

Manfred Stede, *Korpusgestützte Textanalyse, Grundzüge der Ebenen-orientierten Textlinguistik*, Gunter Narr Verlag, Tübingen 2007.

U nakladi izdavačke kuće Gunter Narr Verlag, Tübingen izašla je 2007. godine knjiga Manfreda Stedeja pod naslovom *Korpusgestützte Textanalyse, Grundzüge der Ebenen-orientierten Textlinguistik*. Autor temu iz naslova obrađuje iz dviju različitih perspektiva. Iz perspektive tekstne lingvistike donosi sustavan pregled različitih razina analize teksta, a iz perspektive računalne lingvistike upućuje na potrebu jačeg povezivanja navedenih lingvističkih disciplina i na mogućnosti primjene računalne lingvistike u bavljenju tekstnom lingvistikom.

Polazeći od činjenice prema kojoj su tekstovi višeslojni objekti, autor smatra da pojedine razine teksta najprije treba odvojeno promatrati i analizirati, a nakon toga međusobno sustavno povezati. Kako prema autorovim riječima još uvijek ne postoji općeprihvaćeni model za modeliranje međuzavisnosti različitih razina teksta, on smatra da u izradi takvog modela važnu ulogu ima računalna lingvistika. Prvi korak je anotiranje, odnosno obogaćivanje primarnih podataka autentičnih tekstova, podacima koji proizlaze iz lingvističke analize. To mogu biti različite vrste podataka kao što su primjerice vrste riječi, anaforička i kataforička sredstva, govorni činovi itd. Pritom je poželjno koristiti alate za anotaciju koji su prikladni za pojedine razine analize teksta te se pridržavati jasnih smjernica za anotiranje kako bi na taj način obrađeni podatci mogli poslužiti i lingvistima koji te tekstove nisu osobno anotirali. Sukladno metodološkom pristupu *Mehrebenen-Annotation* (višeslojno anotiranje), razvijenom na Sveučilištu u Potsdamu, na kojem je autor knjige profesor primijenjene računalne lingvistike, anotirani podatci o različitim razinama pojedinih tekstova ili tekstnih vrsta korisni su isključivo ako se na osnovi njih može doći do saznanja o međuzavisnosti različitih razina teksta. U tu svrhu se na Sveučilištu u Potsdamu koriste postojeći softveri za anotaciju pojedinih razina, ali se izlazne datoteke automatski pretvaraju u generički format XML, što omogućuje analizu međuzavisnosti razina koje su anotirane različitim alatima. Tako anotirani podatci predstavljaju važan izvor za bavljenje autentičnim tekstovima te oni moraju biti raspoloživi u elektronskom obliku u različitim lingvističkim bazama podataka, što olakšava pretraživanje i obradu dobivenih podataka. Na Sveučilištu u Potsdamu je u tu svrhu razvijen format za razmjenu podataka pod nazivom PAULA te postoji lingvistička baza podataka pod nazivom ANNIS. Na mrežnoj stranici knjige <http://www.ling.uni-potsdam.de/~stede/KorpTA.html> nalaze se, među ostalim, detaljne informacije o raspoloživosti baze podataka ANNIS, upute za konvertiranje podataka u format PAULA te podatci o raspoloživim alatima za anotiranje.

Prvo poglavlje knjige pod naslovom *Einleitung und Übersicht* sadrži kratak prikaz materije obrađene u knjizi. Autor u kratkim crtama opisuje način korištenja alata za anotiranje i upozorava na nužnost korištenja autentičnih kor-

pusa tekstova u bavljenju tekstnom lingvistikom. Pritom smatra da je za rasvjetljavanje fenomena tekstualnosti potrebna dubinska analiza manjih korpusa.

Drugo poglavlje *Kohäsion, Kohärenz und Textualität* posvećeno je kriterijima tekstualnosti, a autor kao najvažnije ističe koheziju i koherenciju. Jezična sredstva za ostvarenje koherencije su rekurencija, supstitucija, leksička kohezija, korištenje određenog i neodređenog člana, tzv. pro-forme, elipsa, metakomunikacijski signali, glagolsko vrijeme i način, konektori i elementi koji oblikuju strukturu. Uz navedene kriterije tekstualnosti koji se odnose na jezičnu strukturu teksta, postoji niz kriterija koji se odnose na komunikacijsku situaciju, kao što su primjerice intencionalnost, prihvatljivost, informativnost, situativnost i intertekstualnost. Navedeni kriteriji ne smiju se promatrati kao definicija te svi oni nisu u jednakoj mjeri potrebni kako bismo određeni kompleksni jezični znak pojmlili i percipirali kao tekst.

