

Željka Brlobaš i Marijana Horvat
Institut za hrvatski jezik i jezikoslovlje
Strossmayerov trg 2, HR-10000 Zagreb

PRIMJENA KONKORDANCIJE U IZRADBI RJEČNIKA

Tekst je pokušaj da se pokaže na koji način konkordancije djela starije hrvatske književnosti do 17. stoljeća mogu poslužiti kao korpus povijesnom rječniku te rječniku suvremenoga hrvatskog jezika.

Osim kompletnoga popisa riječi nekoga djela, s pomoću konkordancije moguće je utvrditi i uporabu riječi u vremenskom slijedu, njihovu dijalekatnu i žanrovsku rasprostranjenost, promatrati ih na stilskoj razini te uočiti gramatičke osobine.

Izradba rječnika temelji se na odabiru reprezentativnoga korpusa. Taj je korpus polazište u izradbi svakoga rječnika, pa tako i povijesnoga rječnika i rječnika suvremenoga jezika. Ovdje se pod pojmom *rječnik* uglavnom razumijeva opći rječnik prirodnoga jezika, odnosno opći rječnik hrvatskoga jezika. Ne treba posebno naglašavati svu težinu puta od korpusa do rječnika jer je to poznato, a ishod je nakraju ipak ostvarljiv i svrhovit.

U današnjoj je informacijskoj znanosti s razvijenom računalnom tehnikom jedno od mogućih pomagala u izradbi rječnikâ i primjena konkordancija. Kako je konkordancija, nakon niza računalnih i programskih radnji, abecedni popis svih riječi zajedno s okolinama u kojima se pojavljuju u tekstu, ona već samim svojim popisom predstavlja radni odmak od odabranoga korpusa. Naziv *riječ* iz prethodne definicije potrebno je zamijeniti nazivom *pojavnica*, što se obično definira kao omeđen slijed slovno-brojčanih znakova u tekstu. *Različnica* je skupina grafijski identičnih pojedinačnih riječi koja se registrira samo prilikom prvoga pojavljivanja jer se sa svakim kasnijim pojavljivanjem smatra pojavnicom. Pod nazivom *lema* misli se na različnicu svedenu na osnovni oblik: za imenice je to nominativ jednine, a za glagole infinitiv, što znači da je *lematizacija* postupak svođenja riječi na osnovni, kanonski oblik. Lematizacija se provodi prema gramatičkim pravilima, a zatim se dobiveni sustavi testiraju na stvarnom korpusu. Time se svaka pojavnica lematizira s izravnim uvidom u okolinu u kojoj se pojavila u

korpusu. Lematizirane pojavnice mogu poslužiti kao ulazni podaci za konačnu obradu i oblikovanje frekvencijskoga rječnika, dok je sustav totalnih konkordancija abecedom složen red svih oblika riječi, dakle svih pojavnica uz dio konteksta i oznaku mjesta u tekstu koji je konkordiran.

Kako konkordancija nema bez korpusa, jer se one izrađuju na temelju određenih tekstova (književnih, povijesnih, pravnih i drugih djela), tako se u odabiru korpusa ne bi smjelo zaobići njihovo postojanje. U nizu *korpus—rječnik* konkordancija zauzima međumjesto: *korpus—konkordancija—rječnik*, preciznije *korpus — nelematizirana konkordancija — lematizirana konkordancija — rječnik* jer svojom strukturom jest odmak od korpusa, ali još uvijek nije rječnik. Put od različnice do natuknice još je vrlo dug.

Budući da se rječnik starijega razdoblja temelji samo na potvrdama, jasno je da bi se korpus trebao temeljiti na svim postojećim tekstovima. Prethodno učinjena računalna bi konkordancija svakoga teksta omogućila da niti jedna riječ ne bude izostavljena te da se vrlo brzo, zahvaljujući abecednom poretku, može pronaći, uspoređivati s drugim riječima i kategorizirati na najrazličitije načine.

Nećemo se zaustavljati na pitanju izbora korpusa; ipak smatramo da se postojeće konkordancije pri izboru ne smiju izostaviti.¹ Također nećemo ulaziti u principe pripreme konkordancija iz perspektive računalne tehnike i programske podrške za njihovo stvaranje. Govorit ćemo o nekim od mogućih načina primjene gotovih konkordancija u izradbi općega povijesnoga rječnika hrvatskoga jezika. Pri tome ćemo se koristiti odabranim, nelematiziranim konkordancijama djela starije hrvatske književnosti do 17. stoljeća: *Razvod istarski* iz 1275., *Odlomak Korčulanskoga lekcionara* i *Žića svetih otaca* iz 14. stoljeća te *Osman Ivana Gundulića* iz 17. stoljeća.

