

K računalniško podprtim empiričnim raziskavam literature

Udo Kuckartz: *Einführung in die computergestützte Analyse qualitativer Daten*. 2., aktualizirana izdaja.

Wiesbaden: VS Verlag für Sozialwissenschaften, 2007. 268 str.

Urška Perenič

Oddelek za slovenske jezike in književnosti, Koroška cesta 160, SI-2000 Maribor
urška.perenic@uni-mb.si

Monografija Uda Kuckartza je predelana oblika izdaje *Computergestützte Analyse qualitativer Daten* iz leta 1999 in je zamišljena kot splošen uvod v računalniško podprto empirično raziskovanje tekstov, in sicer analizo kvalitativnih podatkov. Čeprav so aplikativni primeri, ki jih ponuja, vezani na neliterarne tekste (intervjuje, ankete ipd.) in monografija s tega vidika še najbolj nagovarja sociološko znanost, pa je lahko ob ustreznih modifikacijah in kritični uporabi relevantno študijsko gradivo ter metodološki doprinos z možnostjo bolj učinkovite vsebinske analize za vse tiste, ki se ne glede na disciplinarno področje tako ali drugače ukvarjajo s teksti. Sem spadajo tudi literarni znanstveniki. Kar se tiče empiričnih obravnav literature, imamo tudi v slovenskem prostoru kar nekaj raznorodnih primerov, ki segajo od sistemskih obravnav do kvantitativnih analiz literature (med najzgodnejšimi sta bila P. Scherber in D. Poniž, sledili so M. Hladnik, v doktorski nalogi A. Bjelčevič, M. Pezdirc Bartol, tudi A. Žbogar, medtem ko so se s sistemsko obravnavo ukvarjali M. Juvan, D. Kos, M. Dovič, U. Perenič idr.). V vrsti sorodnih raziskav ni mogoče zaobiti Scherberjevega *Slovarja Prešernovega pesniškega jezika* (1977), ki dela le s konkordancami, in Zupanove monografije *Kaj je Prešeren rekel o —: Poezije in konkordance* (2001) s statistično analizo Prešernovih pesmi in zbirom konkordanc za samostalnike, na podlagi katerih prihaja npr. do spoznanja o gibkost izražanja in globini pesnikove misli. Marko Limbek, sicer matematik in ekonomist, je z računalniško podprto analizo ugotavljal avtorstvo na primeru Mencingerja (*Usage of multivariate analysis in authorship attribution: did Janez Mencinger write the story 'Poštena Bohinčeka'?*, 2008). Če razpolagamo s prikazom sistemskih obravnav literature (Dovič), pa nimamo sorodnega uvoda ali prikaza, v katerem bi bile tako sistematično predstavljene možnosti računalniško podprtih empiričnih kvalitativnih analiz literarnih tekstov.

Knjiga *Einführung in die computergestützte Analyse qualitativer Daten* je razdeljena na 14 poglavij s priloženo, seznamom literature in slovarčkom

glavnih pojmov, vezanih na predstavljene raziskovalne metode. Razdelitev poglavij spominja na delovne korake, ki jih je potrebno izvesti v postopku računalniško podprte kvalitativne analize teksta. Na koncu so poglavja vedno opremljena s praktičnimi napotki, ki jih bralec lahko preizkusi z računalniškim programom MAXQDA, ter vprašanji, s katerimi lahko preveri svoje razumevanje teorij in modelov. Ob ATLAS-u je MAXQDA eden najbolj razširjenih programov za profesionalno tekstno analizo v nemškem prostoru; za zainteresirane uporabnike je dostopen preko spleta v nemški in angleški verziji. Prikaz se bo osredotočil na tista poglavja iz *Uvoda*, ki prinašajo različne možnosti računalniško podprte kvalitativne analize. Zaradi omejenega obsega in namena članka, ki želi dati vpogled v manj uveljavljene empirične pristope k tekstom, se bomo omejili na tiste metode, ki se dajo prenesti na literaturo in utegnejo biti najbolj zanimive in uporabne za literarnega raziskovalca. Ob strani bo – ob zavedanju, da ne gre za nič manj pomembne korake – ostal prikaz računalniških tehnik za analizo ter faze formatiranja, transkribiranja, priprave, skeniranja in importiranja tekstov, kakor so na kratko predstavljene v drugem poglavju. Težišče bo na problematiki kategoriziranja in kodiranja (literarnega) teksta, ki predstavljata osrednja koraka tekstne analize in s katerima se ukvarjajo predvsem 3., 4., 6. in 12. poglavje. Vmesnih poglavij, ki se ukvarjajo z načini ponovnega iskanja, organizacije, primerjave in vizualizacije že obdelanih kvalitativnih podatkov in postanejo zanimiva šele v sklepni raziskovalni fazi vrednotenja rezultatov, se bomo dotaknili, kolikor bodo smiselna pri razumevanju kodiranja teksta.

