

1 UVOD

Slovenščina kot tuji jezik je pomemben del uporabne slovenistike, njeno poučevanje pa posebna strokovna disciplina, ki se v zadnjem času dinamično razvija. Obenem postaja vse močnejše zavedanje, da bi morala biti eden od temeljnih vidikov slovenskega jezikovnega načrtovanja. Govorci slovenščine kot tujega jezika bi morali biti vključeni »pri vseh fazah načrtovanja sodobnih jezikovnih priročnikov slovenščine, od faze oblikovanja besedilnega korpusa do jezikoslovne in leksikografske obdelave gradiva« (Stabej 2004: 12). Da bi ustrezno teoretično in praktično podprli njihovo usvajanje slovenščine ter ga s tem olajšali in pospešili, je treba še bolje spoznati načela, po katerih poteka. Prav to pa omogočajo korpusi usvajanja tujega jezika.

Korpusi besedil se v jezikoslovju uporabljajo že več stoletij. Slovaropisje je od nekdaj potrebovalo obsežne in relevantne vire podatkov. Delo, ki je potekalo ročno, je bilo zamudno in neizčrpno, težko ga je bilo ponoviti ali preveriti, napake pa so bile dokaj pogoste. Zato je bil za razmah besedilno usmerjenega jezikoslovja ključen razvoj računalništva. Korpusno jezikoslovje se je pojavilo v drugi polovici dvajsetega stoletja, vendar se do sredine osemdesetih let ni razširilo. Delo z računalniškimi zbirkami besedil je bilo omejeno na velike računalnike, do katerih je imelo dostop le malo raziskovalcev, ki niso mogli delati brez pomoči računalniškega strokovnjaka; v sedemdesetih letih je iskanje konkordanc – prikaza vseh korpusnih pojavitev določene besede ali besedne zveze skupaj z neposrednim kontekstom – ene besede v milijonskem

korpusu trajalo več kot eno uro (Kennedy 1998: 6). Danes je z vsesplošno prisotnostjo zmogljivih osebnih računalnikov in programske opreme korpusno jezikoslovje del lingvistike, ki se bliskovito širi po vsem svetu na najrazličnejša področja raziskovanja in uporabe.

Korpusno jezikoslovje je empirično, saj temelji na konkretnih podatkih. Tovrstne študije so prevladovale v zgodnjih petdesetih letih in v začetku šestdesetih, nato pa so jih pričenjale izpodrivati intuitivne raziskave generativne slovnice. Spor med predstavniki obeh tokov se je tako zaostрил, da sta se oblikovala ločena strokovna diskurza. Skupina okrog Chomskega je menila, da »naj bi govorec določenega jezika v desetih minutah produciral za jezikoslovce več ilustrativnega gradiva, kot ga je na voljo v večmilijonskih besedilnih korpusih« (Gorjanc 2003: 20), in dvomila o relevantnosti korpusnih podatkov. Drugi tok pa se je že od začetka usmerjal v uporabnost raziskav in se potrjeval v stiku z uporabniki jezikovnega znanja (prav tam). Vanj spada korpusno jezikoslovje, ki poskuša biti z dostopnostjo izhodiščnih podatkov o besedilih, njihovem zajemu in obdelavi čim bolj objektivno, velik obseg gradiva pa zmanjšuje možnost, da bi izjemne in obrobne jezikovne pojave interpretirali kot temeljne in tipične, s čimer povečuje verodostojnost analiz (Čermák 2005: 137). Danes je uporabnost korpusov splošno priznana in razširjena, vendar korpusna lingvistika ne konkurira drugim teorijam. Z njimi je večinoma komplementarna – ne nazadnje korpusi prikazujejo jezik brez vizualnega in družbenega konteksta, zaradi česar je vsaka posplošitev zgolj sklepanje (Hunston 2002: 23). Korpusi nudijo vire podatkov za vse teorije, ki izhajajo iz besedil (Kennedy 1998: 7). Tako se je oblikovala metodologija korpusne analize in novih jezikovnih opisov, ki namesto na intuiciji temeljijo na kvantitativno preverjeni jezikovni realnosti, osmišljeni šele z ustrezno kvalitativno razlago in interpretacijo podatkov (Gorjanc 2005: 7).