U trećem poglavlju pod naslovom *Textfunktion, Textsorte und Texttyp* autor se bavi problematikom određivanja funkcije teksta te definiranja tekstne vrste i tipa teksta. Funkcija teksta uvjetovana je međudjelovanjem pojedinačnih dijelova teksta i jezičnim signalima koji upućuju na funkciju pojedine rečenice ili dijelova teksta te samim tim i na cjelokupnu funkciju teksta. Kriteriji za opisivanje i razlikovanje pojedinih tekstnih vrsta su grafički izgled, dužina teksta, poredak dijelova teksta, signali za segmentiranje teksta, pozdravne i završne formule, odabir riječi: tipični leksemi i kolokacije, vrsta i učestalost rečeničnog ustrojstva i gramatičkih kategorija, tema i način razvijanja teme, odnos tema – rema, funkcija teksta, medij komunikacije, komunikacijska situacija i područje djelovanja. Kao tipove teksta autor navodi deskriptivne, narativne, ekspozicijske, argumentativne i instruktivne. Smatra da postoje korelacije između tekstnih vrsta i tipova teksta, ali se ne može govoriti o čvrstoj i jednoznačnoj povezanosti.

Četvrto (*Referenzielle Struktur*), peto (*Thematische Struktur*), šesto (*Temporale Struktur*), sedmo (*Von Illokutionsstruktur zu Argumentationsstruktur*) i osmo poglavlje (*Rhetorische Struktur*) prikazuju različite razine analize teksta. Završni dio svakog spomenutog poglavlja posvećen je savjetima, uputama i smjernicama za anotiranje odgovarajućih razina teksta.

U četvrtom se poglavlju autor bavi referencijalnom strukturom teksta koju definira kao zbroj svih koreferencijalnih odnosa u tekstu. Autor dalje navodi kako odabir oblika referencijalnog izraza u konkretnom kontekstu ovisi o tome je li, kada i na koji način određeni predmet prethodno već bio spomenut. Kao argument za svoju tezu navodi teoriju *Centering* koja upućuje na međuzavisnost odabira zamjenica i gramatičke funkcije u rečenici. Softverski alat MMAX 2 posebno je prikladan za anotiranje koreferencije. Omogućuje jednostavno označavanje referencijalnih izraza te povezivanje koreferencijalnih izraza s pomoću miša.

Peto poglavlje posvećeno tematskoj strukturi teksta prikazuje različite pristupe utvrđivanju teme teksta kao cjeline te utvrđivanju glavne i sporednih tema. U dijelu u kojem se obrađuje informacijska struktura teksta navodi se kako struktura pojedine rečenice ne ovisi isključivo o gramatičkim uvjetima, već i o čimbenicima vezanim za prenošenje informacija. Kod informacijske struktu-

re teksta ne može se govoriti o tome je li nešto ispravno ili krivo, već jednostavno manje ili više primjereno. Faktori koji utječu na to pripadaju jezičnoj i nejezičnoj domeni te se mogu promatrati s različitih aspekata: semantičkog, pragmatičkog, tipološkog, sintaktičkog i sl. Anotacijska shema za tematsku strukturu morala bi na globalnoj razini obuhvaćati podjelu teksta u tematske jedinice, a na lokalnoj razini način prijelaza s teme na temu. Prikladan alat za navedene svrhe je EXMARaLDA koji je prvenstveno razvijen za transkribiranje govornog jezika, ali se može koristiti i za obradu pisanih tekstova.

Šesto poglavlje opisuje jezično označavanje vremenskih odnosa u tekstu. Tu funkciju obavljaju konektori, glagolska vremena, glagolski vid, način vršenja glagolske radnje i izrazi za označavanje vremena. U slučaju kada vremenski odnosi nisu eksplicitno izraženi koristi se princip ikoniciteta, što znači da čitatelj polazi od toga da se događaji u tekstu opisuju onim redom kojim su se dogodili. Posljednjih godina je u SAD-u razvijena obilata infrastruktura za anotiranje vremenskih odnosa i informacija. Na mrežnoj stranici *www.timeml.org* dostupne su specifikacije i alati za anotiranje pod skupnim nazivom TimeML.

