

# *PŘEDSTAVENÍ KORPUSU SYN2020 A DALŠÍ NOVINKY V ÚČNK*

**Datum konání:** 23. 4. 2021

**Místo konání:** MS Teams

**Název přednášky:** Novinky v ČNK: Korpusy a aplikace

**Přednášející:** Mgr. Adrian Zasina, Ph.D.

**Počet účastníků:** 27

**Zpracovala:** Miroslava Machanová

Mgr. Adrian Zasina, Ph.D., vystudoval obor Slovanské filologie se zaměřením na český jazyk. Doktorský titul získal v oboru Matematická lingvistika na Filozofické fakultě Univerzity Karlovy. Pracuje v Ústavu Českého národního korpusu FF UK, kde se také mimo jiné podílí na projektu paralelního korpusu InterCorp. Mezi jeho hlavní zájmy patří korpusový přístup ve výuce, současná gramatika češtiny a polštiny, korpusová lingvistika a diskurzní analýza genderu.



Hned v úvodu jsme byli svižně seznámeni s webovými stránkami ČNK, kde nám bylo doporučeno navštívit korpusovou Wiki, na které můžeme najít hodně užitečných rad a tipů.

Po krátkém úvodu nám byla představena webová aplikace KonText. První novinkou bylo zjednodušení zadávání jednoduchého dotazu. Nyní můžeme při hledání jednoduchých dotazů používat víceslovné vyhledávání, volitelné regulární výrazy a volitelné atributy. Funkce pokročilého vyhledávání však zůstaly stejné jako v předchozí verzi.

Další část přednášky pan doktor věnoval popisu inovací v korpusu SYN 2020. Ve zkratce sdělil základní informace o korpusu, a poté nás seznámil s první novinkou týkající se anotace. Změny zahrnují novou tokenizaci a morfologické značkování, které nově vychází z projektu NovaMorf. Dále jsme se zabývali dvouúrovňovou lemmatizací, která rozšiřuje atribut lemma o další atribut sublemma. Poté nám bylo popsáno nové morfologické značkování, které se týkalo eliminace 16. pozice a také přidání dalších tří tagů v kategorii slovních druhů. Poslední aktualitou bylo nové značkování sloves a vícenásobná lemmatizace (agregáty).

Obsahem dalšího bodu prezentace byl paralelní korpus InterCorp v 13. Přednášející nám ukázal, jaký je celkový obsah korpusu, a dostalo se nám i praktických ukázek ilustrujících práci se zarovnanými korpusy a aplikací Treq.

Poslední část přednášky pan doktor věnoval dalším dvěma zajímavým aplikacím. První aplikací bylo Slovo v kostce, které slouží k vytvoření rychlého a přehledného profilu hledaného slova vycházející z korpusových dat. Druhá aplikace SMARTool je zatím ve zkušební verzi, ale po jejím plném spuštění budeme moci využít nabídky úrovní češtiny a ty zkombinovat s různými tematickými okruhy, morfologickými analýzami, slovníkem a výpisem zkratk. Tato aplikace je vytvořena hlavně pro zahraniční studenty.

Závěrem celou přednášku hodnotím jako velmi úspěšnou. I přesto, že se konala v online prostoru, pan doktor měl velmi dobře připravenou jak prezentaci, tak praktické ukázky. Po skončení proběhla i krátká diskuze, kde všem zúčastněným byl vyhrazen prostor na jejich dotazy.