

JAK MĚŘIT JAZYK

Datum konání: 9. 10. 2020

Místo konání: on-line skrz aplikaci MS Teams

Název přednášky: Úvod do kvantitativní lingvistiky

Přednášející: doc. Mgr. Radek Čech, Ph.D.

Počet účastníků: 30

Zpracovali: Kubeček Filip, Sedláčková Sára

Podzimní cyklus odborných přednášek, jenž v důsledku pandemie nového koronaviru probíhá virtuálně, zahájil pan docent Radek Čech svým pojednáním o základech kvantitativní lingvistiky, progresivního odvětví, v němž lingvistika představuje experimentální vědu.

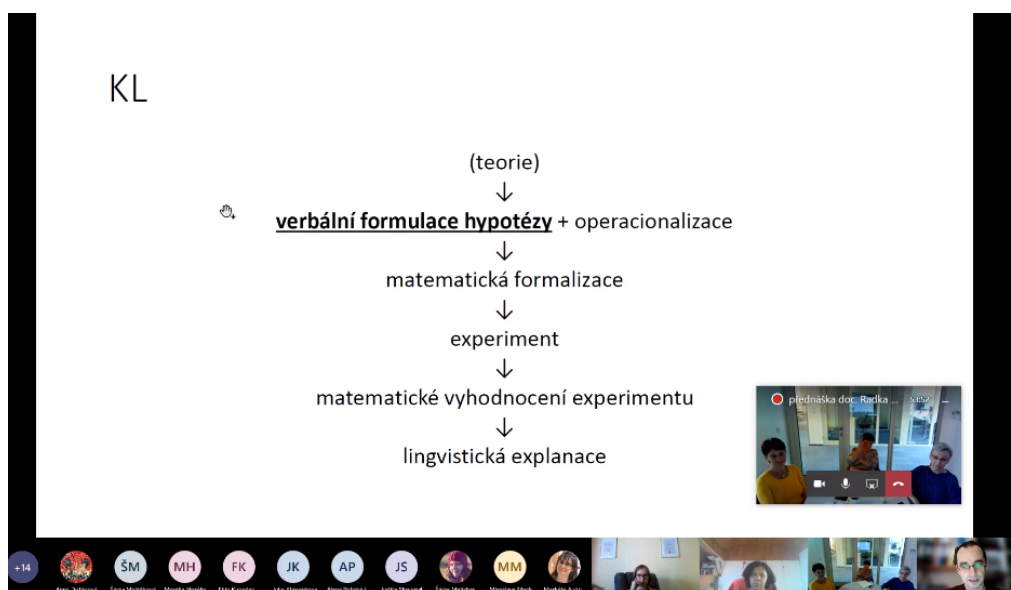
Doc. Mgr. Radek Čech, Ph.D., působí na Katedře českého jazyka Filozofické fakulty Ostravské univerzity a zabývá se syntaxí, textologií, obecnou lingvistikou, analýzou Menzerathova-Altmanova zákona a rovněž vyvíjením lingvistických softwarů. Spolupracuje také s Masarykovou univerzitou nebo Ústavem Českého národního korpusu. Souhrn veškeré jeho činnosti a výsledků je k nahlédnutí na jeho webových stránkách cechradek.cz.



Pan docent začal vymezením pojmu kvantitativní lingvistiky. Jejím cílem je výzkum kvantitativních vlastností přirozeného jazyka, čehož dosahuje pomocí metod matematické statistiky, teorie pravděpodobnosti, teorie informace, teorie fraktálů, entropie nebo teorie chaosu. Po krátkém pojednání o historii oboru jsme byli obeznámeni s jeho nosnými myšlenkami, a sice kladením důrazu na explanaci čili vysvětlení, jak fungují jazykové jevy a proč mají ty které vlastnosti; empirickou čili experimentální metodu, která stojí na práci s daty; a vysvětlení jazykového chování, v němž jsou hledány a zkoumány jeho pravidelnosti. Právě jazykové chování je často nestabilní a náchylné na změny, a proto je na jazyk pohlíženo jako na stochastický systém, v němž se jevy vyskytují s určitou pravděpodobností. I pojetí pravidel

v kvantitativní lingvistice je stochastické a případné odchylky a jejich okolnosti jsou rovněž probádávány.