Sedmo poglavlje donosi pregled različitih pristupa u određivanju ilokucija pojedinih dijelova teksta, različite taksonomije funkcija teksta te prikaz metoda i načina za označavanje argumentacijske strukture teksta. Određivanje smjernica za anotiranje ilokucijske strukture teksta prema autorovim je riječima izrazito teška zadaća i vrlo sklizak teren iz razloga što se, s jedne strane ne smije previše ograničiti intuicija anotatora, a s druge strane, potrebno je omogućiti određenu razinu usporedivosti dobivenih rezultata analize, što podrazumijeva barem identičan inventar ilokucija i određene zajedničke principe za strukturiranje tekstova. Iz navedenih razloga još uvijek ne postoje korpusi koji su anotirani podacima o ilokucijskoj strukturi, a postoji vrlo malo prijedloga konkretnih anotacijskih shema.

Osmo poglavlje prikazuje načine opisivanja sadržajnih veza u složenim jezičnim izričajima u kojima aspekt djelovanja nije nužno u prvom planu. U literaturi postoji čitav niz prijedloga za klasifikaciju odnosa koherencije u tekstu. Neki se autori pritom prvenstveno oslanjaju na introspekciju, neki polaze od jezičnih fenomena, a neki svoje teorije temelje na psiholingvističkim modelima i eksperimentima. Tekstovi iz Potsdamer Kommentarkorpusa anotirani su prema modelu RST (*Rhetorical Structure Theory*) uz korištenje alata RSTTool.

Deveto poglavlje pod naslovom *Minimale Texteinheiten und ihre Verknüpfung* u uvodnom se dijelu bavi problematikom određivanja minimalnih jedinica teksta. Središnji dio poglavlja posvećen je jezičnim indikatorima za određivanje ilokucija minimalnih jedinica teksta, kao što su performativne formule, vrste rečenica, modalni glagoli i prilozi. Završni dio poglavlja obrađuje načine povezivanja segmenta pojedinog teksta. U tu se svrhu najčešće koriste konektori koji se na temelju svojih sintaktičkih obilježja mogu podijeliti u veznike, priložne oznake i prijedloge.

U završnom poglavlju *Interaktionen zwischen Ebenen: Entstehung von Kohärenz* autor ponovo upozorava na ono što bi po njegovu mišljenju trebao biti osnovni princip u bavljenju tekstnom lingvistikom: *divide et explana*. To znači

da je najprije potrebno analizirati pojedinačne fenomene i razine teksta, a nakon toga sustavno ispitati njihovu međuzavisnost koristeći pritom mogućnosti i dosege računalne lingvistike. Na taj se način mehanizmi ostvarivanja koherencije teksta mogu kvantitativno i kvalitativno ispitati temeljitije nego što je do sada bio slučaj.

Knjiga posjeduje indeks pojmova, obilnu bibliografiju na jedanaest stranica te prilog s kratkim pojašnjenjem o XML-u i lingvističkim bazama podataka. Na kraju prvih devet poglavlja nalaze se praktične vježbe za primjenu obrađenih teorijskih pristupa i postavki uz korištenje opisanih alata za anotiranje. Sva poglavlja osim prvog i desetog završavaju prijedlozima za produbljivanje znanja iz obrađenih područja uz kratak opis i komentar što se u kojoj knjizi u kojem poglavlju može pronaći. Autor često upućuje i na mrežne stranice na kojima su dostupni alati za anotiranje ili već anotirani lingvistički korpusi.

Knjiga predstavlja izvrstan temelj za bavljenje tekstnom lingvistikom jer donosi sustavan pregled različitih razina teksta i teoretskih pristupa i postavki. Uz to, izrazito vrijedan doprinos knjige leži u smjernicama za korištenje računalne lingvistike u bavljenju autentičnim tekstovima. Iako autor smatra da računalna lingvistika može pomoći u pronalaženju odgovora za kojima tekstna lingvistika, bez obzira na sva dosadašnja postignuća i saznanja, još uvijek traži, on ne idealizira ulogu računalne lingvistike, već upozorava na određene probleme i ograničenja vezana za korištenje alata te izradu jednoznačnih i preciznih smjernica za anotiranje.

Lucia Miškulin Saletović