

Anita SUJOLDŽIĆ, Lajos SZIROVICZA, Konstantin MOMIROVIĆ,  
Božidar FINKA, Milan MOGUŠ, Petar ŠIMUNOVIĆ i Pavao RUDAN

## PRIMJENA TAKSONOMSKIH ALGORITAMA NA NENUMERIČKE VARIJABLE U PROUČAVANJU LINGVISTIČKE MIKROEVOLUCIJE

Ustanove autora:

Institut za medicinska istraživanja i medicinu rada,  
Zavod za jezik,  
Sveučilišni računski centar SRCE,  
Filološki odjel Filozofskog fakulteta Sveučilišta u Zagrebu

U suvremenom antropologijskom pristupu izučavanju ljudskih populacija nailazimo i na potrebu obuhvatnijeg studija niza ne samo bioloških već i kulturnih, etnoloških, lingvističkih i drugih svojstava ispitivanih skupina, ali i svojstava eko-sustava u kojem se one nalaze i razvijaju (Rudan, 1973, 1975; Rudan i sur., 1977; Rudan i sur., 1978). Već je i dosad primijećeno postojanje izvjesnih analogija između razdiobe pojedinih bioloških i pojedinih lingvističkih svojstava npr., koje ipak nisu — utvrđeno je — u svim istraživanjima bile istovjetne. Štoviše, pokazale su gdjekad čak i posve različite rezultate (npr. Hymes, 1960; Howells, 1966; Spuhler, 1972; Spielman i sur., 1974; spomenimo samo neke). Želeći primijeniti i u našim proučavanjima — vezanima za studij mikroevolucije pojedinih ljudskih skupina — multidisciplinarn pristup (kojim bi sjedinili proučavanja i bioloških, ali i niza lingvističkih, kulturnih i drugih svojstava npr.), prišli smo pokušaju početne analize nekih lingvističkih karakteristika stanovnika Istre kao »modela« (Sujoldžić i sur., 1978a, 1978b), i to stoga što se već niz godina upravo na tom području vrše bio-antropološka istraživanja (Rudan, 1978; Rudan i sur., 1978; Rudan i sur., 1979). Smatramo, naime, kako upravo treba započeti i antropolingvistička istraživanja u nas, budući da je područje naše zemlje (s nizom vrlo bogatih lingvističkih sadržaja) od izuzetnog značenja i za teorijska i za praktična istraživanja. Budući da je pri našem istraživanju primijenjena eksperimentalna metoda, svaki se pokušaj njezine primjene može iskoristiti i u utvrđivanju njezine vrijednosti kao metode koju možemo primijeniti u antropolingvističkim i etnolingvističkim proučavanjima.

Istraživanje se temelji na pretpostavci da u svim jezicima postoji fond riječi vezanih za neke temeljne kategorije opće ljudske kulture, a najotporniji je na promjene, i to za razliku od drugih kulturnih elemenata, kulturnog vokabulara npr. To je oko stotinu riječi (na *Tablici br. 1.* prikazane su riječi kojima je izvršena analiza lingvističke mikroevolucije) koje smo za ukupno 16 sela na području Istre (Grdoselo, 1; Gornji Rabac, 2; Brseč, 3; Žminj, 4; Medulin, 5; Pazin, 6; Boljun, 7; Lupoglav, 8; Nugla, 9; Sv. Martin, 10; Brest, 11; Rovinjsko Selo, 12; Vabriga, 13; Kaštelir, 14; Kaldir, 15; Rakalj, 16) uzeli iz »Dijalektološkog upitnika« za hrvatski ili srpski dijalektološki atlas, pohranjen u Zavodu za jezik u Zagrebu. Podatke su na terenu prikupili P. Šimunović i M. Moguš. Svako je od ukupno 106 riječi dan — bez obzira na semantičku, morfološku ili fonetičku karakteristiku — broj, različit za svaku, pa i najmanju razliku u pojedinom selu. »Lingvističke udaljenosti« između sela procijenjene su s pomoću Hemingovih mjera udaljenosti. Kako su se lingvistička obilježja mogla opisati samo skupom nominalnih varijabli, Hemingove su mjere očito najjednostavniji i najprirodniji (iako ne nužno i najpodobniji sa statističke i taksonomske točke gledišta) instrument za određivanje razdaljine ma koja dva virtualno n-narna vektora. Taksonomska procedura nad matricom Hemingovih mjera udaljenosti izvedena je jednim od algoritama za taksonomsku analizu u prostoru multinominalno distribuiranih nominalnih varijabli, koju su predložili *Kudnjevcev i sur. (1977)*. Taj algoritam formira grupe sela jednim iterativnim postupkom čije je ponašanje moguće kontrolirati izborom početnog minimalnog radijusa i određivanjem frakcije njegova porasta u svakoj iteraciji. Inicijalni objekt definiran je kao objekt s najmanjom prosječnom Hemingovom udaljenošću od svih ostalih objekata. Tako dolazimo do grupiranja pojedinih sela sa sličnijim obilježjima u ispitivanim riječima. Na *Slici br. 1.* prikazane su »izoglose«, nastale uklapanjem u pojedinu skupinu lingvistički sličnijih sela, počevši od onih koja su među sobom najsličnija (selo br. 10 — Sv. Martin i selo br. 3 — Brseč). Na našem materijalu (*Slika br. 1. i Tablica br. 2.*) opaža se postojanje veće sličnosti između udaljenijih sela, priobalnoga te tzv. ravničarskoga područja Istre, nasuprot većoj različitosti između geografski čak i bližih sela (no lingvistički udaljenijih), u planinskoj regiji. Ta se slika poklapa, donekle, s područjima naseljavanja poluotoka Istre, te s razdiobom pojedinih govornih dijalekata i narječja u ispitivanom području. Na *Slici br. 1.* opaža se i impresija u sjeverozapadnom dijelu poluotoka za koju možemo, možda, pretpostaviti da je nastala zbog utjecaja lingvističkih karakteristika slovenskog jezika. U ovom prikazu rezultata i raspravi tu pretpostavku, na žalost, nismo u mogućnosti potvrditi, budući da nam nisu stajali na raspolaganju »Dijalektološki upitnici« tog područja Istre.