Korpusna analiza se je uveljavila tudi v preučevanju usvajanja tujega jezika. Ta proces so v 20. stoletju razlagali z različnimi dejavniki od okolja in genetske zasnove do spoznavnih procesov. V 60. letih dvajsetega stoletja so razlage skušale povezati behavioristično psihologijo s strukturalnim jezikoslovjem, v 70. je bila priljubljena teza, da učeči se samostojno ustvarja jezikovni sistem, v 80. pa so iskali močnejšo povezavo z ostalimi disciplinami (Sharwood Smith 1994: 22–23). Temu ustrezno so bile raznolike metode raziskovanja. Do šestdesetih in sedemdesetih let dvajsetega stoletja je

bila poglavitna kontrastivna analiza, nato jo je nadomestila analiza napak. Analizirane zbirke besedil so vsebovale največ dva tisoč besed in so bile predvsem heterogena zbirališča napak (Nesselhauf 2004: 129). Kontekst ali podatki o pogostnosti pojavov raziskovalcev niso zanimali (Tono 2002: 8), raziskovanje je potekalo ročno. Kmalu pa so se pojavila stališča, da ne bi smeli raziskovati zgolj napak, pač pa celotni jezikovni sistem učečega se (Richards, Sampson 1974: 5). Te poglede je zares utrdilo šele korpusno jezikoslovje s korpusi usvajanja tujega jezika, ki se ne osredotočajo samo na napake, ampak jih zanima vmesni jezik v celoti (Granger 1998a: 5–6). Razmahnilo so se v devetdesetih letih s korpusoma angleščine kot tujega jezika ICLE in LCLE (prim. poglavje 2.4.1). Leta 1999 je bil prvi simpozij na to temo v Hongkongu. Danes večina korpusov nastaja v Evropi in Aziji, nekaj pa tudi v Združenih državah Amerike. Po uvodnem navdušenju, ki ga je spremljalo optimistično prepričanje o širokih možnostih korpusne analize, se je vzporedno z razvojem korpusov pričelo krepiti tudi bolj prizemljeno spoznanje, da je analiza korpusnih podatkov formalna, pogosto omejena na naštevanje kvantitativnih rezultatov, zato jo je za kakovostne raziskave treba kombinirati s kvalitativnimi metodami in jo utemeljevati na novih teoretičnih predpostavkah.

Na Slovenskem so se zametki korpusnega raziskovanja začeli konec devetdesetih let. Ena prvih dosledno empirično zasnovanih raziskav je bila analiza jezikovne zmožnosti 36 udeležencev tečaja slovenščine kot tujega jezika, čeprav gradivo ni dostopno v obliki korpusa (prim. Pirih Svetina 2005). Knjiga, ki je pred vami, pa temelji na doktorski disertaciji *Korpus usvajanja slovenščine kot tujega jezika*, ki sem jo pod mentorstvom red. prof. dr. Marka Stabeja zagovarjala na Oddelku za slovenistiko Filozofske fakultete Univerze v Ljubljani decembra 2009. V okviru njenega nastanka sem spomladi 2005 sestavila, ročno uredila in označila učni korpus usvajanja slovenščine kot tujega jezika. Ker je šlo za preučevanje možnosti za urejanje in označevanje korpusa, je bil miniaturn in je pomenil zgolj nabiranje izkušenj za osrednji del raziskave, poskusni korpus usvajanja slovenščine kot tujega jezika (PIKUST, prim. poglavje 3.4). Ta je bil zbran, da bi pred gradnjo pravega korpusa usvajanja slovenščine kot tujega jezika (KUST) določila in praktično preizkusila natančnejše kriterije za označevanje in zajem besedil. Tovrstni eksperimentalni korpusi so »neke vrste poligoni za raziskovanje in pridobivanje izkušenj glede samega načina tehnične ureditve

korpusa« (Stabej 1998: 97) in so bolj kot jezikoslovnim raziskavam namenjeni treningu oblikovalcev večjih, bolj uravnoteženih korpusov. Kljub temu PiKUST dovolj nazorno ilustrira možnosti, ki jih lahko nudi dejanski KUST.

