

Ranko MATASOVIĆ

Filozofski fakultet, Odsjek za opću lingvistiku

Ivana Lučića 3, HR-10000 Zagreb

TEMELJNI LEKSIK I DUBINSKA GENETSKA SRODNOST¹

U ovom radu raspravlja se o jednoj novoj metodi istraživanja moguće dubinske srodnosti jezika, na primjeru tzv. "nostratičkih jezika", a osobito indoeuropskih, uralskih i kartvelskih jezika. Statistička analiza glasovnih podudarnosti suglasnika u rekonstruiranom "temeljnem leksiku" indoeuropskoga, uralskoga i kartvelskog prajezika pokazuje da se distribucija vjerojatnosti tih podudarnosti ne razlikuje bitno od slučajne distribucije.

§1 Jednom sam razgovarao s profesorom Andrewom Sihlerom o problemima genetske klasifikacije američkih indijanskih jezika. Dok smo se lako složili da je na razinama porodica poput algonkijske, vakašanske ili sališke ta klasifikacija razmjerno nesporna, zaključili smo da na višim razinama vlada priličan kaos i neslaganje čak i među kompetentnim autorima. Ja sam na to dodao da su argumenti na kojima se zasnivaju neke genetske natporodice ili file, koje se navode i u standardnim priručnicima, poput penutijskih jezika (da ne govorimo o Greenbergovu "Amerindijskom") slabiji od onih na kojima je Illič-Svityč zasnovao nostratičku natporodicu jezika. Na to me je Sihler, pretpostavljajući valjda da sam kao lingvist iz ovoga dijela svijeta dobro upućen u rusku lingvističku literaturu, upitao kakav je moj stav o nostratičkoj teoriji, koja ovih dana postaje sve popularnijom i u Americi², no ja mu na to pitanje nisam znao odgovoriti. Premda izuzetno cijenim Illiča-Svityča kao indoeuropeista, zbog njegovih radova o baltoslavenskoj akcentologiji, o njegovu »Opytu sravnenija nostratičeskix jazykov« nisam mogao reći niti da me je uvjerio u ispravnost nostratičke teorije, niti da me nije uvjerio. Količina rekonstruiranih nostratičkih korijena u toj knjizi (i njenim

¹ Zahvalan sam Dorjani Široli na korisnim primjedbama na prvu verziju ovoga članka.

² V. npr. A. Manaster Ramer (1993), zbornik radova o nostratici J. Salmons i B. Josepha (1998), te pomalo "heretičku" verziju nostratičke teorije A. Bomharda (1994). Za kritiku nostratičke teorije v. osobito Serebrennikov 1981, Vine 1991.

nastavcima) uistinu je impresivna, a postupak kojim se do njih dolazi barem na prvi pogled zasniva se na pravilnim glasovnim podudarnostima. Međutim, pažljiviji pogled otkrit će da su mnoge rekonstrukcije dobivene usporedbom korijena u “nostratičkim” jezicima čija se značenja samo približno podudaraju³, da je fonološki sustav pranostratičkoga tipološki nemoguć i prekomplikiran, te je stoga moguće sumnjati da je rekonstruiran na osnovi *ad hoc* glasovnih podudarnosti, te da je indoeuropska građa kojom se Illič-Svityč koristio u velikom broju slučajeva nepouzdana ili netočna. Valja se prisjetiti metodološkog načela koje je formulirao Ives Goddard: »A lexical similarity between only two languages is generally considered insufficiently supported, unless the match is very exact both phonologically and semantically, and it is rare that a match of only one or two phonemes is persuasive. If the meanings of the forms compared differ, then there must be an explicit hypothesis about how the meaning has changed in the various cases. Now, if these strict criteria have been found necessary for etymologies within known linguistic families, it is obvious that much stricter criteria must be applied to word-comparisons between languages whose relationship is in question. In fact it is virtually impossible to prove a distant genetic relationship on the basis of lexical comparisons alone.« (Goddard 1975: 261). No i pored svega toga, čitalac Illič-Svityčeva nedovršenog djela teško se može oteti dojmu da nije moguće da su *sve* njegove rekonstrukcije naprosto pogrešne ili rezultat slučajnosti. Budući da se ne volim naći u situaciji u kojoj ne znam ocijeniti valja li čemu vjerovati ili ne, stao sam razmišljati na koji bi se način nostratička teorija mogla neovisno provjeriti, *sine ira et studio*.

§2 Ne radi se o tome da se pitamo da li su nostratički jezici srodni, jer ne možemo isključiti mogućnost da su *svi* jezici svijeta srodni, niti o tome da pokušamo opovrći da su nostratički jezici srodni, jer u strogom smislu nije dokazivo da neki jezici *nisu* srodni. Postavljati pitanje o uvjerljivosti nostratičke teorije znači pitati se o tome kako treba dokazivati dubinsku jezičnu srodnost među jezicima. Radi se o tome što smatramo dokazom srodnosti:

— sistemske glasovne podudarnosti među jezicima nisu dovoljne⁴, jer one postoje i između jezika koji su bili u intenzivnim međusobnim kontaktima.

³ Što reći o podudarnostima između korijena koji znači ‘slušati’ (ie. *k’lew-) i riječi koja znači ‘uho’ (altajski *k’ul-), ili između kartvelskoga *px(w)- ‘topao’ i ie. *peHw- ‘vatra’, kakve često navode Illič-Svityč i drugi nostratičari? Kada bismo znali da se radi o srodnim jezicima, vjerojatno bismo prihvatili takve etimološke jednadžbe, no na “nostratičkoj” razini prihvaćanje takvih razlika u značenju znatno povećava mogućnost slučajnoga podudaranja.

⁴ Meillet je u svojem klasičnom djelu o poredbenoj metodi (Meillet 1967) zastupao stajalište da su za dokaz genetske podudarnosti dovoljne rekurentne glasovne podudarnosti. Takvo se shvaćanje pronalazi i u nekim suvremenijim priručnicima (npr. Lyovin 1997). O postojanju takvih podudarnosti i između posuđenica v. npr. Holzer 1995.