Kvalitativna tekstna analiza, v katero aktivno poseže tudi literarni raziskovalec, se začena s fazama kategoriziranja in kodiranja. Računalniški program, ki je za tako analizo pomemben z več vidikov, sam po sebi ne pomeni nič brez napore raziskujočega. Ta mora namreč temeljito predihati material za analizo ter smotrno, skladno s problemi in cilji raziskave, izbrati in sporočiti programu enote za obdelavo. Poudariti je treba, da sta obe fazi zaradi večje preglednosti kakor pri Kuckartzu tudi tu predstavljeni ločeno, čeprav sta najožje povezani in ju, kakor bo razvidno iz nadaljevanja, v procesu raziskovanja ni mogoče ločiti. Kategorizacija vključuje skrbno in večkratno prebiranje tekstnega materiala. Na začetku je smiselno oblikovati delovno verzijo teksta z oštevilčenimi odstavki ali drugimi relevantnimi segmenti teksta in označiti mesta, ki utegnejo priti v poštev pri dokazovanju hipotez in poznejši analizi. Kategorije so lahko posamezne besede, besedne zveze, stavki, odstavki ali obširnejši segmenti, ki naj bi bili praviloma tako obsežni, da so razumljivi tudi zunaj konteksta. Označijo se s t. i. marker-tehniko in služijo kot kašipoti k temam v tekstu. Iz njih se da pozneje razbrati semantična polja. Opremljanje tekstnih segmentov s pre-

mišljeno izbranimi analitičnimi kategorijami, ki se razlikujejo glede na raziskovalčeve hipoteze, cilje, namene in interpretativne procese, je smiselno, saj se pri tem identificira za raziskavo ključne fenomene v tekstu, tako da se jih v fazi obdelave in interpretiranja lažje najde in poveže. Temu sledi opremljanje teksta s kodi. Kakor je kategoriziranje aktivnost raziskovalca – seveda v komunikaciji s tekstom –, tako tudi kodi niso samo deskriptivne enote, ampak enote smisla in rezultati intelektualnega raziskovalnega dela, s katerimi se poimenujejo predhodno in skrbno izbrane kategorije. Kodi postanejo v nadaljnjem procesu raziskovanja nekakšni merilni instrumenti, s katerimi se analizira tekst.

V sociološki znanosti obstaja več oblik kodiranja. Bolj ali manj uporabljeni za analizo literarnih tekstov se mi zdijo sledeči predlogi:

1. teoretsko kodiranje,
2. tematsko kodiranje,
3. parafrazirajoče / povzemajoče kodiranje.

Tipičen primer teoretskega kodiranja je t. i. utemeljena teorija (**Grounded Theory**). Izhajamo iz že obstoječe teorije; sestavljena je iz določenega števila kategorij, njihovih lastnosti in odnosov med njimi, ki zajemajo teoretsko relevantne aspekte. V analizi se te vnaprej izdelane teoretske kategorije pripišejo fenomenom v tekstu. Teoretsko kodiranje je sestavljeno iz treh korakov: odprtega, osnega in selektivnega kodiranja, ki spada bolj v območje vrednotenja in interpretiranja rezultatov. Še najbolj pomembno je odprto kodiranje. Pozornost je sicer usmerjena na razvijane konceptualnih kategorij za poimenovanje obdelanega in razčlenjenega teksta, vendar so te po drugi plati večinoma že razvite, saj izhajajo iz vnaprej izbranega teoretičnega koncepta. Pri osnem kodiranju je poudarek na vzpostavljanju povezav med različnimi (teoretskimi) kategorijami, tako da nastajajo paradigme oz. razredi kategorij. S tem se postopoma izgrajuje oz. obnavlja teorija. Čeprav poteka teoretsko kodiranje na ozadju predhodno dane teorije, ne moremo govoriti o mehanskem izvajanju postopka po zgoraj opisanih korakih in popolni izključitvi raziskovalca. Ta je npr. pomemben pri izbiri teoretskih kategorij, ki se upoštevajo na gradivu.

