

Rádi bychom pozvali všechny zájemce z akademické obce, včetně doktorandů a studentů pregraduálního studia, na dvoudílný workshop

InterCorp a Universal Dependencies: nové možnosti výzkumu

Workshop se uskuteční v rámci

[Teoreticko-metodologického semináře Ústavu českého jazyka a teorie komunikace.](#)

Workshop povedou:

Olga Nádvorníková (Ústav románských studií)

Alexandr Rosen (Ústav Českého národního korpusu)

Jako host vystoupí také:

Martin Stluka (Ústav Českého národního korpusu)

Kdy:

20.3. 15:50–17:25 (část 1)

27.3. 15:50–17:25 (část 2)

Kde:

Prezenčně: FF UK, nám. J. Palacha, č. 104

On line: odkaz po registraci v [dotazníku](#)

Další údaje:

[Stránka workshopu](#) s odkazy a literaturou

Část 1:

- Český národní korpus, paralelní korpus InterCorp (stručný úvod)
- Universal Dependencies (představení schématu pro jednotnou lingvistickou anotaci, zaměřené na syntax)
- Jednotná lingvistická anotace v nové verzi InterCorpu

Část 2:

- Jednotná lingvistická anotace v projektu HiČKoK (Historie češtiny v korpusovém kontinuu)
- Míry syntaktické komplexity a lexikální diverzity v nové verzi InterCorpu
- Možnosti využití v kontrastivním a translatologickém výzkumu, výzkumu variability textových typů i v jazykové výuce

Abychom mohli co nejlépe navázat na Vaše dosavadní zkušenosti s korpusovými zdroji a strefit se do Vašich očekávání, prosíme o vyplnění stručného [dotazníku](#).

Děkujeme a těšíme se na Vás,

Olga Nádvorníková

Alexandr Rosen

Abstract

Cílem workshopu je představit paralelní korpus [InterCorp](#) a nové možnosti lingvistického výzkumu, které nabídne jeho právě připravovaná verze 16ud. Workshop bude rozdělen do dvou částí. V první části stručně představíme celý korpus, který v současné době obsahuje 62 jazyků (včetně češtiny), a vysvětlíme základní principy anotace a vyhledávání pomocí jednotného schématu [Universal Dependencies](#) (UD). Ve druhé části workshopu se budeme věnovat možnostem využití takové anotace v pokročilejších dotazech a při generování statistik. Zaměříme se také na míry syntaktické komplexity a lexikální diverzity, nově uváděné v nejnovější verzi u jednotlivých vět a textů. Ukážeme si, jak je lze využít např. v kontrastivním nebo translátologickém výzkumu, ale i ve výzkumu vnitrojazykové variability textových typů (registrů) nebo ve výuce L1/L2. Součástí workshopu bude také krátká prezentace projektu [HiČKoK](#) (Historie češtiny v korpusovém kontinuu), který počítá s jednotnou lingvistickou anotací češtiny napříč všemi obdobími v rámci schématu UD. Účastníci mohou prostřednictvím formuláře předem zaslat dotazy a témata, která je zajímají, a v průběhu workshopu si vyhledávání v korpusu budou moci zkusit rovnou na svých zařízeních.