Na primjer, za onaj dio engleskoga leksika koji je posuđen iz pranordijskoga moguće je uspostaviti glasovne podudarnosti s odgovarajućim riječima u drugim germanskim jezicima kao i za onaj dio leksika koji je naslijeđen iz pragermanskoga. Na sličan način, postoje pravilne glasovne podudarnosti između hrvatskih riječi s početnim *b-* koje su posuđene iz turskoga (npr. *barjak*, *baždar*, *bubreg*, *biser*, itd.) i odgovarajućih riječi u turskome i drugim turkijskim jezicima.

— morfološke podudarnosti same za sebe nisu dovoljne jer mogu biti slučajne, a osim toga, gramatički morfemi češće se posuđuju iz jezika u jezik nego što bi se to moglo zaključiti na osnovi poznavanja samo indoeuropskih jezika (v. Dixon 1981). Primjer slučajnih morfoloških podudarnosti između američkog indijanskog jezika miwok i latinskoga navodi Sihler (Sihler 1998: 138):

	miwok (California)	latinski
1. sg.	-m	-m
2. sg.	-š	-s
3. sg.	-Ø	-t
1. pl.	-maš	-mus
2. pl.	-toš	-tis
3. pl.	-p(u)-	-nt

U mojoj latinskoj gramatici (Matasović 1997: §15, §96) zastupana je teza da je potrebno da uz sustav *glasovnih* podudarnosti postoje i *funkcionalne* podudarnosti, koje se očituju kao podudarnosti morfoloških paradigmi. Premda nijedan od ta dva kriterija uzet za sebe nije dovoljan da u potpunosti isključi posuđivanje i slučaj, zajedno oni zacijelo pružaju dovoljnu sigurnost i najskeptičnijem istraživaču. Ja i dalje stojim kod toga da je to dobra definicija dokaza genetske srodnosti jezika na razini jezične porodice poput indoeuropskih jezika, no tako očito nije moguće dokazati *dubinsku genetsku srodnost*. Jednostavno rečeno, na razini fila i natporodica moraju vrijediti manje strogi kriteriji dokaza, ili pitanje genetske srodnosti na tim razinama valja odbaciti kao neznanstveno.

§3 Možemo zamisliti da će dubinski genetski srodni jezici biti srodni na način kao što su srodni, recimo, albanski, latvijski i novoirski (ne možete zamisliti tri različitija jezika). Treba nam, dakle, metoda pomoću koje će se takvi jezici pokazati vjerojatno srodnima, a sigurno (ili gotovo sigurno) nesrodni jezici poput, recimo, indonezijskoga, svahilija i baskijskoga neće proći na ispitu.

Takvu metodu možda pruža statistička analiza glasovnih podudarnosti temeljnoga leksika (*basic vocabulary*) koja se zasniva na postupku istraživanja slučajnih glasovnih podudarnosti među jezicima koji je predložio Ringe (1992). Ako se znaju frekvencije pojedinih glasova u rekonstruiranom temeljnom rječniku (npr. Swadeshova lista od 100 riječi), moguće je ustanoviti vjerojatnost da

se riječi istoga značenja slučajno podudaraju u početnom suglasniku, ili bilo u kojem suglasniku koje te riječi sadržavaju⁵ (i tako bilo za koji par suglasnika). Ako se zna ta vjerojatnost, lako je izračunati koliko primjera podudaranja za dva jezika mora biti da to (s vjerojatnošću većom od 90 %) *ne bude slučajno*. Pretpostavka je da će među jezicima za koje znamo da su srodni takvih podudarnosti biti dovoljno da s vjerojatnosti većom od 90% zaključimo da nisu slučajne, a za jezike koji nisu srodni (osim na razini ljudske vrste) da će ih biti manje. Kažu da je Voltaire jednom rekao da je etimologija znanost u kojoj konsonanti vrijede malo, a vokali ništa. Mi ćemo se držati te njegove maksime, budući da su suglasnici dijakronijski postojaniji od samoglasnika, te je glasovne podudarnosti suglasnika moguće otkriti i onda kad su podudarnosti samoglasnika davno nestale.

§4 Evo u čemu se sastoji postupak izračunavanja vjerojatnosti slučajnog podudaranja suglasnika u riječima istoga značenja: prvo za svaki jezik izračunamo vjerojatnost P_{xn} da bilo koja njegova riječ sadržava suglasnik X_n . Ta se vjerojatnost dobije tako da se broj pojavljivanja X_n u svim riječima u listi $N(X_n)$ podijeli s brojem riječi u listi (N_r) i pomnoži s brojem suglasnika u prosječnoj riječi u listi (D), pri čemu je $D = N_s$ (broj svih suglasnika u svim riječima u listi) / N_r . Dakle, $P_{xn} = (N(X_n)/N_r) \times D$. Vjerojatnost P_{xn} izračunamo za sve suglasnike oba jezika koje uspoređujemo (L_1 i L_2), a zatim računamo vjerojatnost P_{xy} da će se u riječi istoga značenja u jeziku L_1 pojaviti suglasnik X_n , a u L_2 suglasnik Y_n . Ako su vjerojatnosti P_{xn} i P_{yn} poznate (prema gornjoj formuli), za svaki par suglasnika u L_1 i L_2 izračunamo vjerojatnost P_{xy} koja je jednaka $P_{xn} \times P_{yn}$. Za svaki par suglasnika iz L_1 i L_2 možemo sada očitati kolika je vjerojatnost da se pojavljuju u riječima istoga značenja, odnosno koliko podudarnosti očekujemo će se *slučajno* pojavljivati u riječima istoga značenja. Međutim, ako ustanovimo da u listi od sto riječi jezika L_1 i L_2 očekujemo na osnovi teorije vjerojatnosti da četiri riječi sadržavaju suglasnik X (u L_1) i Y (u L_2), to ne znači da će u nesrodnim jezicima *doista uvijek* četiri riječi sadržavati X i Y . To samo znači da se *vjerojatnost* da će tako biti raspodjeljuje po binomnoj krivulji oko vrijednosti četiri. *Moguće* je da će ih biti i više, ali i manje, no najvjerojatniji je upravo slučaj u kojem će biti upravo četiri takve riječi. Pomoću jednostavnih statističkih formula sada možemo izračunati koliki mora biti broj takvih riječi da mogućnost da sve one *slučajno* sadržavaju suglasnike X i Y bude zanemarivo mala, naprimjer, manja od 10%. Ukoliko za nekoliko parova suglasnika možemo pokazati da je