Sve različnice nisu uvijek i istoznačnice. Neke sadrže leksičke razlike, neke morfološke, a neke i jedne i druge. Većina riječi nema samo jedno značenje nego više njih. To znači da je kod takvih riječi asimetričan odnos izraza i sadržaja jer njihovom formativu nije pridružen samo jedan semem te u različitim kontekstima imaju različito značenje pa ih smatramo *višeznačnim* ili *polisemičnim* riječima. Polisemija je jedna od ključnih pojava semantike. *Polisemička se riječ kao i monosemička smatra jednom jedinicom rječnika. Unutrašnja heterogenost, višestrukost i različitost značenja ne uništavaju identitet riječi. Zajednički oblik i zajednički semantički elementi značenja osiguravaju jedinstvo polisemičke riječi kao lingvističkoga znaka.*² Polisemija omogućuje ekonomičnost jezika jer bi

¹ Ovdje mislimo na konkordancije Milana Mogušića i Željke Bujasa koje su izrađene u okviru projekta *Kompjutorska analiza tekstova stare hrvatske književnosti* Zavoda za lingvistiku Filozofskoga fakulteta u Zagrebu te na konkordancije koje je izradila Dragica Malić iz Instituta za hrvatski jezik i jezikoslovlje.

² Cit. J. Melvinger, *Leksikologija*, str. 27.

se u suprotnome rječnik širio unedogled. Konkordancija nam omogućava brz uvid u značenje svake riječi, zbog konteksta koji ih okružuje. Tako ćemo npr. vidjeti da se u *Žićima svetih otaca* pojavnica *bolizan* kao imenica ženskoga roda pojavljuje u tekstu jedanaest puta: tri puta u očekivanom značenju 'narušenost zdravlja' (*Nigdare telu njeje nemoć ali bolizan / nika ne pride, da vazda zdrava...*), sedam puta u značenju 'žalost, tuga' (...*om / zdihan'ji, ali ča čekaju! Bolizan brez pokoja, plač brez konca ji...*) i jedanput kao 'požuda, pohota' (... *i na / smih, do konca pričasmo bolizan i svršismo nečistoću.*). Ta je riječ potvrđena s jednim primjerom u Odlomku Korčulanskoga lekcionara, također u značenju 'narušenost zdravlja' (... *in'ju, Isukrste, koga nigdare bolizan pritisnet, / zlamenujući se.*) te s tri primjera u *Osmanu*, u kojemu je zabilježen samo ijekavski lik: *boljezan*, i to u prenesenom značenju: 'ljubavni bol, pohota, požuda' (... *da srce otvori. / Gori, ali plame krije / gorke trpeć vik boljezni; / ljubi, ali još ne smije / objaviti sve ljuvezni.*), dok se u značenju 'tuga, žalost' sedam puta pojavljuje leksem *bolest* (... *proz gomile turskijeh kosti / jaše paša i naprijeda / od bolesti i od žalosti /zavrć riječi plač mu ne da.*), tri puta s varijantom *boles* (... *koliko je privelika / boles ka mi srce jada ...*). Iz primjera dobivamo osnovne podatke za buduću natuknicu: vrsta riječi, denotativna i konotativna značenja, mogućnost konkretnih potvrda iz teksta, ispitivanje omjera i određivanje zakonitosti u uporabi dubleta — ijekavskih i ikavskih elemenata te praćenje pojavljivanja i razvoja riječi u vremenskom slijedu (što već imamo u Akademijinu *Rječniku* koji *boljezan* bilježi od 13. stoljeća, a *bolest* od 15. stoljeća).³

Za razliku od višeznačnih riječi kod kojih su istome formativu pridružena različita ali bliska značenja, *homonimi* značenjski najčešće nisu ni u kakvoj svezi. Većina tradicionalnih klasifikacija homonima razlikuje *pune homonime*, *homografe*, *homofone* i *homoforme*. Na primjeru homografa, riječi različitoga značenja čiji se formativi isto pišu, a drukčije izgovaraju, pokazat ćemo svrsishodnost konkordancije. Da bismo mogli odrediti ispravnu natuknicu i odgovarajuću vrstu riječi, moramo imati uvid u okružje u kojem se te riječi pojavljuju, osobito ako se radi o starim tekstovima. Pokazat ćemo to na primjeru pojavnice *gori*:

1. ... *tko bi gori, eto je doli ...*,
2. ... *i ne manjijem ognjem gori ...*,
3. ... *ptica u moru, riba u gori ...*,
4. ... *n'je, veći Boga zabiti, jošće gori desperati se ali Boga ne pomnit.*⁴

Usporedimo li kontekste u kojima se pojavljuje pojavnica *gori*, dolazimo do sljedećih podataka: u prvom se primjeru *gori* pojavljuje u funkciji priloga

³ Usp. *Rječnik hrvatskoga ili srpskoga jezika*, str. 540. za *boljezan* i 526. za *bolest*

⁴ Prva su tri primjera iz *Osmana*, a posljednji je iz *Žića svetih otaca*.

gòre, u drugom kao treće lice jednine prezenta glagola gòrjeti, u trećem je riječ o lokativu jednine imenice gòra. U četvrtom je primjeru gori uporabljen kao komparativ pridjeva lđš, zão, što se sa sigurnošću može tvrditi tek uvidom u širi kontekst, odnosno u tekst: *Velik grih jest nesaznan'je, veći Boga zabiti, jošće gori desperati se ali Boga ne pomnit*. Takvi su rezultati, dobiveni iz konkordancija, vrlo korisni jer mogu poslužiti pri obradbi četiriju rječničkih natuknica.

Pojavnica *ljut* iz konkordancije *Osmana* I. Gundulića primjer je kojim se želi pokazati stilistička uporaba navedenoga pridjeva. Može se, dakako, postaviti pitanje koliko je jedno stilističko promatranje pojavnice iz konkordancije uopće korisno pri izradbi povijesnoga rječnika, ali se odgovori vrlo brzo pronalaze.

Pojavnice pridjeva *ljut* najprije će u rječniku omogućiti uspostavu navedene natuknice i njezinu gramatičku obradbu. Uvidom u konkordancije pratit će se vremenski slijed uporabe navedenoga pridjeva. Na temelju našega korpusa konkordancija ustanovljeno je da tekstovi *Razvoda istarskoga* (13. stoljeće) i *Odlomka Korčulanskoga lektionara* (14. stoljeće) ne sadrže taj primjer. *Žića svetih otaca* iz 14. stoljeća imaju jednu potvrdu pridjeva *ljut* u svezi s imenicom *konj* gdje pridjev ima preneseno značenje:

... on jest kakono konj *ljuti* želeći samicu najti ..

Najviše potvrda pridjeva *ljut* pronalazi se u konkordanciji Gundulićeva *Osmana* (17. stoljeće) pa se time vremenski slijed može povezati i s pripadnošću teksta određenoj književnoj vrsti, odnosno žanru. U tekstu manirističko-barokne poetike dobiva se i najveća stilistička raslojenost i uporaba odabranoga primjera. Pridjev *ljut* pojavljuje se u svezi s nekom imenicom 44 puta. Tim pojavnica pridjeva moguće je dalje odrediti imenice uz koje se pojavljuju, te im također utvrditi učestalost: pridjev *ljut* najčešće stoji uz imenicu *zmija* (5 puta), te uz imenice *boj* i *zmaj* (4 puta). Dvapat se pridjev pojavljuje uz imenice *jadi*, *sila*, *sablja*, *glad*, *strila* i *rana*. Učestalost pojavljivanja uz te imenice može se objasniti ulogom pridjeva *ljut* kao stalnoga epiteta. Svaka od ostale dvadeset i tri pojavnice pridjeva *ljut* odnosi se na jednu imenicu i time također pokazuje određeni stilistički izbor i svezu s određenom imenicom.

Na temelju navedenoga promatranja sveze *pridjev – imenica* moguće je pridjevu u rječničkoj natuknici definirati određena denotativna i konotativna značenja koja se opet mogu lakše odrediti ako u konkordanciji potražimo imenice koje se javljaju s navedenim pridjevom i usporedimo s kojim se još pridjevima pojavljuju te imenice.

