci korpusů mluveného jazyka.

## **Závěr**

Profesor Čermák poskytl přítomnému obecenstvu velmi zajímavý a cenný vhled do historie Českého národního korpusu a okolností, které vznik tohoto projektu doprovázely. Svou přednášku obohatil i o několik osobních zkušeností a vzpomínek, kterými posluchačům přiblížil pohled na toto dílo očima jednoho z jeho hlavních realizátorů. Část věnovaná budoucnosti ČNK pak představila neméně zajímavý náhled na široké možnosti, které další rozvoj české korpusové lingvistiky skýtá.

*Obr. 2. Závěrečné poděkování přednášejícímu*



## Použité zdroje

- Filozofická fakulta UK, 2017. Prof. PhDr. František Čermák, DrSc. In: *Filozofická fakulta Univerzity Karlovy* [online]. FF UK, © 2017. Cit. 29. 3. 2017. Dostupné z: <http://www.ff.cuni.cz/fakulta/cestne-afiliace/emeritni-clenove-akademicke-obce-ff-uk/prof-phdr-frantisek-cermak-drsc/>
- SCHMIEDTOVÁ, Věra, 1999. Český národní korpus. *Ikaros*. Elektronický časopis o informační společnosti [online]. *Ikaros*, 3(5) [cit. 29. 3. 2017]. ISSN 1212-5075. Dostupné z: <https://ikaros.cz/cesky-narodni-korpus>