⁵ Kada bismo promatrali samo početne suglasnike, kao što je činio Ringe, ne bismo mogli pokazati postojanje statistički relevantnih, pravilnih glasovnih podudarnosti između dvaju jezika od kojih je jedan razvio sustavnu prefiksaciju, a drugi nije. U tom slučaju glasovno podudarni početni suglasnici korijena u tim jezicima ne bi bili otkriveni.

vjerojatnost njihova slučajnog podudaranja u riječima istoga značenja u L_1 i L_2 manja od 10%, to znači da u temeljnom leksiku jezika L_1 i L_2 postoje sustavne glasovne podudarnosti, tj. da su L_1 i L_2 (dubinski) genetski srodni. Formula za binomnu raspodjelu dobro je poznata u statistici (Vranić 1970): $b(x, n, p) = \binom{n}{x} p^x q^{n-x}$. U toj ćemo formuli zamijeniti n sa brojem riječi koje promatramo (npr. $n = 100$ za Swadeshovu listu od 100 riječi), p će nam biti vjerojatnost podudaranja riječi istoga značenja u nekom suglasniku (tu vrijednost moramo izračunati za svaki par suglasnika koji želimo ispitati), dakle $p = P_{xy}$, dok će x biti broj činjenično opaženih riječi koje se podudaraju. Izračunavajući $b(x, n, p)$ za sve činjenično opažene slučajeve podudaranja, lako ćemo utvrditi ima li među njima takvih za koje je $b(x, n, p) < 10\%$, tj. podudarnosti čija je vjerojatnost slučajnoga pojavljivanja manja od 10%. Samo ćemo takve podudarnosti priznavati kao neslučajne, a postojanje većeg broja takvih glasovnih podudarnosti smatrat ćemo dokazom dubinske genetske srodnosti jezika.

Važno je istaći da ćemo uspoređivati *riječi*, a ne korijene, kao što najčešće čine nostratičari. Korijeni su lingvistička apstrakcija, a značenja koja im se pripisuju (npr. u rječniku Pokornoga) često su preširoka, pa uspoređujući korijene ne možemo unaprijed odabrati točno određen popis jedinica koje ćemo uspoređivati. Kada bismo uspoređivali korijene na nostratičkim jezicima, ne bismo imali jasnoga kriterija pomoću kojega bismo utvrdili da li je opravdano usporediti korijen koji znači 'velik' s korijenom, na drugom prajeziku, kojemu je približno značenje 'jak'. Ili korijen koji znači 'velik' i korijen koji znači 'biti zdrav'? Ili 'jak' i 'brz'? Premda su zamislive semantičke promjene kojima bi se mogla objasniti etimološka srodnost korijena tako različitih značenja, moramo odustati od njihove usporedbe u prajezicima čiju genetsku srodnost tek treba dokazati, ukoliko u našoj analizi želimo isključiti slučajne glasovne podudarnosti. Kada jednom otklonimo sumnju u genetsku srodnost istraživanih jezika, riječi koje se po značenju ne podudaraju potpuno predstavljat će zahvalnu građu za etimološka istraživanja.

§5 Za riječi koje ulaze u analizu tražimo:

a) da ulaze u Swadeshovu listu 100 riječi koje se najrjeđe posuđuju; valja odmah istaći da nikakvim semantičkim ni funkcionalnim kriterijem nije moguće odrediti klasu riječi koje se nikad ne posuđuju: u lingvističkoj literaturi zabilježeni su čak i slučajevi posuđivanja osobnih zamjenica⁶.

⁶ Sumo (čibčanski jezik Južne Amerike) je iz jezika miskito posudio zamjenicu za 1. l. sg. Jezik Mednyj Aleut je posudio zamjenice iz ruskoga: *ungučī-ju, unguči-iš, unguči-it, unguči-im, unguči-iti, unguči-jat* umj. rus. *ja sižu, ty sidiš, on sidit, my sidim, vy sidite, oni sidjat*. (Campbell 1997: 245). Na sličan način, do posuđivanja zamjenica može doći zbog tabua. Takav slučaj opisuje Dixon (1981). U jeziku Western Desert oko 1950 umro je neki Nganunya i zato je zabranjena zamjenica *ngayu* 'ja'. Zamijenila ju je zamjenica *nganku*, ali onda je umro netko s imenom sličnim toj zamjenici, pa se u upotrebu vratilo *ngayu*, ali i posuđenica iz engleskoga, *mi* (29).

- b) da budu riječi koje imaju (gotovo) svi jezici (dakle kulturno neovisne riječi);
- c) da se mogu rekonstruirati za sve uspoređivane prajezike;
- d) da je njihovo značenje postojano, tj. da se rijetko mijenja u vremenu.

Iz temeljnoga rječnika izbacujemo riječi poput ‘mama’, ‘tata’, ‘ja’, ‘ti’, jer znamo da se one podudaraju i u jezicima koji gotovo sigurno nisu genetski srodni (osim na razini vrste *Homo Sapiens*). Iz toga da se u gruzijskom ‘mati’ kaže *deda*, a ‘otac’ *mama*, dok u kolokvijalnom engleskom te riječi glase *mom* i *dad* ne može se ništa zaključiti o genetskoj srodnosti tih jezika. Nešto veći problem predstavljaju osobne zamjenice 1. i 2. lica, za koje također vrijedi da u velikom broju jezika svijeta sadržavaju bilabijalne ili dentalne nazale⁷. Premda objašnjenje te činjenice nije poznato, ne možemo je iskoristiti za zaključivanje o genetskom srodstvu jezika.

Za “nostratičke” jezike primjere ćemo uzimati samo iz provjerenih priručnika, i pored toga, samo ako se barem dva priručnika slažu (osim za indoeuropski, gdje se smatram dovoljno kompetentnim prosuditi i predložiti svaku rekonstrukciju)⁸. Nažalost, takvi pouzdani priručnici ne postoje za sve grane “nostratičkoga”, već samo za dvije: uralsku i kartvelsku. Radi se o knjigama Sinor (1988), Rédei 1986-88 i Collinder (1977) za uralski, te Klimov (1964) i Fähnrich (1995) za kartvelski. Premda to za sada nije slučaj, opisanu metodu bilo bi moguće primijeniti i na druge grane “nostratičkoga”, kada bi njihovi prajezici bili podjednako uspješno rekonstruirani.