Rezultati naših preliminarnih istraživanja mislimo da ipak upozoravaju na slijedeće:

- kako postoji određeno grupiranje pojedinih sela s obzirom na istovjetnosti lingvističkih karakteristika u ispitivanim riječima;
- kako je nalaz približno istovjetan onom koji je uočen u dosadašnjim lingvističkim istraživanjima s obzirom na karakteristike pojedinih dijalekata i narječja u Istri;
- kako geografska konfiguracija poluotoka — s obzirom na tzv. ravninarsku i tzv. planinsku regiju — donekle odgovara razdiobi sličnosti i različnosti ispitivanih govornih jezika, narječja i dijalekata;
- kako uz pomoć taksonomskih postupaka nad nenumeričkim varijablama možemo, možda, pratiti neke karakteristike lingvističke mikroevolucije, napose lingvističke migracije.

Preliminarna istraživanja prikazana u ovoj raspravi ipak treba uzeti s velikom rezervom, budući da:

- analiza obuhvaća samo 106 riječi iz tzv. »bazičnog vokabulara«, te da te riječi možda i nisu dostatne da bi se u potpunosti okarakterizirao jezik određenog sela;
- riječi tzv. »kulturnog vokabulara« — koje su vjerojatno više pod utjecajem brzih promjena tijekom evolucije — neće pokazivati istovjetnost nalaza s obzirom na »izoglose«, ali će zbog toga pokazivati možebitnu mozaičnost lingvističke mikroevolucije;
- istraživanjima treba obuhvatiti i ostale regije poluotoka Istre, kako bi se — možda — utvrdilo postojanje i jednog drugog centra (ili pak više njih) oko kojeg bi započelo istovjetno okupljanje lingvistički sličnijih sela;
- vjerojatno postoji i više »mjesta okupljanja«, napose ukoliko u analizu uzmemo riječi tzv. kulturnog vokabulara, riječi materijalne kulture, lascivne riječi i sl.;
- kako bi trebalo prići klasifikaciji razlika među analiziranim riječima.

Valja, naime, istaknuti kako je ovom analizom bilo kakva razlika — na fonološkoj, akcenatskoj, morfološkoj ili leksičkoj razini — okarakterizirana samo kao »postojanje razlike«, dok bi valjalo pristupiti klasifikaciji razlika prema njihovu značenju. Valjalo bi stoga skupiti podatke koje bi bilo moguće hijerarhizirati s obzirom na tri različita nivoa. Vjerujemo kako bismo u tom slučaju s pravom dobili potvrdu o mogućnosti primjene nenumeričkih taksonomskih algoritama u proučavanju lingvističke mikroevolucije. Nalaz bi tada, vjerojatno, bio donekle drugačiji. Stoga je ovaj rad samo preliminarni izvještaj o našem pokušaju primjene lingvističkih antropoloških istraživanja, ali i poziv da se i na našim populacijama prate problemi vezani uz mikroevoluciju, i to: analogni problemi i bioloških i kulturnih sustava ispitivanih skupina.

