

Milan Moguš

## LISTAJUĆI KOMPJUTORSKU KONKORDANCIJU KRLEŽINIH »BALADA«

U okviru kompjutorskih programa Zavoda za lingvistiku Filozofskoga fakulteta u Zagrebu obrađuju se tekstovi starije hrvatske književnosti i pismenosti. Rad je u ovoj fazi usmjeren prvenstveno na izradu kompjutorskih konkordancija spomenutih djela kao osnovu za najraznovrsnija leksička, stilistička i gramatička istraživanja. Neki su rezultati takva pothvata već objelodanjeni.<sup>1</sup>

Pri izradi samih kompjutorskih konkordancija elaborirani su i doradivani programi kao plod zajedničkoga rada elektroničara i jezikoslovaca.<sup>2</sup> Svladavši početna, relativno jednostavna konkordiranja, prelazilo se na teža, tekstualno neujednačena djela gdje, na primjer, osim osnovnog versificiranog teksta postoje i autorove prozne bilješke uz pojedine stihove, ili gdje se prozni tekst miješa sa stihovima, ili gdje uz hrvatski tekst dolaze i latinski sklopovi itd., a sve je to — u želji da se uz svaki jezični podatak nađe točna lokacija i kolokacija — trebalo precizno programirati. U tom je pogledu dosada najstroženiji posao bio kod konkordiranja *Razvoda istarskoga*. Naime, trebalo je posebno izraditi transliteracijski postupak iz glagoljice u latinicu u skladu s tastaturom kompjutorskog terminala, zatim predočiti, osim totalne konkordancije, sve razlike između temeljnog *Kršanskog rukopisa* i kasnijeg *Momjanskog prijepisa*, tj. posebno registrirati sve različite oblike iste riječi u ta dva teksta, ili odsutnost pojedine riječi, ili dodavanje neke riječi, ili čak drugačiji redosljed riječi.

Dakako, uspjeh je rodio novu želju: ići u elektroničkoj obradi još dalje, i to na tekstu koji bi — i formalno i poetski — bio složeniji predložak. Izbor je pao na Krležine *Balade Petrice Kerempuha*.

Elektronička je obrada *Balada* zamišljena tako da obuhvati nekoliko glavnijih sklopova, npr. izradu potpune konkordancije, izradu unazadne konkordancije, izradu frekvencijskih lista leksemâ itd. Unutar će se tih cjelina izra-

<sup>1</sup> Željko Bujas: *Kompjutorska konkordancija Gundulićeva »Osmana«*, izd. Liber, sv. 1. i 2., Zagreb, 1975, str. 1—492 i 493—939. — Milan Moguš i Željko Bujas: *Kompjutorska konkordancija hrvatskih djela Marka Marulića*, izd. Institut za lingvistiku Filozofskoga fakulteta, Zagreb, 1974, str. 1—650. — Milan Moguš i Željko Bujas: *Kompjutorska konkordancija Karnarutićevih djela*, izd. Instituta za lingvistiku Filozofskoga fakulteta, Zagreb, 1975, str. 1—321. — Milan Moguš i Željko Bujas: *Kompjutorska konkordancija »Razvoda istarskoga«*, izd. Institut za lingvistiku Filozofskoga fakulteta, Zagreb, 1976, str. I—XVIII i 1—383.

<sup>2</sup> U projektu su sudjelovali prof. dr Željko Bujas i prof. dr Milan Moguš kao voditelji, mr Maja Bratanić-Čimbur kao asistent i dipl. ing. Milutin Cihlar kao autor elektroničkih programa.

divati programi za posebne, specijalne zadatke, kao što su npr. različite selektivne konkordancije. Ukratko: nastojat ćemo stvoriti podlogu za razlučivanje ili povezivanje što je moguće više elemenata na svim jezičnim razinama.