Pomoću konkordancije brže se utvrđuju fonološke inačice, kao što su *kroz* – *proz* kojima obiluje *Osman* (*kroz planinu staru uzidi : proz gomile turskijeh kosti*) ili *brez* – *prez* – *prezi* iz *Žića svetih otaca* (...kada ostanu *brez vode na suhi umru...* : *Ki rasrdit brata svooga // prez kažuna?* : ...oko tvoje desno ne zamnet, *prezi / kažuna ne srdi se njemu*.). Razlike se u starim tekstovima očituju i po

refleksima poluglasa, npr. *a* u *pravadan* (Žića svetih otaca — ...ki brez griha su, *prava/dni*...; *Odlomak Korčulanskoga lektionara* — ...dostojno i *pravadno* jest...) nasuprot varijanti s *e*: *pravedan* (*Osman* — ...nastoj bit *pravedan*). Pored već spomenutih razlika u refleksima jata [npr. *bolizan* : *boljezan* ili *tijelo* : *tilo* : *telo* — ...ovo *tijelo* cječ *smiljenja*... : ...psi njegovo mrtvo *tilo*... (*Osman*) : *Nigdare telu njeje nemoć*... (Žića svetih otaca)] u starim je tekstovima vidljiva varijantnost suglasničkih skupova *čt*, *št* i *šč*. Kako su oni pod jakim utjecajem crkvenoslavske pravopisne norme, u mnogim se slučajevima ne bilježi promjena suglasničkoga skupa *čt* u *št*. To znači da pored primjera *ništa*, *nište* i *ništar(e)* postoji i *ničtare*, odnosno *nišće*, te *poštenje* uz *počten'je*, ili *praštati* uz *praščati* i slično. Ilustrirat ćemo to primjerima:

št : **čt** : **šč**

- a) **ništar(e)** : **ničtare** : **nište** : **nišće** : **ništa** — ... *a ništar* mu ne rabi : *Noj ništare* ne jimiše ni *udržaše* : ...*nijednim zakonom* ne ti *vzeti ničtare* : ...*v koj nište* jest *koristno*... : *Učenik* bo *nišće* mogaše *znati*... : *I zato ništa* ni *bolje nere mlčati* (Žića svetih otaca)
- b) **poštenje** : **počten'je** — ...*a poštenje* oni i *silu*... (*Osman*) : ...*samu muku prašća*, da i *počten'ju* učini *dostojna* (Žića svetih otaca)
- c) **praštati** : **praščati** — ...*ne prašta* se *vik krivina* (*Osman*) : ...*samu muku prašća* (Žića svetih otaca).

Kako će se u budućem rječniku obraditi takve natuknice, stvar je koncepcije rječnika, ali se to mora imati na umu.

Konkordancija, pogotovo odostražna, daje cjelovit pregled morfoloških oblika, pa tako brzo i jednostavno možemo doći do podatka potrebnog u obradbi natuknice, prema pravilima obradbe utvrđenim za svaku vrstu riječi posebno.

Na kraju bismo zaključile da je primjena već izrađenih konkordancija od velike pomoći u leksikografskome poslu. Pitanje je samo kako što brže i efikasnije prirediti tekstove za konkordiranje, s obzirom na to da je pojedinačan unos tekstova u računalo dugotrajan proces, uglavnom bez mogućnosti skeniranja jer se radi o starom jeziku, starom pismu i staroj grafiji.

Literatura

- Bratanić, Maja. 1992. Korpusna lingvistika ili sretan susret. *Radovi Zavoda za slavensku filologiju* 27, str. 145–159.
- Halonja, Antun. 1996. *Strojna obrada korpusa i analiza problema višeznačnosti*, diplomski rad. Zagreb : Filozofski fakultet.
- Malić, Dragica. 1997. *Žića svetih otaca*, Hrvatska srednjovjekovna proza. Zagreb : Matica hrvatska — Institut za hrvatski jezik.
- Melvinger, Jasna. 1989. *Leksikologija*. Osijek : Pedagoški fakultet.

- Moguš, Milan. 1993. Piščev rječnik. *Rječnik i društvo*: zbornik radova sa znanstvenog skupa o leksikografiji i leksikologiji, str. 249–254.
- Neweklowsky, Gerhard. 1994. Od konkordancije do rječnika, *Filologija* 22–23, str. 355–366.
- Smith, George W. 1991. *Computers and human language*. New York : Oxford University Press.
- Rječnik hrvatskoga ili srpskoga jezika*, I. dio. Zagreb : JAZU, 1880–1882.

The use of concordances in dictionary compilation

Summary

The authors analyze selected unlemmatized concordances of early Croatian literary works (before the 17th century) – *Razvod istarski* from 1275, *A Fragment of the Korčula Lectionary* and *Žića svetih otaca* from the 14th century and *Osman* by Ivan Gundulić from the 17th century – and determine possible ways in which these concordances could be used in the compilation of a general historical dictionary of the Croatian language.

Ključne riječi: konkordancija, pojavnica, različnica, lema, povijesni rječnik
Key words: concordance, token, lemma, historical dictionary