## L I T E R A T U R A

1. HOWELLS, W. W.: 1966. Population distances: biological, linguistic, geographical, and environmental, *Curr. Anthropol.* 7, 531.
2. HYMES, D. H.: 1960. Lexicostatistics so far, *Curr. Anthropol.* 1, 3.
3. KUDNJEVCOV, V., MOMIROVIĆ, K., SZIROVICZA, L., GREDELJ, M., KNEP, A.: 1977. Neobjavljena analiza.
4. RUDAN, P.: 1973. O nekim problemima suvremene biološke antropologije, *Lij. vjesnik*, 95, 59.
5. RUDAN, P.: 1975. Problemi suvremene biološke antropologije — uz djelo J. Hiernauxa »Jednakost ili nejednakost rasa?«, p. 5, Školska knjiga, Zagreb.
6. RUDAN, P.: 1978. Biological structure of the Istrian population — study of »distances« by anthropometric traits, dermatoglyphic properties and blood group gene frequencies, *Coll. Antropol.* 2, 53.
7. RUDAN, P., MAVER, H., DIMOV, D., KOVAČEVIĆ, M., SUJOLDŽIĆ, A., ŠKRINJARIĆ, I.: 1978. O biomedicinskom pristupu antropološkoj znanosti u nas, te o časopisu »Collegium Antropologicum«, *Lij. vjesnik* 100, 253.
8. RUDAN, P., MAVER, H., VLAHOVIĆ, P.: 1977. Anthropology as a natural science — problems, investigations and prospects, *Coll. Antropol.* 1, 26.
9. RUDAN, P., RUNJAK, T., MACAROL, B., PIŠL, Z., ŠKRINJARIĆ, I., GOMZI, M., SCHMUTZER, L.J., ZUŠKIN, E.: 1978. The analysis of quantitative dermatoglyphic traits in the Istrian population — a contribution to the study of »biological structure«, *Acta med. iug.* 32, 385.
10. RUDAN, P., SCHMUTZER, L.J., ŠKRINJARIĆ, I., MACAROL, B.: 1979. Quantitative characteristics of digito-palmar dermatoglyphic in the study of biological structure of population, ref. The Second workshop on medical genetics, Zagreb.
11. SPIELMAN, R. S., MIGLIAZZA, E. C., NEEL, J. V.: 1974. Regional linguistic and genetic differences among Yanomama Indians, *Science* 184, 637.
12. SPUHLER, J. N.: 1972. Genetic, linguistic and geographical distances in Native North America, u: Weiner, J. S., Huizinga, J.: The assessment of population affinities in man, p. 72, Clarendon Press, Oxford.
13. SUJOLDŽIĆ, A., FIO, D., RUDAN, P.: 1978. Glotokronologija u antropološkim proučavanjima mikro-evolucije, *Zbornik na Materijali, Simpozijum biomedicinska kibernetika*, p. 2.35, Skopje.
14. SUJOLDŽIĆ, A., SZIROVICZA, L., MOMIROVIĆ, K., FINKA, B., MOGUŠ, M., ŠIMUNOVIĆ, P., RUDAN, P.: 1978. Primjena taksonomskih algoritama na nenumeričke varijable u proučavanju lingvističke mikro-evolucije, Referat, XVII. kongres antropološkog društva Jugoslavije, Pirot.

Tablica br. 1.

Riječi kojima je izvršena analiza:

1. glava	37. koljeno	73. mast
2. obraz	38. čovjek	74. glad
3. čelo	39. živjeti	75. mjera
4. tijelo	40. umrijeti	76. međaš
5. krv	41. ime	77. sjeme
6. prišt	42. otac	78. mlin
7. čir	43. mati	79. ovca
8. gnoj	44. dijete	80. kozji
9. nos	45. sin	81. konj
10. oko	46. kći	82. kobila
11. zjenica	47. djed	83. kola
12. obrva	48. unuka	84. mlijeko
13. suza	49. sestra	85. pas
14. spavati	50. muž	86. jaje
15. san	51. žena (supruga)	87. divlji
16. uho	52. žensko	88. pčela
17. čuti	53. laž	89. uš
18. usta	54. momak	90. drvo
19. zub	55. djevojka	91. lišće
20. jezik	56. košulja	92. cvijet
21. kost	57. prati	93. suša
22. zglob	58. kuća	94. sjekira
23. grudi (prsa)	59. krov	95. kamen
24. sise (dojke)	60. vrata	96. put
25. njedra	61. vatra (oganj)	97. cesta
26. rebro	62. žar	98. međa
27. pleća	63. pepeo	99. magla
28. ruka	64. mjehur	100. kiša
29. žulj	65. dim	101. sunce
30. lakat	66. daska	102. zvijezda
31. dlan	67. nož	103. mjesec
32. prst	68. dno	104. uštap
33. nokat	69. jesti	105. sušica
34. noga	70. hraniti	106. grob
35. íci	71. žrvanj	
36. bedro	72. kruh	