Dosada smo uspjeli izvršiti prvi zadatak, tj. izradili smo prvu knjigu pod naslovom *Kompjutorska konkordancija Krležinih »Balada Petrice Kerempuha«* (Zagreb, 1977).<sup>3</sup> Knjiga ima dva dijela: u prvom se dijelu nalazi potpun tekst svih balada, što zaprema 30 dvostupačnih stranica velikoga formata. Redoslijed balada teče prema »Zorinu« izdanju *Sabranih djela Miroslava Krleže* (sv. 10, Zagreb, 1970). Svaka balada ima svoj broj (od 1 do 34), zatim oznaku iz koje se vidi nalazi li se riječ u naslovu, podnaslovu ili stihu te broj retka (u naslovu i podnaslovu) odnosno broj stiha. Smatrali smo važnim ovo raščlanjivanje, osobito odvajanje podnaslovâ kao zasebne jedinice jer se u njima mogu naći — kao što je poznato — najraznovrsniji tekstovi od citatâ i posvetâ do mota na hrvatskom, latinskom ili njemačkom jeziku. Naslovi, podnaslovi i stihovi imaju, kako je registrirao kompjutor, 2906 redaka, odnosno 15.041 riječ. Tu masu od preko petnaest tisuća riječi konkordirao je veliki kompjutor IBM 370/155 na magnetskoj vrpici za oko 20 minuta. To znači da je za veoma kratko vrijeme kompjutor izdvojio iz teksta svaku riječ u onom obliku u kojem se nalazi u tekstu, pridodao joj obostrani kontekst tako da natuknica bude u sredini i zatim točno abecedirao sve te retke prema izdvojenoj natuknici te precizno označio sa strane potpunu lokaciju u tekstu *Balada* (tj. broj balade, broj stiha, redak u naslovu ili podnaslovu). Otiskivanje s magnetske vrpce na kompjutorski ispis trajalo je također dvadesetak minuta. Tako smo dobili na 280 strana potpunu konkordanciju Krležinih *Balada*.

Iako je, kao što je rečeno, ovom konkordancijom tek započeta kompjutorska analiza Krležina kajkavskog opusa, određeni se rezultati mogu već sada iznijeti. Pa — počnimo listati konkordanciju.

Pođimo od podatka da je konkordancija *Balada* totalna, tj. elektronički je konkordiran cjelokupan vokabular sa svim oblicima i svim potvrdama. Isticanje se spomenutoga podatka može činiti u prvi čas suvišnim jer se ne radi o ručnom (i mučnom) konkordiranju, a za stroj je banalna stvar hoće li sređivati nekoliko tisuća ili nekoliko desetaka tisuća ili nekoliko stotina tisuća jedinica. Međutim, nije problem u tome. Nego: ima li svrhe nanizati, na primjer, kod veznikâ svih 110 potvrda za *a*, odnosno 541 potvrdu za *i*, ili kod prijedloga 272 potvrde za *na*, odnosno 172 potvrde za *v* itd. Ako se tomu dodaju neke mnogobrojnije potvrde kod ostalih vrsta riječi (npr. prilog *kak* sa 325 potvrda, zamjenica *kaj* sa 138 potvrda, 3. lice singulara prezenta *je* sa 395 potvrda itd.), dobivamo masu koja može djelovati tako da od stabala ne vidimo šume. Postavlja se, dakle, pitanje svrsishodnosti potpunoga ispisa, pogotovu sada kad — zahvaljujući kompjutoru — potpuni ispisi ne predstavljaju nikakve tehničke teškoće.

Dosadašnje analize pokazuju da treba ići na totalne konkordancije jer u takvim potpunim ispisima potvrđâ, koliko god dugački bili, gotovo nema balasta. Navest ću nekoliko ilustracija.

<sup>3</sup> Radni se primjerak nalazi u Zavodu za lingvistiku Filozofskoga fakulteta u Zagrebu.