Abychom si dokázali prakticky představit, jak pracuje lingvista, který se pokouší o nějakou explanaci, popsal nám pan docent proces, který je pro to nutné absolvovat. Zdůraznil přitom, že jde ve své podstatě o obecný postup jakéhokoliv experimentálního výzkumu. Je možné vycházet z teoretické úvahy (pro kvantitativní lingvistiku je známá třeba teorie Zipfova), z níž lze odvozovat empiricky testovatelné hypotézy (např. „čím je slovo frekventovanější, tím je větší jeho polysémie“). Ty však lze formulovat i na základě vlastní zkušenosti, intuice nebo nápadu. Bez hypotézy není možné ve výzkumu pokračovat, neboť je nezbytné vědět, co je předmětem bádání.

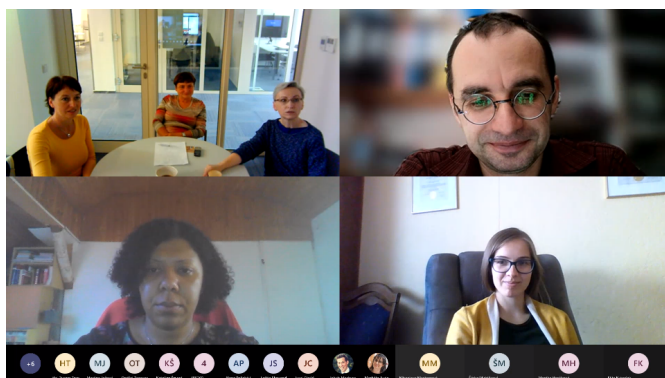


V rámci následné operacionalizace se definují proměnné, se kterými se bude pracovat (např. „Co je slovo? Jak se bude polysémie kvantifikovat?“), přičemž je nutné mít na paměti, že různé operacionalizace téže hypotézy mohou výrazně ovlivňovat podobu výsledku. Dalším krokem je matematická formalizace, jež spočívá ve vyjádření sledovaných vlastností čísly. V kvantitativní lingvistice se často projektuje do distribučních modelů, z nichž lze dobře vyčíst případné pravidelnosti. Stěžejní částí je samotný experiment, který v lingvistickém prostředí často spočívá v práci s korpusem nebo analýze dat pomocí softwarů modelujících funkce pro testované jevy. Hlavní výhodou experimentálního přístupu je jeho replikovatelnost – každý, kdo experiment zopakuje, dojde k týmž výsledkům. Nevýhodou je naopak nezbytný, pro experimentální vědy však typický redukcionismus a také specifické technické nároky, s nimiž se v tradiční lingvistice běžně nesetkáváme. Experiment může vést i k vyvrácení hypotézy, což však nemusí znamenat celou její neplatnost, ale jen odhalení nutnosti stanovení dodatečných podmínek. Na základě matematického vyhodnocení experimentu dojde následně buď k odhalení mechanismu, nebo jen vlivu náhody.

K tomu se využívají statistické testy významnosti, které určují, za jakého výsledku už o náhodě mluvit nelze. V rámci nich se pracuje se dvěma pojmy; nulovou hypotézou pro rozdíl zapříčiněný náhodou a alternativní hypotézou pro závislost danou mechanismem. Výsledkem celého procesu je lingvistická explanace, kterou pan docent ukázal na vztazích mezi délkou slov, jejich podobností a velikostí inventáře fonémů, které vedou k co nejmenšímu úsilí mluvčího i posluchače, ale zároveň k jejich vzájemnému porozumění.

Ke konci přednášky jsme se dozvěděli o Mezinárodní asociaci kvantitativní lingvistiky, jejich konferencích a časopisech, knižních řadách a dalších publikacích věnujících se tomuto odvětví. Na úplný závěr před prostorem k diskusi pan docent předestřel budoucnost kvantitativní lingvistiky, jejíž metodologie se

díky svému rozmachu v posledních letech stává běžnou praxí a lze předpokládat sílení této tendence. Díky kvantitativní lingvistice může nadále docházet k formování obecných zákonitostí a modelování jazykového vývoje. Byly by ale potřeba důkladnější analýzy teorií, z nichž se vychází.



Přednášku pana docenta Čecha hodnotíme nadmíru přínosně, a to nejen díky možnosti vhledu do oboru kvantitativní lingvistiky jako takového, ale především díky důkladnému objasnění experimentálního badatelského procesu, jehož znalost se hodí pro široké spektrum výzkumů, zvláště v oblasti počítačové lingvistiky. Díky názorným příkladům jsme si dokázali dobře představit, jak se v dané části postupuje, jakých zásad je vhodné se držet a čemu se vyvarovat – například neshromažďovat data bez vytyčené hypotézy. Počítáme, že nabyté informace využijeme při dalším studiu nebo práci.