Tematsko kodiranje ima s teoretskim nekaj skupnih točk, sicer pa menim, da gre bolj za zdravorazumsko tehniko, saj se pri raziskovanju izhaja iz tihega dogovora, da je vsak raziskovalec v luči teoretsko-zgodovinskih postavk sposoben izbrati relevantne enote za analizo, izluščiti iz teksta ključne tematske aspekte za analizo in za njihovo poimenovanje ne potrebuje kakšnega trdnega teoretskega okvirja. Po drugi plati je ves čas prisotna določena negotovost glede metodične korektnosti. Preizkušeno obliko tematskega kodiranja po avtorjevem mnenju predstavlja Hopfova metoda. Tematsko kodiranje oz. razčlemba teksta sicer temelji na (teoretskem)

predznanju, pri čemer ne gre za izgrajevanje, ampak bolj za preizkušanje in nadgradnjo doslej obetavne teorije. Analitične kategorije zato niso nekaj fiksnega, ampak se oblikujejo v stiku z gradivom. Po določitvi formalno-tematskih kategorij za analizo sledi kodiranje teksta. Posamezni segmenti teksta se zdaj dosledno označijo in povežejo. Sledi poglobljena analiza z vrednotenjem rezultatov.

Podobna je povzemajoča kvalitativna vsebinska analiza. Kuckartz se opre na Mayringov predlog, ki je nastal v delni opoziciji z vsebinsko analizo, kakršna se je uveljavila v komunikacijskih raziskavah. Tudi tu igrajo osrednjo vlogo kategorije. Kvalitativna vsebinska analiza skuša jezikovni material sistematično razčleniti in analizirati. Na ozadju teorije razvije kategorialni sistem, predloži pa tudi vidike, ki so predmet analize. Kategorialni aparat je namenjen temu, da je analiza sistematična, skozi kategorije pa se gradivo skrči do te mere, da se ohranijo le ključne vsebine, ki naj bi še vedno odlikavale celotno gradivo. Mayringova analiza poteka v več korakih. Začne z določitvijo analitičnih enot, čemur sledi parafraziranje vsebinsko relevantnih segmentov. Zatem določi nivo abstrakcije parafraz. Naslednja stopnja je selekcija oz. črtanje pomensko enakih parafraz. Sledi povezovanje kategorij, tj. parafraz, in interpretacija. Korak za korakom se gradivo skrči in spravi na ustrezen nivo abstrakcije. V primerjavi s tematskim in posebej teoretskim kodiranjem je ta pristop tudi po Kuckartzovem mnenju najbolj induktiven, saj izhaja zlasti iz tekstov. V ospredju je torej skrbno in metodično kontrolirano povzemanje teksta z izgradnjo kategorialnega sistema, ki vodi do osrednjih semantičnih polj teksta.

Teksta se je mogoče lotiti tudi po leksikalni poti, in sicer skozi iztočnice, gesla in ključne besede. Iz tega se lahko oblikuje t. i. KWIC-lista (Keyword-in-Context), ki vsebuje ključne besede, besedne verige, v katerih se iskana beseda pojavlja, in pripadajoči kontekst. Na ta način dobimo npr. pregled besed s pripadajočimi konteksti in pomenskimi polji, ki strukturirajo tekst.

Sorodna leksikalni je vsebinska analiza teksta skozi sezname besed, indekse in slovarčke. Podlaga za analizo je predhodno izdelani slovar, ki je sestavljen iz kategorij, tem pripadajočih besed in njihovih kombinacij. Z njim se lahko sistematično spremlja pojavljanje iztočnic v tekstu. Pri računalniško podprti analizi se sproti beleži frekvenca kategorij. Priporočljivo je oblikovanje seznama vseh relevantnih in iskanih besed, urejenih po pogostosti, abecedi ali kako drugače in opremljenih z referencami, tj. mesti v tekstu, ki se nanašajo nanje. V primerjavi z zgornjo vrsto analize, kjer se kodirajo predvsem vpadljive in izstopajoče besede in kombinacije iz teksta, je za to kodiranje značilna večja mera sistematičnosti pri urejanju analitičnih kategorij, medtem ko vidim njegovo uporabnost prav tako v

izluščenju semantičnih polj teksta. Iz pripadajočih kontekstov in pogostosti pojavljanja posameznih kategorij bi se dalo razbrati motivne in tematske kroge. Ti so lahko literarnem raziskovanju zanimivi iz različnih razlogov; med drugim se po tej poti odprejo vprašanja komunikacijske funkcije teksta, njegove vključenosti v družbene diskurze.

Kuckartz v podporo kvalitativni analizi predlaga še sprotno opremljanje teksta z opombami, ki lahko preprečijo izgubo vmes porojenih idej raziskovalca in lahko pozneje pomagajo pri vrednotenju rezultatov. V poglavju 7 je pomen pripisan beleženju in sprotnemu organiziranju lastnih idej, pisanju komentarjev in vtisov o gradivu, ki se porajajo v osrednjih fazah kodiranja in kategoriziranja. To podpira tudi računalniška funkcija »Memos«. Opombe gradijo most k zaključni fazi raziskovanja, tj. k vrednotenju in interpretiranju rezultatov, in so lahko pomemben pripomoček pri potrjevanju ali zavračanju hipotez, razvijanju teorije ali sestavljanju zaključnega poročila s predstavitevijo rezultatov.