Rekosmo da u *Baladama* ima 541 potvrda za veznik *i*. Zahvaljujući činjenici da su sve potvrde grupirane na jednome mjestu sa svim kolokacijama, možemo u potpunosti razmotriti uporabu veznika *i*.

Veznik *i* je ne samo najfrekventniji veznik u Krležinim kajkavskim pjesmama, nego najučestalija riječ. Iz kolokacijâ se razabire da daleko češće vezuje dvije riječi negoli dvije rečenice, a da može stajati i na početku rečenice. Ali to je tek prva, izvanjska slika veznika *i*. Kad se malo dublje zaviri u to vezivo, kad se svi primjeri podvrgnu pažljivijoj analizi, onda se počne pokazivati da se Krležino vezivanje obavlja unutar istovrsnih gramatičkih i semantičkih grupa. Konjunkcija *i* povezuje u najvećem broju slučajeva kod imenica i pridjeva jedninu s jedninom (*oslek i volek*), množinu s množinom (*pekve i zdele*), padež s padežom (*s Francuzi i z Madžari*), a kod glagola infinitiv s infinitivom (*piti i jesti*), prezent s prezentom (*žge i peče*), imperativ s imperativom (*sedni i napiši*), isto tako toponim s toponimom (*Ptuj i Ormož*, *Sisek i Bužin*) itd., itd. Tako je i na semantičkoj osi, tj. spajaju se samo istoklasne jedinice, a ne raznoklasne. Primjerice, na jednoj se strani nalaze *veliki i zmožni, poglaviti i velemožni, veleučeni i poglaviti, špani i kapetani, biskupi i stožerniki, hatmani i koloneli*, a na drugoj *pobožni i podložni, kokošar i tat, kramari i kukavci, grintavci i slinavci, cafute i fakini, fačuk i smujin sin*. Nema povezivanja predstavnika jedne grupe s predstavnicima druge. Do takvih se podataka može doći zahvaljujući upravo potpunom ispisu gdje momenat slučajnosti otpada. A takvi podaci, opet, počinju dobivati još veću vrijednost kad se utvrdi da su već od hrvatske srednjovjekovne književnosti — kako čitamo npr. u glagoljaškim tekstovima Pariškoga kodeksa<sup>4</sup> — na jednoj strani bili *grđinali i biskupi, biskupi i opati*, a na drugoj *mala bratja i koludri predikavci*. Tada i taj segment pokazuje da su u pravu oni malobrojni istraživači koji su brižljivim analizama došli do spoznaje da je izražajna, sadržajna i misaona srodnost davne glagoljaške poezije prismo povezana s Krležinim kajkavskim baladama.<sup>5</sup>

Takva se istraživanja mogu sada, kad postoji totalna konkordancija, provesti kod svih oblika koji imaju više od jedne potvrde. Jer, tek na temelju neke vrste paradigmatškog odnosa svake jedinice sa svakom istoga glasovnog kostura otkriva se prava funkcija pojedine natuknice. Tako, na primjer, sada je relativno lako odvojiti sve primjere prezenta *da* od veznika *da* i, pošto se to učini, može se odmah pristupiti opisu različite sintaktičke upotrebe veznika *da* i njegova različita položaja u rečenici. Može se vrlo brzo uočiti da jedanput stoji *da* u izričnoj rečenici (*tak je na svetu da za najvećšeg suca smert se s kosum vre okolo smuca*), drugi put u namjernoj (*a da bi ga oprali od smertnoga greha, s terbuha su mu zrezali za orgule meha*), treći put u načinskoj (*zmutra vočje lakom ftič, kak stvoril da ga je hudič*), četvrti put u dosta neobičnom redu riječi vrši zajedno s glagolom zapovjednu funkciju (*čuli su se glasi kak z terbuha basi: Vivat Gašpar, v nebe dobre došel da si!*) itd.

