

PRAVOPIS I RAČUNALO

Slaven Batnožić - Branko Ranilović - Josip Silić HRVATSKI RAČUNALNI PRAVOPIS

*Gramatičko-pravopisni računalni vodič (spelling checker)
(Matica hrvatska - SYS, Zagreb 1996)*

Pojam je *spelling-checkera* u svijetu informatičara i kompjutorske tehnologije već poodavno dobro znan, jasan i jednoznačan. Tako je u Websterovu rječniku kompjutorskih termina *spelling-checker* definiran kao kompjutorski program, obično priključen na programe za obradu teksta, koji uspoređuje napisane riječi s listom riječi te upozorava korisnika na moguće pogreške u pisanju.¹

Unatoč općepoznatosti i općeprihvaćenosti danas gotovo nezaobilazna pomagala u pisanju na kompjutoru korisnici se računala i računalnih programa na hrvatskome govornom području nisu dosada imali prilike služiti takvim dodatnim "uslugama" svojih više ili manje sofisticiranih strojeva za pisanje.

Odgovor na pitanje zašto je tome tako i prije no u našem neprijepornu kaskanju

¹ Usp. *Webster's New World Dictionary of Computer Terms* (Completely Revised & Updated), fourth edition, New York, 1993, str. 395: *A computer program, usually associated with word processing, that*

za informatički razvijenijim zemljama treba tražiti u strukturi hrvatskoga jezika. (Pritome dakako valja imati na umu način funkcioniranja računalnoga pravopisa: on riječi koje se javljaju u zadanome tekstu nastoji "prepoznati"; "prepoznat" će dakako samo one koje su prethodno određene kao ispravne, dakle samo one koje se nalaze na njegovu popisu prihvatljivih riječi.) "Jezici zemalja u kojima su stvoreni prvi programi za pravopisnu kontrolu uglavnom pripadaju germanskomu (anglosaksonskom) i romanskomu krugu. Pripadaju skupini analitičkih jezika. Njihova je karakteristika da promjenjive riječi nemaju mnogo tvorbenih oblika. Padeži (za imenice, pridjeve i zamjenice) ili vremena i načini (za glagole) tvore se u pravilu tako da se ispred riječi stavljaju prijedlozi ili pomoćni glagoli. Osnovna se riječ ne mijenja ili joj se dodaju kakvi (brojem vrlo ograničeni) nastavci. Takva jezična tvorba povoljna je za programsku kontrolu, jer broj morfoloških izvedenica ne povećava znatno broj osnovnih rječničkih jedinica. (...) Hrvatski pripada skupini sintetičkih jezika, koje karakterizira velik broj morfoloških inačica iste osnovne riječi. Ako npr. kao osnovnu riječ označimo *imenica u nominativu jednine*, tada ona ima još trinaest morfoloških inačica - ukupno po sedam za jedninu i množinu. Ili: ako je osnovna riječ *nominativ jednine pridjeva*

u muškom rodu, tada taj pridjev možemo razviti u najmanje 126 oblika za muški, srednji i ženski rod, za jedninu i množinu te za komparativ i superlativ, tamo gdje ih ima, i dodati još podatke za određeni i neodređeni oblik. Ili: ako je osnovna riječ *glagol u infinitivu*, tada možemo nanizati najmanje 29 različitih oblika po vremenu ili načinu, rodu i broju, ne računajući one koje tvorimo pomoćnim glagolima i glagolskim pridjevom."²

Čak i samo naslućujući kakve to poteškoće zadaju sintetički jezici, kakav je i hrvatski, računalnom pravopisu, veliko je zadovoljstvo što možemo ustvrditi: napokon postoji *spelling-checker* i za hrvatski jezik! **Josip Silić** te informatičari **Branko Ranilović** i **Slaven Batnožić** autori su *Hrvatskoga računalnog pravopisa (Gramatičko-pravopisnog računalnog vodiča)* koji se u knjizi i na disketi pojavio u izdanju **Matice hrvatske** i **SYS-a** (Zagreb 1996).