§6 Za sada je premalo riječi temeljnoga leksika koje se mogu rekonstruirati za sva tri testirana nostratička jezika. Pažljivom analizom izvora pronašao sam točno dvadeset takvih riječi, koje navodim u tablici A. Za usporedbu su u tablici B dane riječi istoga značenja na albanskom, latvijskom i irskom (genetski srodnim, ali vrlo izmijenjenim jezicima), a u tablici C na hrvatskome, svahiliju i indonezijskome (jezicima za koje nemamo nikakva razloga vjerovati da su srodni):

⁷ Usp. Nichols 1992, 1993. Nichols čak tvrdi da postoji globalna arealna distribucija jezika kojima zamjenica 1. l. jd. počinje s *m*- i onih kojima ista zamjenica počinje s *n*-. Prva skupina jezika navodno pokriva najveći dio Staroga Svijeta, a druga jezike obiju Amerika, Pacifika i pacifičkih obala Azije.

⁸ I među najpouzdanijim priručnicima često ima neslaganja; budući da razmatrana metoda zanemaruje samoglasnike, nisam smatrao relevantnim neslaganjima slučajeve u kojima jedan od priručnika za kartvelski ili uralski rekonstruira drukčiji vokalizam; primjerice, na naše rezultate ne utječe činjenica što Collinder uralsku riječ za ‘ime’ rekonstruira kao **nimi*, a Rédei kao **nime*, i sl. Također sam zanemarivao razlike koje se tiču samo notacije (npr. Sinor ima prauralsko **š* gdje Rédei ima **ç*), a u vrlo rijetkim slučajevima gdje korišteni izvori imaju rekonstrukcije koje se razlikuju u samo jednom suglasniku, morao sam prosuditi po svojoj (ograničenoj) kompetenciji.

A	engleski	indoeuropski	kartvelski	uralski
1.	blood	H ₂ esn-	zixsl-	weri (u.-f.)
2.	bone	H ₃ eH ₁ s-	dz ₁ dz ₁ wal-	lix ₁
3.	ear	H ₂ ews-	q'ur-	kəw ₁
4.	eye	H ₃ ek ^w -	twal-	šilmä
5.	fire	puH ₂ n-	dzec ₁ x-	tuli
6.	heart	k'rd-	gul-	šüdämi (u.-f.)
7.	liver	yek ^w n-	qwidz ₁ l	miksə
8.	moon	meH ₁ n	dute	kix ₁
9.	mouse	muHs	tagw-	šijiri
10.	nail	H ₃ nogh-	prcxa-	künci
11.	name	H ₁ nomn	dz ₁ axe-	nimi
12.	night	nok ^w t-	gam-	üji (u.-f.)
13.	no (not)	ne	ma-	e...k
14.	sun	suH ₂ l-	mz ₁ e	kəjə
15.	tear	dak'ru	c ₁ leml-	kuñil
16.	three	treyes	sam-	kolme (u.-f.)
17.	tongue	dng'hu-	nen-	käxli
18.	two	dwoH ₁	yor-	kektä
19.	water	wedn-	s ₁ im-	weti
20.	who	k ^w o-	ma-	ke-
B	engleski	albanski	latvijski	irski
1.	blood	gjak	asinis	fuil /ful'/
2.	bone	asht	kauls	cnámh /kna:v/
3.	ear	vesh	auss	cluas /kluəs/
4.	eye	sy	acs	súil /su:l'/
5.	fire	zjarr	uguns	tine /t'in'ə/
6.	heart	zemër	sirds	croí /kri:/
7.	liver	mëlçi	aknas	ae /e/
8.	moon	hënë	mēness	gealach /g'aləx/
9.	mouse	mi	pele	luchóg /luxo:g/
10.	nail	thua	nags	ionga /iŋə/
11.	name	emër	vārds	ainm /an'v/
12.	night	natë	nakts	oíche /i:x'ə/
13.	no (not)	nuk	ne	ní /n'i:/
14.	sun	diell	saule	grian /g'r'iən/
15.	tear	lot	asara	deoir /d'or'/
16.	three	tre	trīs	trí /t'r'i:/
17.	tongue	gjuhë	mēle	teanga /t'əŋə/
18.	two	dy	du	dó /do:/
19.	water	ujë	ūdens	uisce /us'k'ə/
20.	who	kush	kas	cé /k'e:/

C	engleski	svahili	bahasa indonesia	hrvatski
1.	blood	damu	derah	krv
2.	bone	fupa	kaki	kost
3.	ear	sikio	telinga	uho
4.	eye	jicho	mata	oko
5.	fire	moto	api	vatra
6.	heart	moyo	djantung	srce
7.	liver	ini	hati	jetra
8.	moon	mwezi	bulan	mjesec
9.	mouse	panya	tikus	miš
10.	nail	ukucha	kuku	nokat
11.	name	jina	nama	ime
12.	night	usiku	malam	noć
13.	no (not)	siyo	tidak	ne
14.	sun	jua	matahari	sunce
15.	tear	pasua	matuh	suza
16.	three	tatu	tiga	tri
17.	tongue	ulimi	lidah	jezik
18.	two	wili	dua	dva
19.	water	maji	air	voda
20.	who	nani	siapa	tko

Analiza tablica pokazuje da je među indoeuropskim jezicima, koliko god neslični oni bili, ipak očuvano četrnaest riječi koje su etimološki podudarne barem u dva od tri uspoređivana jezika (brojevi 1, 3, 4, 6, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20)⁹ dok je čak 5 riječi podudarno u sva tri jezika (brojevi 13, 16, 18, 19, 20). Naravno, glasovne podudarnosti među tim riječima nije moguće otkriti na korpusu od samo 20 riječi, no povećavanjem popisa na 100 ili 200 te bi se podudarnosti utvrdile opisanom statističkom metodom. Neformalnom analizom fonetskih sličnosti lako ćemo ustanoviti da među uspoređivanih dvadeset riječi na indoeuropskim jezicima ima između pet i deset njih koje su fonetski slične barem na dva jezika¹⁰. S druge strane, usporedba dvadeset riječi istoga značenja u tri analizirane grane “nostratičkoga” ne obećava mnogo: između indoeuropskih i kartvelskih riječi istoga značenja uopće nema takvih koje bi bile fonetski slične, dok se usporedbom uralskih i ie. riječi nameću samo tri fonetske sličnosti, među riječima br. 11 (‘ime’), 19 (‘voda’) i 20 (‘tko’)¹¹.