Bistveno pri predstavljenih pristopih k tekstom se mi zdi, da teksti niso nikoli razumljeni kot neskaljena in manifestna odslikava pišočega ali recipientovih reakcij, ampak da je vseskozi prisotno zavedanje, da sta kategorizacija in opremljanje teksta s kodi v veliki meri domeni raziskovalca. Ker kategorije in kodi ne padejo »kot mana iz nebes«, ampak so rezultat težkega dela ter raziskovalčeve komunikacije in pogajanja s teksti na ozadju védenja, idej, teorij, delovnih hipotez in ciljev (intelektualno kodiranje), je jasno, da so rezultati take analize vselej intelektualni dosežki raziskovalca. Vsebinska analiza je učni in ustvarjalni proces raziskujočega, ki v spremenjenem socio-kulturnem kontekstu lahko izbere druge kategorije za analizo. Kot taka ne vodi do ontološke resničnosti, skuša pa na ozadju skupnega védenja, ki velja znotraj stroke oz. raziskovalne skupine, ter s pozornostjo, ciljno izbiro analitičnih kategorij tvoriti izjave o kompleksnih (literarnih) procesih preteklosti. Ti so zdaj lahko osvetljeni z drugačnega zornega kota. Pri predmetu raziskovanja lahko odprejo doslej spregledane vidike. Rezultati analize se dajo poleg tega z uporabo istega postopka intersubjektivno preveriti.

Prednost take analize je tudi, da se da po enakih pravilih raziskovati množico tekstov, rezultate pa zatem povezati s statističnimi metodami, organizirati in vizualizirati skozi računalniške funkcije, ki jih ponuja MAXQDA (predvsem poglavji 8 in 10). S takim razčlenjevanjem se gotovo ne da zajeti vse kompleksnosti teksta, pomenov in semantične diferenciacije. Nevarnost kategorizacije in kodiranja je, da se tekst spartikularizira. Toda če vemo, da je za sleherno metodo in teoretski diskurz značilno opazovanje izbranih segmentov in lastnosti, premišljena in konsistentna uporaba te metode ne more biti ovira.

Kar se tiče računalniške podpore pri analiziranju, se je mogoče strinjati, da program s celim spektrom funkcij za izgradnjo paradigem, klasiifikacij, tipologij, vzorcev in vizualizacijo omogoča hitrejšo, zanesljivejšo, pregledno in sistematično obdelavo podatkov in celih korpusov ter da je metodološko kontrolirana analiza in vrednotenje rezultatov nedvomno prednost, ki jo je vredno izkoristiti. Po drugi plati pa se izkaže, da so formalno-strukturni elementi za tekstno analizo vedno prepuščeni raziskovalcu. Tudi U. Kuckartz je mnenja, da vsem računalniškim programom navkljub ne obstaja nobena avtomatična analiza teksta, da pa je z njihovo uporabo tako zaradi hitrosti kot velikega repertoarja funkcij za obdelavo teksta nedvomno dosežena nova in višja stopnja tekstne analize.

Leta 2008 je bila pri založbi VS Verlag für Sozialwissenschaften v Wiesbadnu – prav tako zaradi očitnega uspeha pri (študentskem) bralstvu in odziva na polju empiričnih raziskav – drugič natisnjena knjiga *Qualitative Evaluation; Der Einstieg in die Praxis*, tokrat v sodelovanju Kuckartza s Thorstnom Dresingom, Stefanom Rädikerjem in Clausom Steferjem. Z ozirom na faze računalniško podprte kvalitativne analize teksta, ki vključuje že omenjene faze priprave in importiranja teksta, njegove kategorizacije s kodiranjem ter vrednotenja rezultatov, jo razumem kot nadaljevanje tu predstavljene knjige, saj skuša zgornje metode aplicirati na konkretno gradivo in pokazati, kako je skozi nje mogoče izpeljati kvalitativno raziskavo teksta. To pripomore k boljšemu razumevanju predstavljenih oblik kodiranja, v razvid pa pridejo tako prednosti kot slabosti takih analiz. Knjiga je dragocena, saj reflektira računalniško podprte kvalitativne pristope k tekstom. Odločitev za metodo je seveda še vedno prepuščena presoji raziskovalca glede na zastavljeno znanstveno vprašanje.

Maj 2009