

Korpusi v jezikoslovju: Zgodovinski vpogled

Evropa in Severna Amerika

Ženevski strukturalizem: Ferdinand de Saussure (1857-1913), Cours de linguistique générale, 1916

Amerika: sinhroni pristop – opazovanje, zapisovanje in analiza govornega jezika (antropologija): Franz Boas (1858-1942), Handbook of American Indian Languages, 1911; Edward Sapir (1884-1939), Language, 1921

V Evropi mo

an vpliv ženevskega strukturalizma: Praški lingvisti

ni krožek, 1926 (magi

nost navadnega jezika); Vilém Mathesius, Roman Jakobson, Nikolai Sergeevi

Trubetzkoy

Anglija – ameriški vpliv: John Rupert Firth

Severna Amerika: Tradicija ameriških antropologov (magi

nost govornega jezika): Leonard Bloomfield, Language, 1933; Avram Noam Chomsky

(1928-) (generativna oz. tvorbo-pretvorbna slovnica-jezikovna oz. slovni

na zmožnost omogo

a tvorbo in razumevanje neskon

nega števila stavkov oz. izjav??); Syntactic Structures, 1957 - delitev jezikoslovja izrazito na

dva pola: chomskijansko in njegovo opozicija (vrsta šol, ki temeljijo na realnosti

komunikacije)

Gradivno usmerjeni jezikoslovni opisi

- raziskave mrtvih jezikov – korpus razpoložljivih besedil
- analiza jezikov brez pisne tradicije – govorni korpusi (Franz Boas)
 - na
 - rtno zbiranje jezikovnih podatkov
 - od kon
 - nega nabora jezikovnih podatkov (mrtvi jeziki) do zbiranja in izbiranja besedil kot osnove jezikovne analize
 - od opazovanja pisnega k opazovanju govornega jezika

N. Chomsky in “njegova opozicija”

N. Chomsky zavra

a obsežne na

rtno zbrane jezikovne podatke : jezikoslovje, utemeljeno na jezikovnih podatkih kot

“organizirana opozicija”; zavra

a veliko koli

ino jezikovnih podatkov za temelj jezikoslovnih raziskav; prve poskuse gradnje korpusa so ozna

ili kot popolna izguba

asa, saj naj bi govorec dolo

enega jezika v 10 min produciral za jezikoslovce ve

ilustrativnega gradiva, kot ga je na voljo v ve

milj. besedilnih korpusih.

Dileme

- Vprašanje absolutne zadostnosti korpusa ali absolutne zadostnosti intuicije, ki jo predvideva Chomsky, je s

- asovne distance gledano umetno ustvarjena dilema. Delo s korpusom res temelji na veliki koli
- ini zbranih besedil, ki so podlaga za jezikovno analizo, a se ne odreka intuiciji.
- intuicija rojenega govorca: idealni govorec kakšna koli
- ina izpri
- anih besedil in stavkov je potrebna, da lahko ustrezno ponazarja jezikovno zmožnost chomskyjevskega idealnega rojenega govorca?... Zanašanje na intuicijo je v generativni slovnici privedlo tudi do definiranja idealnega govorca. Predvsem sociolingvisti
- ne raziskave in utemeljitev funkcijskega pristopa v jezikoslovju so pokazale nevzdržnost takega izhodiš
- a. Takoj ko je bil v izhodiš
- e postavljen vidik nemonolitnosti jezikovne pojavnosti, je idealni govorec nezadosten, saj njegovo jezikovno mnenje ne more zajeti celotne jezikovne pojavnosti in njene variabilnosti.
- normativisti
- ni pristop
 - jezikovna raba : odklon od norme : napaka □ analize korpusov ang. So pokazale na vrsto jezikovnih rab, ki bi jih z vidika jezikovne intuicije ozna
 - ili kot nemogo
 - e, z normativnega vidika pa ozna
 - ili kot napake; tovrstne analize so opozorile na v
 - asih prevelike posplošitve v skladu z željo po sistematizaciji.

“KJER SE LOMIJO KOPJA”

- jezikovna realnost in uporabno jezikoslovje: razvoj uporabnega jezikoslovja je zahteval podatke o jezikovni rabi, in sicer tako naravnih govorcev nekega jezika kot prvega kakor tudi drugega ali tujega jezika
- jezikovna realnost in procesiranje naravnih jezikov □ navezava procesiranja naravnih jezikov kot potencialnega prihodnjega sistema analize in sinteze besedil naravnih jezikov na realne korpusne podatke je nujna, zato da bi v kon
- ni fazi tovrstne aplikacije res lahko u
- inkovitno delovale

Korpusi prve generacije

Dva osnovna principa

- splošni korpus
- specializirani korpus (korpus za posebne namene)

□ Prvi korpus še ne pravi korpus: *SEU (Survey of English Usage)*, 1955, 1959 □ še klasi ni nera

unalniški, britanska angleš

ina; polovica besedil je transkripcija govora; kasneje prenesen v elektronsko obliko (V TEM ČASU GRADIVNA ZBIRKA ZA SSKJ-klasi

na listkovna kartoteka)

□ *Korpus Brown*, 1961-1964; ameriška angleš

ina, prvi elektronski korpus, standard je milijon besed (V TEM ČASU GRADIVNA ZBIRKA ZA SSKJ-klasi

na listkovna kartoteka)

□ Na lingvisti

nem inštitutu FF Univerze v Zagrebu - *prevod polovice korpusa Brown; desetletje kasneje*
Mogušev korpus

□ *Korpus LOB* (Lancaster, Oslo, Bergen) 1970-1978: angleška varianta korpusa Brown