Najprije međutim valja naznačiti odnos pravopisa knjige (pa onda i aktualnoga hrvatskoga pravopisa: Babić-Finka-Moguševa *Hrvatskoga pravopisa*) s jedne strane te računalnoga pravopisa s druge. Da bi se taj odnos bolje razjasnio, važno je razlikovati pravopis kao normu i pravopis kao knjigu u kojoj je ta norma (*pr*) opisana. Tada se može reći: Gledano s aspekta norme *Hrvatski računalni*

compares typed words against a word list and informs the user of possible spelling mistakes. A sophisticated spelling checker can have a base dictionary of well over 100,000 words and can provide the user with the ability to create special-purpose dictionaries of words not included in the base dictionary. Kako će biti kasnije pokazano, hrvatski računalni pravopis o kojem je ovdje riječ ima i tu posebnu, dodatnu mogućnost.

2 Usp. Batnožić-Ranilović-Silić, *Hrvatski računalni pravopis (Gramatičko-pravopisni računalni vodič)*, Matica hrvatska, SYS, Zagreb 1996, str. 11 i 12.

pravopis ne donosi većih novosti. (Potvrđeno je time načelo konzervativnosti norme, ali i ugođeno korisničkoj komociji: neće trebati učiti novi pravopis!) Naprotiv, namijenjen drugom mediju (računalu), *Hrvatski je računalni pravopis* postojeću pravopisnu normu na sebi svojstven način "obradio"; ponajprije moralo se svakako odustati od načela "itd."! Pritom su se nužno aktualizirala pojedina pitanja hrvatskoga pravopisa, a morali su biti razriješeni i neki problemi.

Valja odmah jasno naglasiti da pravopis primjeren radu na računalu, pa odatle i kompjutorskom "mozgu", ne dotiče sve pravopisne probleme. Budući da mu je svrha provjera pravopisa (*spelling*) u tekstovima pisanim na računalu, temeljno pitanje koje razrješava (točnije: može razriješiti) jest ustanovljavanje, i potom prepoznavanje pravopisnog tipa: fonološko-morfonološkoga (fonemsko-morfemskoga). (To je dakako moguće samo ako je računalni pravopis "upoznat" sa što većim brojem leksičkih jedinica i njihovih oblika.) Jedinice veće od riječi (npr. rečenice, stoga pisanje rečeničnih znakova), ali i pisanje velikog početnog slova u nekim pozicijama (npr. u vlastitim imenima, jer da bi se utvrdilo što je vlastito ime, nisu dostatna jezična znanja i to se ne da svesti na "logiku" računala), pisanje (stranih) vlastitih imena ostaju nedostupni strojnom načinu "razmišljanja", pa onda i neprovjerljivi u klasičnom postupku računalne pravopisne provjere. Računalo će međutim, uputi li ga se u to, vrlo dobro poslužiti i za još poneku "tehničku" provjeru. Od koristi će tako biti pri rastavljanju riječi na slogove. Bilo je stoga potrebno utvrditi jasna i nedvosmislena opća pravila rastavljanja hrvatskih riječi

na slogove - a utvrđeno je i u knjizi navedeno njih osam - koja su potom upotrijebljena za izradu programskog algoritma. (Računalu se, nakon što je hrvatski odabran za glavni jezik, može zadati da riječi rastavlja automatski ili se to po potrebi može činiti ručno.) Općoj će urednosti na računalu pisanih tekstova pridonijeti i činjenica da će *spelling-checker* kao na nepoznate reagirati i na one riječi iza kojih dolazi koji rečenični znak koji nije napisan na odgovarajući način (iza njega ne slijedi prazno mjesto, bjelina), npr. *razgovarati, pa*. (U knjizi su popisana pravila kada se ostavlja, a kada ne ostavlja razmak u tekstu i u brojevima.)

Da bi računalni pravopis hrvatskoga jezika mogao biti upotrebljiv, trebalo ga je učiniti "pametnim". Nije mu pritom bilo dostatno ponuditi bilo kakav, ma kako velik popis riječi hrvatskoga jezika, već je iznimno iscrpnu listu riječi trebalo višestruko umnažati upisivanjem svih oblika promjenljivih riječi. Računalni se pravopis po tome razlikuje od bilo kojeg "papirnatog" pravopisa, pravopisa knjige, koji u svome rječniku u načelu donosi samo one riječi i samo one njihove oblike koji su zanimljivi s pravopisnoga stajališta. (Svaki korisnik *spelling-checkera* može stvarati i osobni rječnik u koji će dodavati riječi i izraze - češća osobna imena, toponime, termine vlastite struke - koji će u sljedećim pravopisnim provjerama biti prepoznati kao ispravni.)