<sup>4</sup> Dragica Malić: *Jeziik najstarije hrvatske pjesmarice*, Znanstvena biblioteka HFD, Zagreb, 1972.

<sup>5</sup> Radoslav Katičić: »Zapisi s izvorišta«, *Slovo*, br. 25—26, Zagreb, 1976, str. 393—406.

Ali konkordancija ne pruža samo mogućnost da se u potpunosti obuhvate sve funkcije iste riječi, nego i da se razmotre različiti oblici za isti sadržaj. Riječ je, dakako, o dubletama. Taj je element veoma važan u Krležinu kajkavskom opusu. Kompjutorska konkordancija *Balada* omogućuje da se za vrlo kratko vrijeme, bez onog dugog mukotrpnog ekscerpiranja, izradi kompletna slika leksemnih dubleta. Ali već je sada moguće dopuniti neke konstatacije u vezi s pojedinim dubletama<sup>6</sup>. Tako se, na primjer, sa sigurnošću mogu registrirati ne samo sve različitosti početnoga skupa čr- (kao čr-, čar-, čer-, cer-), nego se iz kompletnog materijala vidi a) da daleko najviše leksema ima slijed čer-, b) da se pridjev u značenju 'crn' ostvaruje samo kao čar-, a njegove izvedenice najčešće kao čer- i cer-, c) da i pridjev u značenju 'crven' (ukoliko ne glasi *rdeči*) ima početno č- s dvije grupe dubletnih realizacija: u jednoj je čarv- (*čarvendač*) i červ-, a u drugoj črl- i čerl- (nikada črv-, odnosno čarl-). Ovakve se dublete leksemâ mogu sada sveobuhvatnije analizirati i s dijakronijskoga gledišta. Primjerice, bez ikakve bojazni da nam je nešto u dokumentaciji promaklo možemo zagledajući u konkordanciju sasvim pouzdano reći da u *Baladama* postoje dublete *dan* i *den*, odnosno *danas* i *denes*, da je *dan* frekventnije od *den* (odnos 34 : 5), ali je *denes* frekventnije od *danas* (11 : 2) i, konačno, da nema hibridâ *danes* ili *denas* jer ih nema ni u kajkavskim organskim govorima. Slična je situacija i s leksemom u značenju 'misa' što se realizira dubletno kao *maša* i *meša* (s frekvencijom 7 : 1), ali je više izvedenicâ i češće su od manje frekventnog oblika *meša* negoli od frekventnijeg *maša*. Dakako, mnoge riječi nemaju takvih dubleta iako bi ih — teoretski — mogle imati, kao npr. *deska*, *melin* 'mlin', *megla*, *pes* i dr. Ali ako *pes* nema dublete sa *pas* ima sa — *cucek*. I tu bi već moglo početi ono stilističko »fino tkanje« jer su u konkordanciji svi *pesi* i svi *cucki* na okupu. Pokazuje se, naime, da su kolokacije tih dubleta ponešto drugačije. *Cucek* je *lovni* i *verni*, a *pes* je *stekli*, *peklenski*, *kervavi* i *carli*. Razlika je sasvim očita. A da se uoči ta razlika, ili one o kojima sam govorio, treba samo nekoliko minuta iako je pred nama korpus od 15.041 riječi.

Vjerojatno ne treba posebno ni naglašavati koliko su ovakve parcijalne analize samo priprema, samo dobra podloga da se krene dalje, u zamašnije pothvate kakav je npr. izrada rječnika Krležinih *Balada*. A kad uskoro dobijemo i ostale proizvode elektronički sređenih podataka ovoga teksta, bit će, nadam se, stvoreni svi preduvjeti da se počne svestranije i meritornije govoriti o izboru i dometu riječi Krležine »Reči«.

<sup>6</sup> Uspor. Milan Moguš: »Što pokazuju dublete u Krležinim *Baladama*«, *Forum*, sv. 9, Zagreb, 1973, str. 376—386.