Besedila učečih se z različnimi jezikovnimi ozadji in nivoji znanja, ki sem jih vključila v PiKUST, so mi bila enostavno dostopna na Centru za slovenščino kot drugi/tuji jezik Filozofske fakultete Univerze v Ljubljani, s katerim sem sodelovala kot lektorica na tečajih slovenščine za tujce in kot organizatorica teh tečajev. Tekste sem dobila od udeležencev tečajev in v arhivih Izpitnega centra. Označila in uredila sem jih na Univerzi v Bergnu na Norveškem, na kateri sem od oktobra 2006 do januarja 2007 gostovala s štipendijo Evropske unije Marie Curie v programu BATMULT na centru Aksis. Pri vprašanih, kot je označevanje napak, sem za modele uporabljala že obstoječe korpusne usvajanja, posebej korpus norveščine kot tujega jezika ASK (prim. poglavje 2.4.7), v katerega sem imela največ vpogleda. Vendar so bili zgledi koristni predvsem po tehnični plati. Rešitev za angleščino ali norveščino ni mogoče avtomatično prenesti na slovensko gradivo, saj ima vsak jezik svojo strukturo, posebnosti in s tem povezane specifične napake. Pri opredeljevanju kategorizacije in načel za označevanje napak za slovenščino sem torej delno upoštevala priporočila tujih graditeljev korpusov, delno slovenistično tradicijo, v veliki meri pa sem izhajala iz problemov, ki so se pokazali v zbranem gradivu PiKUST-a.

Ko je bila kategorizacija v grobem postavljena, mi je pri podrobnejšem umeščanju napak in dilemi o navajanju pravih oblik pomagala anketa med slovenistično-jezikoslovno populacijo. Vprašalniki so bili sicer anonimni, toda ker je izobraževalni seminar Centra za slovenščino kot drugi/tuji jezik, na katerem sem anketo izpeljala, namenjen učiteljem slovenščine kot tujega jezika, prevajalcem ter drugim strokovno usposobljenim in zainteresiranim, je upravičeno sklepanje, da so odgovarjali ljudje, ki se profesionalno ukvarjajo s slovenščino kot tujim jezikom in so tudi potencialni uporabniki KUST-a. Vrnjenih je bilo 46 vprašalnikov, v katerih so morali anketirani dvajset avtentičnih napak tujcev, večinoma iz učnega KUST-a ali iz besedil, napisanih na tečajih slovenščine, razvrstiti po posameznih kategorijah in jim pripisati pravilno obliko. Primeri so bili izbrani tako, da naj bi bila polovica razmeroma enostavnih, ostali pa so ponujali več možnosti. Vnaprej predvidenih ali pravih rešitev ni bilo, saj je šlo predvsem za pridobivanje mnenja strokovno

usposobljene publike. Med pregledovanjem odgovorov se je izkazalo, da v enem od primerov zaradi skopega konteksta dejansko ni bilo napake, zato je bil izključen iz analize.

Pričujoča knjiga je sestavljena iz več delov. Poglavje 2 prinaša kratko opredelitev korpusov usvajanja tujega jezika, njihovega položaja znotraj korpusnega jezikoslovja in uporabljene terminologije. Obsežen del je pregled obstoječih korpusov usvajanja tujega jezika. Opozoriti je treba, da je v tiskani obliki dosegljivih le nekaj splošnih podatkov, tako da sem se morala zanašati na bolj spremenljive spletne vire. Eden od osrednjih problemskih sklopov, to je oblikovanje korpusov usvajanja tujega jezika, je obravnavan v poglavju 3. Pregledu rešitev po posameznih parametrih v obstoječih korpusih in opisu izvedbe v PiKUST-u sledi predlog za zasnovo uravnoteženega KUST-a v poglavju 4. Poglavje 5 se osredotoča na drugo veliko vprašanje raziskave, označevanje napak. Te so pojmovane kot del kreativne produkcije učečih se, del normalnega učnega procesa (Pirih Svetina 2005: 41). Teoretičnemu razmisleku in pregledu napak v tujih korpusih sledi predlog označevanja napak za slovenščino s primeri iz PiKUST-a. Poglavje 6 pa prinaša kratek pregled področij, na katerih se uporabljajo korpusi usvajanja tujega jezika, med katerimi je treba izpostaviti integrirano kontrastivno analizo ter aplikacijo korpusnih ugotovitev na izdelavo učnih gradiv, kot so slovarji, slovnice in učbeniki.

Zaključek uvoda je primerno mesto za zahvalo: red. prof. dr. Marku Stabeju za konkretno in splošno mentorstvo, sodelavkam Centra za slovenščino kot drugi/tuji jezik in norveškega Aksisa za pomoč na različnih frontah, ekipi Trojine za »šolarske« debate in vsem mojim za vse ostalo. Najpomembnejši pa so seveda vsi anonim(izirajni) udeleženci tečajev in izpitov iz znanja slovenščine, katerih besedila so z vključitvijo v PiKUST postala del raziskave – brez njih in njihove slovenščine tega dela res ne bi bilo.