Ni jedan postojeći hrvatski rječnik, ni gramatika, nisu imali na taj način obrađenu hrvatsku leksičku građu, pa se i nisu mogli upotrijebiti kao osnova za izradu računalnoga pravopisa. Trebalo je zato stvoriti novi opis hrvatskoga jezika koji bi se temeljio na hrvatskom jeziku kao

sustavu, ali se pritom nije smjela zanemariti situacija u jezičnome standardu (i to u svim njegovim funkcionalnim stilovima) i u konkretnim komunikacijskim situacijama.

Tek za ilustraciju - jer i nije lako zamisliti kolik je to posao - može se navesti da je uneseno preko 70.000 natuknica, osnovnih riječi, a kad su pridodani i svi morfološki oblici promjenljivih, dobiveno je više od 2.000.000 oblika riječi (u klasičnoj će pravopisnoj provjeri svi oni biti "prepoznati" kao ispravno napisane riječi hrvatskoga jezika). Morfološki je opis nužno uključio i morfonološke promjene tipa *iskati* : *iščem* : *ištem*, ali se nije smjela zanemariti komunikacijska realnost (npr. *iskan*, *zamašćen*, *iskorišten*). Pritom je utvrđeno više od 700 morfoloških modela.³ Nužno je bilo voditi računa i o varijantama: ako je u komunikacijskoj situaciji uglavnom nevažno rabimo li oblike *kojeg*, *kojega* (gen. jd.), *hrvatskom*, *hrvatskome*, *hrvatskomu* (dat. jd.), *nokti*, *noktiju*, *nokata* (gen. mn.) i sl., računalu oni i nisu istovrijedni i poima ih kao različite riječi. Zato da bi ih sve "propustio" kao ispravne, računalni program sa svima mora biti "upoznat".

Stvoren je tako dosad nedvojbeno najbogatiji gramatički tezaurus, koji je i kao zaseban dio priključen *Hrvatskome računalnom pravopisu* (knjizi tek dijelom,

au cijelosti disketi): moguće je jednostavno provjeriti sve morfološke oblike promjenljivih riječi te ustanoviti kojem morfološkom modelu pripadaju. *Gramatički tezaurus* zadobiva i sasvim praktičnu namjenu jer zacijelo predstavlja najpotpuniji morfološki rječnik hrvatskoga jezika. (U prilog mu ide i spretnost i jednostavnost njegove uporabe: dovoljno je upisati bilo koju riječ - i to ne mora nužno biti imenica u nominativu jednine, a glagol u infinitivu - i ako je to jedna od 70.000 unesenih osnovnih riječi ili koja od preko 2.000.000 njihovih morfoloških jedinica, začas se mogu dobiti svi podaci, od vrste riječi do svih paradigmatskih oblika.)

Napokon s gledišta se funkcioniranja računalnoga pravopisa kao poseban problem nameću neka pravopisna pravila. U prvom su redu to ona koja pravopisnu normu opterećuju (nepotrebnim) dvojnostima tipa *zadaci/zadatci* ili *bregovi/brjegovi*.⁴ Računalni pravopis, ostajući vjeran logici "razmišljanja" stroja, donosi alternativna rješenja: *zadaci*, *zadatci*, *izvaci*, *izvadci*, *napuci*, *naputci*; *bregovi*, *brjegovi*, *crepić*, *crjepić*, *sprečavati*, *sprječavati*, *vredniji*, *vrjedniji*, ali ide i korak dalje. "Oglušujući" se na nedosljednosti hrvatske pravopisne norme i pojedina kompromisna rješenja (npr. *letci* prema *letak*, ali *oci* prema *otac*; *brježuljak*, *brjegovit*, ali *vremena*, *napredan*, *neposredan*, *potreba*), on kao

³ Broj je modela gotovo nepredvidivo rastao neovisno o poznatim "klasičnim" gramatičkim modelima. Naime ako prema *iskorištavati* dolazi *iskorišten*, a prema *zamastiti* *zamašćen*, jasno je da glagoli *iskorištavati* i *zamastiti* ne mogu biti pridruženi istome modelu. Nadalje posebni su modeli morali, primjerice, biti uvedeni za poznatija vlastita imena (uglavnom gradova i zemalja) koja u pravilu imaju isključivo jedninu ili množinu (npr. *Vinkovci*); neke su riječi tipa *doba* uključene i u više od jednog modela (i kao sklonjive i kao nesklonjive), itd.

