

Název práce: **Evaluace chybové anotace v žakovském korpusu češtiny**
Autor: Barbora Štindlová
Ústav: Ústav českého jazyka a teorie komunikace, Filozofická fakulta, Univerzita Karlova
Vedoucí disertační práce: prof. PhDr. Karel Šebesta, CSc.

Abstrakt:

Předkládaná práce se obecně týká tématu češtiny jako cizího jazyka a částečně zasahuje do oblasti korpusové lingvistiky, neboť se věnuje problematice žakovských korpusů, především pak otázkám jejich chybového značkování a možnostem evaluace anotačních schémat. Žakovské korpusy se staly významným zdrojem pro poznání žakovského mezijazyka a významným stimulem pro různé oblasti studia a výuky cizího, resp. druhého jazyka. Jsou využívány zejména pro kontrastivní srovnávání jazyka roditelů a nerodilých mluvčích, resp. srovnávání žakovských mezijazyků a pro tzv. počítačem podporovanou chybovou analýzu žakovského jazyka. Pro tento typ analýzy má zcela zásadní důležitost tzv. chybové značkování. Chybové značkování je u každého korpusu, pokud jej používá, založeno na chybové typologii, jejíž vymezení je v mnoha teoretických aspektech problematické. Z toho důvodu je důležitým krokem při výstavbě žakovského korpusu zhodnocení spolehlivosti a validity navrženého anotačního schématu.

Disertační práce se zaměřuje především na technické aspekty a specifické problémy při elektronizaci rukopisů, na možnosti chybové anotace projevů nerodilých mluvčích a problematiku její evaluace. Zároveň však věnuje značný prostor i metodologii, koncepci a účelu budování žakovských korpusů, protože téma korpusu nerodilých mluvčích a jeho využití je v českém prostředí relativně nové a je vhodné jej podrobněji představit. V první části (A) jsou stručně shrnuty základní přístupy k otázkám nabývání cizího, resp. druhého jazyka a podrobněji představeny proměny teorie chyby v jazyce nerodilých mluvčích. V části (B) předkládám shrnutí aktuálního stavu problematiky a uvádím podrobný přehled existujících korpusů jazyka nerodilých mluvčích založený na dotazníkovém šetření a podrobné analýze dostupných žakovských korpusů. Třetí část práce (C) představuje budovaný žakovský korpus češtiny nerodilých mluvčích (CzeSL) a soustředí se především na problematiku přepisu dat. Čtvrtá část disertace (D) se zabývá evaluací konceptu chybové anotace navržené pro žakovský korpus CzeSL. Pro zhodnocení spolehlivosti anotačního schématu byl zvolen výpočet tzv. koeficientu mezianotátorské shody kappa. Výsledky měření mezianotátorské shody, analýza anotačních problémů, návrhy jejich řešení a zhodnocení anotačního schématu včetně chybové taxonomie jsou jedním z hlavních výsledků této práce.

Klíčová slova: žakovský korpus, chybová anotace, elektronizace rukopisů, mezianotátorská shoda, mezijazyk