⁹ Dvojbeno je da li su etimološki podudarni alb. *gjak* i latv. *asinis* (br. 1), te alb. *diell* i latv. *saule* (br. 14). U oba slučaja meni se čini da se radi o nesrodnim riječima, premda ima autora koji misle suprotno.

¹⁰ Zavisno od toga koliko su nam strogi kriteriji “fonetske sličnosti”. Primjerice, nekome će se albanski *vesh* činiti sličnim latvijskomu *auss*, a nekome neće, premda su te riječi etimološki identične.

¹¹ Sličnosti su među tim uralskim i ie. riječima uočene još davno, usp. Collinder 1934.

Fonetske sličnosti, naravno, nisu nikakav dokaz genetske srodnosti među jezicima. Činjenicu da takvih sličnosti ima između riječi temeljnoga leksika ural-skih i ie. jezika, a nema između kartvelskih riječi i riječi drugih dviju rekonstruiranih prajezika, možda možemo objasniti većom tipološkom podudarnosti rekonstruiranih fonoloških sustava ie. i ural-skog prajezika, te podudarnim pravilima o strukturi riječi u ta dva prajezika, uslijed čega su i slučajne fonetske sličnosti među njihovim riječima češće. Osim toga, pogled na tablicu C pokazat će nam da i između barem dviju indonezijskih riječi i njihovih indoeuropskih istoznačnica postoje fonetske sličnosti, koje ne ukazuju na dubinsku genetsku srodnost, već na činjenicu da se i riječi temeljnoga leksika mogu posuđivati, a mogu biti i slučajno slične: indonezijski *nama* 'ime' posuđenica je iz sanskrta, a sličnost između riječi *dua* i latvijskoga *du* posve je slučajna.

§7 Proširenjem popisa rekonstruiranih riječi za još četrdeset jedinica, i ograničenjem na samo dvije grupe, indoeuropsku i ural-sku, možemo provjeriti da li su fonetske sličnosti među riječima tih dvaju prajezika rezultat pravilnih glasovnih podudarnosti, koje bi možda mogle biti otkrivene i statistički ovjerene u duljoj listi riječi. U razmatranje ćemo dakle uključiti sljedećih četrdeset riječi od kojih gotovo sve pripadaju duljoj Swadeshovoj listi od 200 riječi temeljnoga leksika:

1. 'big' uf. *enä* ie. *meg*H₂-
2. 'cloud' uf. *pilwi* ie. *nebhos*-
3. 'dawn' uf. *koji* ie. *H₂ewso*-
4. 'die' u. *kaxh* ie. *mer*-
5. 'dream' uf. *alma* ie. *swepno*-
6. 'drink' uf. *juxi* ie. *pe*H₃-
7. 'dry' u. *koški* ie. *sawso*-
8. 'earth' uf. *mixi* ie. *dheg'hom*-
9. 'eat' uf. *sewi* ie. *H₁ed*-
10. 'egg' u. *muna* ie. *H₂ōwyo*-
11. 'four' uf. *neljä* ie. *k^wetwore*-
12. 'full' uf. *täwi* ie. *pl*H₁*no*-
13. 'feather' u. *tuxli*, ie. *potn*-
14. 'foot' uf. *jilka* ie. *pod*-
15. 'grind' uf. *pänV*- ie. *mel*H₂-
16. 'hear' uf. *kuuli* ie. *k'lew*-
17. 'knee' u. *puxli* ie. *g'enu*-
18. 'know' u. *tumti* ie. *g'ne*H₃-
19. 'left' u. *waxa* ie. *laywo*-
20. 'light' uf. *wilki* ie. *lew*k-
21. 'long' uf. *pidi* ie. *dl*H₃*gho*-

22. 'louse' uf. *täji* ie. *gnid-*
23. 'lung' u. *täwiw* ie. *plewmo-*
24. 'middle' uf. *küti* ie. *medhyo-*
25. 'mouth' u. *soxi* ie. *H₃eH₁s-*
26. 'naked' uf. *panca* ie. *nog^w-*
27. 'new' uf. *wud'i* ie. *newo-*
28. 'old' uf. *so^{dx}a* ie. *seno-*
29. 'root' u. *wanca* ie. **wrH₂d-*
30. 'see' uf. *näki* ie. *weyd-*
31. 'sleep' uf. *oda* ie. *ses-*
32. 'smoke' u. *künti* ie. *dhuH₂mo-*
33. 'snake' uf. *küji* ie. *H₃eg^whi-*
34. 'snow' u. *lomi* ie. *sneyg^wh-*
35. 'star' u. *kunśa*, ie. *H₂ster-*
36. 'stone' uf. *kiwi* ie. *H₂ek'mon-*
37. 'tooth' uf. *pi^{vi}i-* ie. *H₃don-*
38. 'voice' uf. *äni* ie. *wok^w-*
39. 'what' uf. *mi* ie. *k^wi-*
40. 'woman' uf. *näxi* ie. *g^wenH₂-*

Valja, međutim, istaći sljedeće: 1. broj od 60 uspoređivanih riječi još uvijek je premalen da bi se mogla izvesti dovoljno pouzdana statistička analiza; 2. među novopridodanim riječima ima i takvih koje bi lakše mogle biti posuđenice (npr. riječ za 'snijeg' ili 'zmiju' 3. U produljenoj listi ima više riječi čija indoeuropska i uralska rekonstrukcija nije sasvim sigurna: primjerice, da li je ie. riječ za 'zub' počinjala s *H₃ ili s *H₁? Da li je ie. riječ za 'lijev' *laywo- ili *sewyo-? Znači li ie. *nebhos 'nebo' ili 'oblak'? Sve to čini proširenje naše liste manje iskoristivim za donošenje čvrstih zaključaka o dubinskom jezičnom srodstvu.