⁴ Usp. Babić-Finka-Moguš, *Hrvatski pravopis*, Zagreb 1994, str. 56-57, 45.

ispravne "propušta" i alternativne oblike *otče, otci, otcima; sudče, sudci, sucima* te *vrijemena* i sl. Onome je koji piše prepušteno dakle da odabere na koji će način biti pismen. Računalni ga pravopis neće onemogućavati da u tome bude dosljedan i logičan.

U knjizi su pridodanoj pravopisu na disketi (doduše u našoj bi se "predinformatičkoj" svijesti prije moglo govoriti o disketi pridodanoj knjizi) prikazana i ona pravopisna pitanja koja i nisu baš pogodna za računalnu provjeru: pisanje rečeničnih znakova, velikoga početnog slova i stranih osobnih imena. Osnove gramatičke norme (poglavlja *Od glasa do teksta, Vrste riječi*, ali i izvaci *Iz gramatičkoga tezaurusu*) mogu biti od koristi u razumijevanju jezične strukture, praktična pomoć u odčitavanju oznaka u *Gramatičkome tezaurusu*, ali i opravdavaju podnaslov knjige: *Gramatičko-pravopisni računalni vodič*. Nadalje podizanju bi razine opće pismenosti svakako mogli pripomoći savjeti za pisanje poslovnoga pisma (cjelina *Poslovno pismo*) s iscrpnim uputama o formi poslovnog pisma,

njegovu sadržaju, načinu oslovljavanja i obraćanja i sl.

Evo kako je u knjizi *Hrvatskoga računalnog pravopisa* riješeno pitanje pisanja rečeničnih znakova, velikog početnog slova i stranih osobnih imena.

Ako je u provedbi pravopisnih pravila kojima je određen fonološko-morfonološki tip hrvatske interpunkcije zamijećena iznimna dosljednost, ona neće iznenaditi ni u provedbi odabranog, (logičko-)semantičkog interpunkcijskog načela.⁵ Tako propisujući "čistu" (logičko) semantičku interpunkciju, knjiga donosi iscrpna i jasna objašnjenja uz koje se veznike i kada zarez piše. Veznici su podijeljeni po vrstama i s obzirom na tipove rečenica u kojima se javljaju. Kriteriji su o kojima ovisi (ne)pisanje zarez, dakako, semantički.⁶ Samom se organizacijom pravila unekoliko odstupilo od tradicije pisanja pravopisnih knjiga (polazi se dakle od veznika i određuje se koja značenja oni mogu zadobivati u iskazu i tekstu te se potom utvrđuje piše li se ili ne uz njih rečenični znak), ali se nedvojbeno dobilo na preglednosti i jednoznačnosti.

5 O interpunkcijskim načelima, *strukturnom, (logičko-)semantičkom i ritmo-melodijskom* usp. npr. L. Badurina, *O interpunkcijskim načelima*, Fluminensia, god. 8, br. 1-2, Rijeka 1996, str. 257-264.

6 Može se to potkrijepiti primjerice pravilima o pisanju zarezu u sastavnim rečenicama: veznik i "a) ne odvaja se zarezom kad sastavlja istodobne sadržaje dijelova složene rečenice koje vezuje (...); b) odvaja se zarezom kad sadržaj druge rečenice pridružuje sadržaju prve rečenice kao njegovu posljednicu"; veznik **pa** "najčešće se ponaša kao i pod b), pa se obično odvaja zarezom"; veznik **te** "a) ne odvaja se zarezom kad se ponaša kao i pod a) (tada obično zamjenjuje i pri nizanju sadržaja od kojih se sadržaj koji pridružuje donekle razlikuje od sadržaja koji vezuje veznik i) (...); b) odvaja se zarezom kad se ponaša kao veznik i pod b)"; veznici **ni** i **ni** - **ni** te **niti** - **niti** "općenito se ne odvajaju zarezom" (str. 47-48). Primjeri bi se i dalje mogli nizati, ali bit će spomenuto još samo jedno pravilo: "Dijelovi se jednostavne rečenice (subjekt, predikat, objekt i priložna oznaka) ne odvajaju jedni od drugih zarezom" (str. 62) - izravno se to suprotstavlja dosadašnjoj normi koja je, čini mi se, ustuk ritmo-melodijskom načelu, onome koji piše prepuštena da odluči hoće li ili neće pisati rečenični znak i uzdaval se pritom u njegov osjećaj za dobru mjeru (usp. npr. Babić-Finka-Moguš, *Hrvatski pravopis*, Zagreb 1994, str. 113).

