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 govorjenega jezika (antropologija): Franz Boas (1858-1942), Handbook of American Indian Languages, 1911;

Edward Sapir (1884-1939), Language, 1921

V Evropi močan vpliv ženevskega strukturalizma: Praški linvistični krožek, 1926 (magičnost navadnega jezika); Vilém Mathesius, Roman Jakobson, Nikolai Sergeevič Trubetzkoy

Anglija – ameriški vpliv: John Rupret Firth

Severna Amerika: Tradicija ameriških antropologov (magičnost govorjenega jezika): Leonard Bloomfield, Language, 1933; Avram Noam Chomsky (1928- ) (generativna oz. tvorbno-pretvorbna slovnica-jezikovna oz. slovnična zmožnost omogoča tvorbo in razumevanje neskončnega 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)
  + načrtno zbiranje jezikovnih podatkov
    - od končnega nabora jezikovnih podatkov (mrtvi jeziki) do zbiranja in izbiranja besedil kot osnove jezikovne analize
    - od opazovanja pisnega k opazovanju govorjenega jezika

N. Chomsky in “njegova opozicija”

N. Chomsky zavrača obsežne načrtno zbrane jezikovne podatke : jezikoslovje, utemeljeno na jezikovnih podatkih kot “organizirana opozicija”; zavrača veliko količino jezikovnih podatkov za temelj jezikoslovnih raziskav; prve poskuse gradnje korpusa so označili kot popolna izguba časa, saj naj bi govorec določenega jezika v 10 min produciral za jezikoslovce več ilustrativnega gradiva, kot ga je na voljo v večmilj. besedilnih korpusih.

Dileme

* Vprašanje absolutne zadostnosti korpusa ali absolutne zadostnosti intuicije, ki jo predvideva Chomsky, je s časovne distance gledano umetno ustvarjena dilema. Delo s korpusom res temelji na veliki količini zbranih besedil, ki so podlaga za jezikovno analizo, a se ne odreka intuiciji.
* intuicija rojenega govorca: idealni govorec🡪kakšna količina izpričanih 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 idelanega govorca. Predvsem sociolingvistične raziskave in utemeljitev funkcijskega pristopa v jezikoslovju so pokazale nevzrdžnost takega izhodišča. Takoj ko je bil v izhodišče postavljen vidik nemonolitnosti jezikovne pojavnosti, je idealni govorec nezadosten, saj njegovo jzikovno mnenje ne more zajeti celotne jezikovne pojavnosti in njene variabilnosti.
* normativistični pristop
  + jezikovna raba : odklon od norme : napaka 🡪 analize korpusov ang. So pokazale na vrsto jezikovnih rab, ki bi jih z vidika jezikovne intuicije označili kot nemogoče, z normativnega vidika pa označili kot napake; tovrstne analize so opozorile na včasih 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 končni fazi tovrstne aplikacije res lahko učinkovitno 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 klasični neračunalniški, britanska angleščina; polovica besedil je transkribcija govora; kasneje prenesen v elektronsko obliko (V TEM ČASU GRADIVNA ZBIRKA ZA SSKJ-klasična listkovna kartoteka)

🡪 Korpus Brown, 1961-1964; ameriška angleščina, prvi elektronski korpus, standard je milijon besed (V TEM ČASU GRADIVNA ZBIRKA ZA SSKJ-klasična listkovna kartoteka)

🡪 Na lingvističnem inštitutu FF Univerze v Zagrebu - prevod polovice korpusa Brown; desetletje kasneje Mogušev korpus

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