Ako se, usprkos svim ogradama, za sve suglasnike u ie. i uralskom koji se pojavljuju u više od četiri riječi izračuna vjerojatnost da se pojavljuju u riječima istoga značenja, dobivaju se za tezu o dubinskom genetskom srodstvu tih jezika prilično nepovoljni rezultati. Evo nekih usporedbi:

$P(\text{Ie. } *k^w : \text{ural. } *k) = 0,1$ (vjerojatnost pojavljivanja *k^w u ie. riječi) \times $0,417$ (vjerojatnost pojavljivanja *k u uralskoj riječi) = $0,0417$. Imamo 2 podudarnosti (riječi koje znače 'tko' i 'jetra', a po formuli o binomnoj raspodjeli to znači da je vjerojatnost da je to slučajno $b(2, 60, 0,0417) = 0,2602$, tj. više od 25% što ne zadovoljava kriterij koji smo postavili¹².

¹² Kao neslučajne prihvaćamo samo skupove podudarnosti čija je vjerojatnost slučajnoga pojavljivanja manja od 10%, v. gore.

$P(\text{ie. } *n : \text{ural. } *n) = 0,333 \times 0,183 = 0,061$. Imamo četiri podudarnosti (riječi koje znače 'žena', 'gol', 'ime' i 'nokat', a po binomnoj formuli vjerojatnost je toga broja podudarnosti 12,89 %.

$P(\text{ie. } *m : \text{ural. } *m) = 0,183 \times 0,17 = 0,031$. Imamo dvije podudarnosti, 'ime' i 'zemlju', no to je događaj čija je vjerojatnost čak 27,38%.

$P(\text{ie. } *k' : \text{ural. } *k) = 0,067 \times 0,417 = 0,028$. Imamo tri podudarnosti, riječi za 'su-zu', 'čuti' i 'kamen', što je po binomnoj formuli događaj s 14,88% vjerojatnosti.

Na koncu, najpovoljniji slučaj za indo-uralsku hipotezu čine riječi koje i u uralskom i u ie. prajeziku sadržavaju *w. Vjerojatnost da će se ti suglasnici pojaviti u riječi istoga značenja jest $0,058 = 0,267 \times 0,217$. Posvjedočeno je čak sedam riječi koje i u uralskom i u indoeuropskom sadržavaju *w: to su riječi za 'vodu', 'uho', 'nov', 'svjetlost', 'pluće', 'korijen' i 'lijev'. Po binomnoj formuli slučajna je vjerojatnost toga događaja samo 3,59%, što zadovoljava naš kriterij, no valja imati na umu da se radi o jednoj jedinoj glasovnoj podudarnosti.

§8 Glavni rezultat ovoga rada jest da za sada nema čvrsta dokaza da je indoeuropska jezična porodica genetski srodna ijednoj drugoj jezičnoj porodici na svijetu. Iz toga ne treba zaključiti da indoeuropski jezici najvjerojatnije *nisu* srodni nijednoj drugoj posvjedočenoj skupini jezika. Jezična srodnost vrlo je prirodno stanje na lingvističkoj karti svijeta, kako danas, tako zacijelo i prije pet-šest tisuća godina kada se govorio ie. prajezik. Ie. prajezik je najvjerojatnije imao genetskih srodnika, i malo je vjerojatno da su baš svi oni u međuvremenu izumrli bez traga. Kada je tako, postavlja se pitanje – kako uopće formulirati hipoteze o dubinskoj genetskoj srodnosti jezika? Ako na temelju našeg istraživanja kažemo da bi ie. prajezik *mogao* biti srodan uralskome, to zacijelo nije pogrešno, ali time nismo mnogo rekli: *svi* bi jezici svijeta mogli biti genetski srodni. Pa ipak, premda za sada ne raspoložemo čvrstim dokazima o genetskim odnosima u to doba, to ne znači da se moramo potpuno kloniti *vjerojatnosnih* sudova o njima: ukoliko teorijska fizika može živjeti s vjerojatnosnim hipotezama, ne vidim zašto lingvistika ne bi mogla. Stvari dakle stoje ovako: indoeuropski prajezik vjerojatnije je genetski srodan uralskom prajeziku nego kartvelskome, a također je vjerojatnije srodan kartvelskome nego, primjerice, sjeverozapadnokavkaskome¹³. I to je više-manje sve što se o tome za sada može znati.

¹³ V. dodatak ovom članku.

DODATAK: Dijakronijski stabilna tipološka svojstva kao pokazatelji
 genetske nesrodnosti jezika

Jedna od prvih stvari koju student nauči na svakom tečaju poredbenopovijesne lingvistike jest da tipološke podudarnosti nisu nikakav dokaz o postojanju genetskih veza među jezicima. Iz nepobitne činjenice da je baskijski ergativan jezik poput abhaskoga nije moguće zaključiti da su ti jezici genetski srodni, ništa više nego što je moguće zaključiti da su srodni hindski i australski aboriginski jezik dyirbal, u kojima također postoji ergativna konstrukcija. Posve ista, i “najezotičnija” tipološka obilježja mogu nastati posve neovisno na različitim stranama svijeta, u jezicima među kojima nema nikakve genetske veze. Međutim, moguće je zamisliti da postoje tipološka svojstva koja imaju takvu dijakronijsku stabilnost da je malo vjerojatno da će ih jezici koji su genetski povezani izgubiti¹⁴. Kada bismo pronašli takav skup tipoloških svojstava, ispitujući njihovu prisutnost u jezičnim porodicama čija je genetska srodnost dokazana, još uvijek ne bismo mogli na osnovi prisutnosti tih svojstava u rekonstruiranim prajezicima zaključivati da su ti prajezici srodni: već smo rekli da bilo koje tipološko obilježje može neovisno nastati u bilo koja dva genetski nesrodna jezika. No kada bismo mogli pokazati da nekoliko takvih dijakronijski stabilnih svojstava postoji u prajeziku A, a ne postoji u prajeziku B, mogli bismo zaključiti da A i B vjerojatno *nisu* srodni, osim možda na tako dubokoj razini koju nikada ne bismo mogli dosegnuti nikakvom metodom.