Ni u pisanju velikoga početnog slova nije iznevjerena tradicija hrvatskoga pravopisa; tradicija je prije napuštena u formulacijama pravopisnih pravila i organizaciji cijeloga pravopisnog poglavlja. Pravila koja su se inače protezala na povećem broj stranica sada su sažeta na svega tri stranice (str. 82-84)! Pisanje se velikoga početnog slovo na početku rečenice posebno ne razmatra (pitanje je to automatizma u pisanju: nakon rečeničnog znaka kojim se označava rečenični kraj piše se veliko slovo, usp. i str. 25); veliko početno slovo u riječima kojima se iskazuje poštovanje svedeno je na četiri pravila, pa su tako u ukupno trideset i jednome pravilu predviđene situacije u kojima se piše veliko početno slovo.

Što se tiče pisanja stranih osobnih imena, *Hrvatski je računalni pravopis* izričito: "Strana osobna imena iz jezika koji se služe latinicom pišu se onako kako se pišu u jezicima iz kojih dolaze (...) Strana se pak imena koja dolaze iz jezika koji se ne služe latinicom pišu prema odgovarajućim transkripcijsko-transliteracijskim pravilima" (str. 85). To je, uostalom, u skladu s hrvatskom pravopisnom tradicijom. Deklinacija je i tvorba pridjeva od njih i gramatički i pravopisni problem. Imena su svrstana u hrvatske sklonidbene tipove s obzirom na svoje završetke; hrvatski se gramatički nastavci aglutiniraju na stranu osnovu (još jedan "izuzetak" suvremene pravopisne norme nije ignoriran, ali ga se nastojalo ublažiti - gen. *D'Annunzia*, a ne samo, kako je propisano Babić-Finka-Moguševim *Hrvatskim pravopisom*, *D'Annunzija*); odnosni se pridjevi od stranih imena naseljenih mjesta, kao i oni

od stranih osobnih imena, tvore isključivo s transkribiranom imenskom osnovom (dakle samo *njujorški*, *kembrički*, *lionski*, kao, uostalom, i *geteovski*, *leopardijevski*, *šekspirijanski* i sl.).

Ipak bi se na kraju trebalo vratiti osnovnoj namjeni *Hrvatskoga računalnog pravopisa*: dragocjenoj pomoći koju on pruža u pisanju na računalu (namijenjen je radu u programu MS Word 6x i MS Word 7x). Obilje je podataka u zapisu na disketi organizirano tako da ne zauzima mnogo prostora na disku osobnog računala (PC 386 ili 486, minimum), a da ujedno bude osigurana potrebna brzina pretraživanja. Doista sada nije ništa lakše no pravopisno provjeriti tekst ili u *Gramatičkom tezaurusu* pronaći koju riječ ili neki njezin oblik. Napokon višegodišnji je naporan posao Josipa Silića i Branka Ranilovića te umijeće Slavena Batnožića da se, primjenom posebnih algoritama, golema građa sažme i time omogućiti potrebna brzina pretraživanja nagrađen iznimnim priznanjem: hrvatski je *spelling-checker* udovoljio visokim standardima što ih je postavio *Microsoft*, vodeća svjetska softverska tvrtka. Hrvatski se jezik time našao na popisu svjetskih jezika u *Microsoftovim* programima, i to između katalonskog i danskog! O praktičnim pak prednostima uporabe sofisticiranog programa moći će svjedočiti njegovi korisnici. Na nama je da se nadamo da će se podići razina hrvatske (javne) pismenosti; iako *Računalni pravopis* nepismenoga neće opismeniti, pismenima će i polupismenima omogućiti da budu (još) pismeniji.

Lada Badurina