

PLIN041 Vývoj počítačové lingvistiky

Raná korpusová lingvistika

Mgr. Dana Hlaváčková, Ph.D.

Raná korpusová lingvistika (Early corpus linguistics)

konec 19. st – 50. léta 20. st.

- strukturalistická tradice v Evropě, americký deskriptivismus (studium indiánských jazyků)
 - „korpusová“ metodologie
 - metody založené na **zkoumání souborů textů a na empirii**
- shromažďování jazykového materiálu
 - rozsah je důležitým parametrem
- v této době se ještě nemluví o korpusech a korpusové lingvistice – *archiv, kartotéka, deníky, seznamy, slovníky*
- žánrová vyváženost souboru textů
- zkoumání významů slov a homonymie
- problematika slovní jednotky a lemmatizace
- morfologické, syntaktické i sémantické analýzy jazyka na základě textového materiálu

Raná korpusová lingvistika

- **počátky moderní lexikografie** – excerpční lístky (ručně, na stroji), výpisky z beletrie, novin, zapojení slova v kontextu (konkordance)
 - **frekvenční studie** – **W. Käding**, 1897–1898 (11 mil. slov), na dlouhou dobu nejrozsáhlejší jazykový materiál v podobě frekvenčních seznamů a frekvenčního slovníku
 - **výuka jazyka pro cizince** – frekvenční seznamy slov, frekvenční slovníky a navazující slovníky a učebnice k výuce jazyka pro cizince, např. **E. Thorndike** – *The Teacher's Word Book*, 1921

Raná korpusová lingvistika

- **zápisy dětské mluvy** – rodičovské deníky, studium akvizice jazyka (1876–1926), od 1927 analýzy jazyka, později malý vzorek dětí a dlouhodobé sledování
- **William Thierry Preyer** (1841–1897)
- narodil se v Anglii, studoval a žil v Německu
- vystudoval chemii a fyziologii v Heidelbergu a medicínu v Bonnu
- působil v Jeně jako ředitel fyziologického ústavu
- zakladatel dětské psychologie
 - založena na empirickém pozorování a experimentech
 - k výzkumu využívá rodičovské deníky
 - významné dílo *Die Seele des Kindes* – vývojová psychologie

Raná korpusová lingvistika

- **komparativní lingvistika** – srovnávání významů slov z různých jazyků, studium jazyka Bible a dalších kanonických textů (užívání konkordancí)
- **zapisování indiánských jazyků**
- **Franz Boas** (1858–1942), pův. Němec, zakladatel moderní americké **antropologie**, studie indiánských kmenů

Raná korpusová lingvistika

- Franz Boas

- vystudoval fyziku a geografii
- při výpravě do severní Kanady ho okouzlil jazyk a kultura domorodých kmenů
- emigroval do USA – profesorem antropologie na Columbia University
- byl proti „vědeckému rasismu“ – např. stavba lebky se řídí rasou
- stavba kostry je ovlivněna okolním prostředím a výživou
- chování lidí není výsledkem biologické predispozice, ale ovlivněno sociálním prostředím a výchovou

Raná korpusová lingvistika

Kritika

- kolem 1950 – **Noam Chomsky** – generativní lingvistika, odpor ke korpusovému přístupu k jazyku, korpusy nejsou v lingvistice potřebné, poskytují *pokřivená data*
- předpočítačové období – ruční hledání v rozsáhlých datech je příliš pracné
- **X** rozvoj počítačové techniky