Istraživanjem takvih dijakronijski stabilnih svojstava osobito se intenzivno bavila u zadnje vrijeme Johanna Nichols (1992, 1993). Koristeći njezin popis dijakronijski stabilnih svojstava¹⁵, pogledajmo ima li razloga za sumnju da su neki od nostratičkih jezika vjerojatno nesrodni. Evo o kojim se tipološkim obilježjima radi:

- razvijen konsonantski sustav (preko cca. 25 suglasnika)
- sustavi imenskih klasa (i gramatičkoga roda)
- postojanje opreke inkluzivnih i ekskluzivnih zamjenica
- osnovna morfosintaktička konstrukcija (akuzativna, ergativna ili aktivna)

¹⁴ To ne mora značiti da se ta obilježja ne mogu posuđivati iz jezika u jezik, tj. da ne mogu imati i arealnu, uz dijakronijsku, stabilnost. Naprimjer, postojanje opozicije inkluzivnih i ekskluzivnih zamjenica vjerojatno je upravo takvo obilježje koje se lako posuđuje iz jezika u jezik, ali se, jednom steknuto, teško gubi (Nichols 1993).

¹⁵ U nastavku teksta u potpunosti ću se oslanjati na rezultate Johanne Nichols, jer vlastitih tipoloških istraživanja dijakronijske stabilnosti nisam provodio. Moji su zaključci dakle onoliko opravdani koliko su njezini rezultati valjani, a ja sam prilično uvjeren da jesu, budući da su zasnovani na egzaktnoj metodologiji.

- velik broj *head-marking* obrazaca u jeziku¹⁶
- osnovni tip adpozicija (prepozicije / postpozicije)
- postojanje kategorije broja kod imenica
- postojanje pasiva kao kategorije.

Usporedimo sada s obzirom na navedena svojstva one “nostratičke” prajezike čiji je gramatički ustroj dovoljno dobro poznat.

	<i>ie.</i>	<i>kartv.</i>	<i>sem.</i>	<i>ural.</i>
više od 25 suglasnika	+ ¹⁷	+	+	—
inkl/ekskl. zamjenica	— ¹⁸	— ¹⁹	— ²⁰	—
osnovna sintak. konstrukcija	nominativna	aktivna	nominativna	nominativna
adpozicije	postpozicije ²¹	postpozicije	prepozicije	postpozicije
sg./pl. kod imenica	+	+	+	+
pasiv	—	+	+	+
imenske klase	+	—	+	—
<i>head-marking</i>	—	+/- ²²	+/-	+/-

¹⁶ *Head-marking* obrasci su sintaktičke konstrukcije u kojima je morfološki obilježena glava sintaktičke konstrukcije. Naprimjer, to je semitski *status constructus* ili indeksiranje izravnoga objekta na glagolu u kartvelskim jezicima.

¹⁷ Točan broj konsonanata u *ie.* prajeziku nije utvrđen zbog neslaganja oko broja guturala i laringala, no prema većini teorija taj broj prelazi dvadeset i pet suglasnika.

¹⁸ Prema nekim shvaćanjima i u *ie.* prajeziku je postojala opozicija inkluziva i ekskluziva, koja se očituje u razlici između osnove *no- (lat. *nos*) i *we- (skr. *vayam*) osobnih zamjenica 1. l. mn.

¹⁹ Opreka inkluzivnih i ekskluzivnih zamjenica postoji u svanskom, no nije posve sigurno da li je to arhaizam koji treba pripisati kartvelskom prajeziku ili rezultat utjecaja nahsko-dagestanskih jezika.

²⁰ Opreka inkluzivnih i ekskluzivnih zamjenica 1. l. mn. postoji u nekim čadskim jezicima, dijelu afroazijske file kojoj pripadaju semitski jezici. Nije moguće utvrditi da li je riječ o arhaizmu, tj. da li je ta opreka postojala u praafrazijskome.

²¹ Premda većina *ie.* jezika ima prepozicije (prijedloge), podudarnosti između hetitskoga i najstarijega vedskoga čine vjerojatnim zaključak da je *ie.* prajezik imao većinom, ili isključivo postpozicije (poslijeloge).

²² Dok je *ie.* prajezik izraziti predstavnik *dependent-marking* jezika (jezika koji morfološki obilježuju zavisni član sintaktičke konstrukcije), ostali jezici s našega popisa imaju podjednak broj *head-* i *dependent-marking* obrazaca.

Premda se tablica dijakronijski stabilnih svojstava ie. prajezika ne slaže s tablicom tih svojstava niti u jednom drugom navedenom “nostratičkom” jeziku, najveće su podudarnosti sa semitskim prajezikom. Iz toga ne treba izvlačiti dalekosežne zaključke, ali prilično je jasno da tipologija dijakronijski stabilnih obilježja ne govori *protiv* teze o genetskom zajedništvu navedenih “nostratičkih” jezika. S druge strane, ako bismo pogledali istu takvu tablicu za abhasko-adigejske (sjeverozapadnokavkaskе) jezike, uvidjeli bismo da između njih i ie. prajezika nema gotovo nikakve podudarnosti: abhasko-adigejski jezici imaju ergativnu sintaksu, prepozicije, nemaju opreke jednine i množine kod imenica, nemaju morfosintaktički izražen pasiv, i predstavljaju prototipske *head-marking* jezike. Odatle nam opravdanje da zaključimo da su ie. jezici *vjerojatnije* srodni, primjerice, kartvelskima, nego sjeverozapadnokavkaskim jezicima.

