

# Počítačová a korpusová lingvistika v Rusku

**Datum konání:** 7. 5. 2021

**Místo konání:** Online prostřednictvím MS Teams

**Název přednášky:** Počítačová a korpusová lingvistika v Rusku (v Petrohradě)

**Přednášející:** Doc. PhDr. Maria Khokhlova, Ph.D.

**Počet účastníků:** 22

**Zpracovali:** Koutová, Martina

Docentka Khokhlova působí na katedře Matematické lingvistiky Petrohradské státní univerzity. Zaobírá se především počítačovou a korpusovou lingvistikou a statistickými metodami. Aktuálně zkoumá kolokace, např. pro využití při výuce ruštiny jako cizího jazyka. Kromě výzkumu i vyučuje předměty související s korpusovou lingvistikou, statistickými metodami a lingvistickými databázemi. Je držitelkou ceny vedení města Petrohradu za pedagogickou a výzkumnou činnost. Absolvovala stáž na Fakultě informatiky Masarykovy univerzity.



V úvodu nás paní docentka seznámila s katedrou Matematické lingvistiky, která byla založena roku 1962 a nechala se inspirovat pracemi Pražského lingvistického kroužku. Mezi nejdůležitější projekty na katedře Matematické lingvistiky patří RussNet (WordNet pro ruštinu), SKAT (korpus hagiografických textů), Ingrie (databáze toponym) a databáze ruských kolokací.

Následoval přehled ruských korpusů a bližší představení Ruského národního korpusu, který byl inspirován korpusy Českého národního korpusu. Jedná se o největší korpus ruštiny a skládá se z několika subkorpusů. Z těchto subkorpusů nám docentka Khokhlova

představila základní korpus, který rozhraním připomíná korpusy ČNK, syntaktický korpus a multimediální korpus. Třetí z těchto subkorpusů byl nejzajímavější. Obsahuje části filmů, zpráv, pořadů, veřejných projevů, přednášek apod. a má označovaná gesta a řečové akce. Z gest např. orientace dlaní, emoce, směr pohybu, jméno gesta, autentičnost gesta a další. Z řečových akcí pak mluvený jazyk, sociální situace, způsob mluvení, druhy řečových akcí (otázky a odpovědi, souhlas, pejorativní promluvy apod.). Podle těchto značek pak můžeme v korpusu vyhledávat a přímo si i přehrát ukázkou, kde se daný jev vyskytuje. Dokonce existuje i Multimediální paralelní korpus, kde si například můžeme porovnat dvě různá zpracování téže divadelní hry.

Pro výuku ruštiny jako cizího jazyka může sloužit Ruský žákovský korpus, který obsahuje psané a mluvené texty cizinců a imigrantů, kteří se učí rusky. Podobným korpusem je pak Korpus ruských studentských textů.

Po korpusech se docentka Khokhlova zaměřila na slovníky, při jejichž tvorbě se použilo materiálu z Ruského národního korpusu. Mezi tyto slovníky patří: Slovník ruské idiomatiky, Slovník slovesné kolokability abstraktních jmen, Nový frekvenční slovník ruštiny.

Další zajímavostí byla existence jazykové poradny pro ruštinu. Na rozdíl od Jazykové poradny Ústavu pro jazyk český AV ČR v té ruské s nikým nehovoříte. Ruská jazyková poradna obsahuje slovníky, články a pravidla pravopisu, ve kterých můžete samostatně vyhledávat.

Přednáška byla zakončena nástroji pro zkoumání sémantických vztahů mezi slovy, analýzu sentimentu a následným prostorem pro dotazy a diskuzi.

Docentka Khokhlova velmi pěkně a přehledně přiblížila nástroje vyvíjené na Petrohradské státní univerzitě. Díky této přednášce mohli studenti získat přehled o tom, jak se tvoří korpusy a nástroje zpracovávající jazyk v zahraničí. Bylo zajímavé je porovnat s českými nástroji a zjistit, na jaké nástroje působila inspirace Pražským lingvistickým kroužkem nebo Českým národním korpusem a u jakých nástrojů zašli v Rusku ještě dál než my. Díky přednášce se také studenti mohli zamyslet se nad tím, jaké nástroje pro češtinu chybí nebo jaké by se ještě mohly vylepšit. Toto se týká například multimediálního korpusu. Nic podobného v České republice nemáme, proto není divu, že studenty zaujal nejvíce.

Další témata, která studenty zaujala, můžeme vidět na následujícím obrázku:

# Co Vás dnes zaujalo?

fascinující rusko  
gesta  
úžasná přednáška  
multimediální korpus  
multimediální korpusy  
skvělá přednáška a výstup  
analýza sentimentu  
nejdelší chodba  
profesionalita