BIBLIOGRAFIJA

- Abondolo, Daniel (ed.) 1998. *The Uralic Languages*, London: Routledge.
- Bomhard, Allan 1994. *The Nostratic Macrofamily: a Study in Distant Linguistic Relationship*, Berlin: Mouton de Gruyter.
- Burrow, T. & M. B. Emenau 1984. *A Dravidian Etymological Dictionary*, Oxford: Clarendon.
- Campbell, Lyle 1997. Prikaz knjige Greenberg 1987, *Language* 64 / 1988: 591-615. *American Indian Languages. The Historical Linguistics of Native America*, Oxford: Oxford University Press.
- Campbell, L. & M. Mithun 1979. *The Languages of Native America: Historical and Comparative Assessment*, Austin : University of Texas Press.
- Collinder, Björn 1934. *Indo-Uralisches Sprachgut*, Uppsala.
- Collinder, Björn 1977. *Fenno-Ugric Vocabulary: an Etymological Dictionary of the Uralic Languages*, Hamburg: Buske.
- Čop, Bojan 1974. *Indouralica*, Ljubljana: SAZU.
- Décsy, Gyula 1965. *Einführung in die finnisch-ugrische Sprachwissenschaft*, Wiesbaden: Harrassowitz.
- Dixon, R. M. W. 1981. *The Languages of Australia*, Cambridge: Cambridge University Press.
- Dörfer, Gerhard 1973. *Lautgesetz und Zufall. Betrachtungen zum Omnikomparatismus*, Innsbrucker Beiträge zur Sprachwissenschaft, Innsbruck.
- Dolgopolsky, Aaron B. 1964. Gipoteza drevnejšego rodstva jazykovyx semej Severnoj Eurazii s verojatnostnoj točki zrenija, *Voprosy jazykoznanija* 2/1964: 53-63.
- Ehret, Christopher 1995. *Reconstructing Proto-Afroasiatic (Proto-Afrasian)*, Berkeley, LA and London: University of California Press.
- Eyuboglu, Ismet 1988. *Türk dilinin etimoloji sözlüğü*, Istanbul: Sosyal Yagular.
- Fährnich, J. 1995. *Etymologisches Wörterbuch der Kartwel-Sprachen*, Leiden: Brill.

- Goddard, I. 1975. Algonquian, Wiyot, and Yurok: proving a distant genetic relationship. In: *Linguistics and anthropology: in Honor of C. F. Voegelin*, ed. by Dale Kinkade, Kenneth L. Hale, and Oswald Werner, Lisse: The Peter de Ridder Press, 249-262.
- Greenberg, Joseph 1987. *Language in the Americas*, Stanford: Stanford University Press.
- Holzer, Georg 1995. *Das Erschließen unbelegter Sprachen*, Frankfurt a/M: Peter Lang Verlag.
- Illič-Svityč, V. M. 1971. *Opyt sravnenija nostratičeskix jazykov. Vvedenije. Sravnitel'nij slovar'*, Moskva: Nauka.
- Ivanov, V. V. 1980. Prajazyki kak objekty opisaniya v izdanii 'Jazyki mira'. *Teoretičeskie osnovy klassifikacii jazykov mira*, Moskva: Nauka, 181-204.
- Klein, Ernest 1987. *A Comprehensive Etymological Dictionary of the Hebrew Language for Readers of English*, London: Macmillan.
- Klimov, G. 1964. *Etimologičeskij slovar' kartvel'skix jazykov*, Moskva: Nauka.
- Lakó, György 1967-78. *A magyar szókészlet finnugor elemei; etimológiai szótár*, Budapest: Akadémiai kiadó.
- Levitskaja, L. S. (ur.) 1989. *Etimologičeskij slovar' tjurkskix jazykov* (1), Moskva: Nauka.
- Lyovin, Anatole 1997. *An Introduction to the Languages of the World*, Oxford: Oxford University Press.
- A Magyar nyelv történeti-etimológiai szótára*, Akadémiai kiadó, Budapest 1967-78.
- Matasović, Ranko 1997. *Kratka poredbenopovijesna gramatika latinskoga jezika*, Zagreb: Matica hrvatska
- Matisoff, James 1990. On Megalocomparison. *Language* 66 : 106-120.
- Manaster Ramer, Alexis 1993. On Illič-Svityč's Nostratic Theory. *Studies in Language* 17: 205-250.
- Meillet, Antoine 1967. *La méthode comparative en linguistique historique*, Paris: Champion.
- Nichols, Johanna 1992. *Linguistic Diversity in Space and Time*, Chicago: The University of Chicago Press
- Nichols, Johanna 1993. Diachronically Stable Structural Features. In: H. Andersen (ed.) *Historical Linguistics*, Amsterdam: Benjamins, 337-355.
- Oswalt, R. 1991. A Method for Assessing Distant Linguistic Relationships. In: Sidney M. Lamb & E. Douglas Mitchell, eds. *Sprung from some Common Source*, Stanford: Stanford University Press.
- Rédei, K. 1986-88. *Uralisches etymologisches Wörterbuch*, Budapest: Akadémiai Kiadó.
- Ringe, Donald A. Jr. 1992. On Calculating the Factor of Chance in Language Comparison. *Transactions of the American Philosophical Society* 82: 1-110.
- Ringe, Donald A. Jr. 1993. A Reply to Professor Greenberg. *Proceedings of the American Philosophical Society* 137: 91-109.
- Ringe, Donald A. Jr. 1995. Nostratic and the factor of chance. *Diachronica* 12: 55-74.
- Ruhlen, Merritt 1987. *A Guide to World's Languages, Vol. 1: Classification*, Stanford: Stanford University Press.

- Salmons, J. C. & Brian Joseph (eds.) 1998. *Nostratic: Sifting the Evidence*, Amsterdam: Benjamins.
- Serebrennikov, A. J. 1981. Počemu nel'zja verit' storonnikam 'nostratičeskoj' teorii. *Voprosy jazykoznanija* 3: 1-9.
- Sevortjan, E. V. 1974-1980. *Etimologičeskij slovar' tjurkskih jazykov (A-D)*, Moskva: Nauka.
- Sihler, Andrew 1998. *Principles of Historical Linguistics*, (rukopis), Madison.
- Sinor, Denis (ed.) 1988. *The Uralic Languages*, Leiden: Brill.
- Trask, R. L. 1996. *Historical Linguistics*, London: Arnold.
- Vine, Brent 1991. Indo-European and Nostratic. *Indogermanische Forschungen* 96: 9-35.
- Vranić, Vladimir 1970. *Vjerojatnost i statistika*, Zagreb: Tehnička knjiga.

The distant genetic relatedness of languages

Summary

In this paper a new method for assessing the distant genetic relatedness of languages is proposed, and applied to three "Nostratic" protolanguages. However, the statistical analysis of sound correspondences in the reconstructed basic vocabularies of PIE, Uralic, and Kartvelian protolanguages shows that the distribution of probabilities of those correspondences does not differ significantly from the expected distribution of probabilities of random correspondences.

Ključne riječi: genetska srodnost, temeljni leksik, indoeuropski, uralski, kartvelski, nostratički

Key words: genetic relatedness, basic vocabulary, Indo-European, Uralic, Kartvelian, Nostratic