Tablica br. 2.

Vrijednosti Hemingovih mjera udaljenosti između pojedinih parova sela:

R	P	HU	R	P	HU	R	P	HU
1.	2/1	0,1226	41.	10/5	0,0943	81.	14/3	0,2075
2.	3/1	0,2358	42.	10/6	0,1792	82.	14/4	0,2169
3.	3/2	0,1792	43.	10/7	0,1886	83.	14/5	0,3301
4.	4/1	0,2169	44.	10/8	0,2169	84.	14/6	0,3113
5.	4/2	0,1320	45.	10/9	0,2924	85.	14/7	0,2547
6.	4/3	0,2075	46.	11/1	0,2547	86.	14/8	0,3018
7.	5/1	0,5094	47.	11/2	0,1132	87.	14/9	0,1792
8.	5/2	0,1037	48.	11/3	0,1320	88.	14/10	0,1886
9.	5/3	0,1603	49.	11/4	0,1509	89.	14/11	0,2547
10.	5/4	0,1415	50.	11/5	0,1981	90.	14/12	0,3301
11.	6/1	0,2452	51.	11/6	0,2264	91.	14/13	0,5094
12.	6/2	0,1415	52.	11/7	0,2264	92.	15/1	0,3584
13.	6/3	0,1981	53.	11/8	0,2735	93.	15/2	0,1886
14.	6/4	0,4056	54.	11/9	0,3113	94.	15/3	0,2452
15.	6/5	0,1698	55.	11/10	0,2452	95.	15/4	0,2735
16.	7/1	0,2547	56.	12/1	0,4245	96.	15/5	0,2358
17.	7/2	0,1226	57.	12/2	0,1226	97.	15/6	0,3490
18.	7/3	0,2075	58.	12/3	0,1698	98.	15/7	0,3490
19.	7/4	0,3773	59.	12/4	0,1792	99.	15/8	0,4056
20.	7/5	0,1690	60.	12/5	0,5471	100.	15/9	0,2547
21.	7/6	0,5849	61.	12/6	0,2264	101.	15/10	0,3301
22.	8/1	0,3490	62.	12/7	0,1792	102.	15/11	0,3490
23.	8/2	0,1792	63.	12/8	0,2075	103.	15/12	0,2452
24.	8/3	0,2547	64.	12/9	0,1415	104.	15/13	0,3962
25.	8/4	0,3301	65.	12/10	0,1320	105.	15/14	0,4905
26.	8/5	0,1698	66.	12/11	0,2075	106.	16/1	0,5943
27.	8/6	0,3773	67.	13/1	0,3584	107.	16/2	0,1037
28.	8/7	0,4528	68.	13/2	0,1320	108.	16/3	0,2169
29.	9/1	0,1698	69.	13/3	0,2075	109.	16/4	0,1226
30.	9/2	0,1132	70.	13/4	0,1320	110.	16/5	0,6981
31.	9/3	0,1037	71.	13/5	0,3773	111.	16/6	0,1886
32.	9/4	0,1698	72.	13/6	0,2264	112.	16/7	0,1981
33.	9/5	0,1226	73.	13/7	0,2169	113.	16/8	0,1981
34.	9/6	0,2358	74.	13/8	0,2075	114.	16/9	0,1226
35.	9/7	0,2169	75.	13/9	0,1037	115.	16/10	0,0943
36.	9/8	0,2075	76.	13/10	0,1037	116.	16/11	0,1886
37.	10/1	0,1603	77.	13/11	0,2358	117.	16/12	0,5566
38.	10/2	0,0849	78.	13/12	0,3584	118.	16/13	0,3773
39.	10/3	0,0754	79.	14/1	0,4056	119.	16/14	0,3584
40.	10/4	0,1698	80.	14/2	0,1509	120.	16/15	0,2358

R = redni broj

P = parovi sela

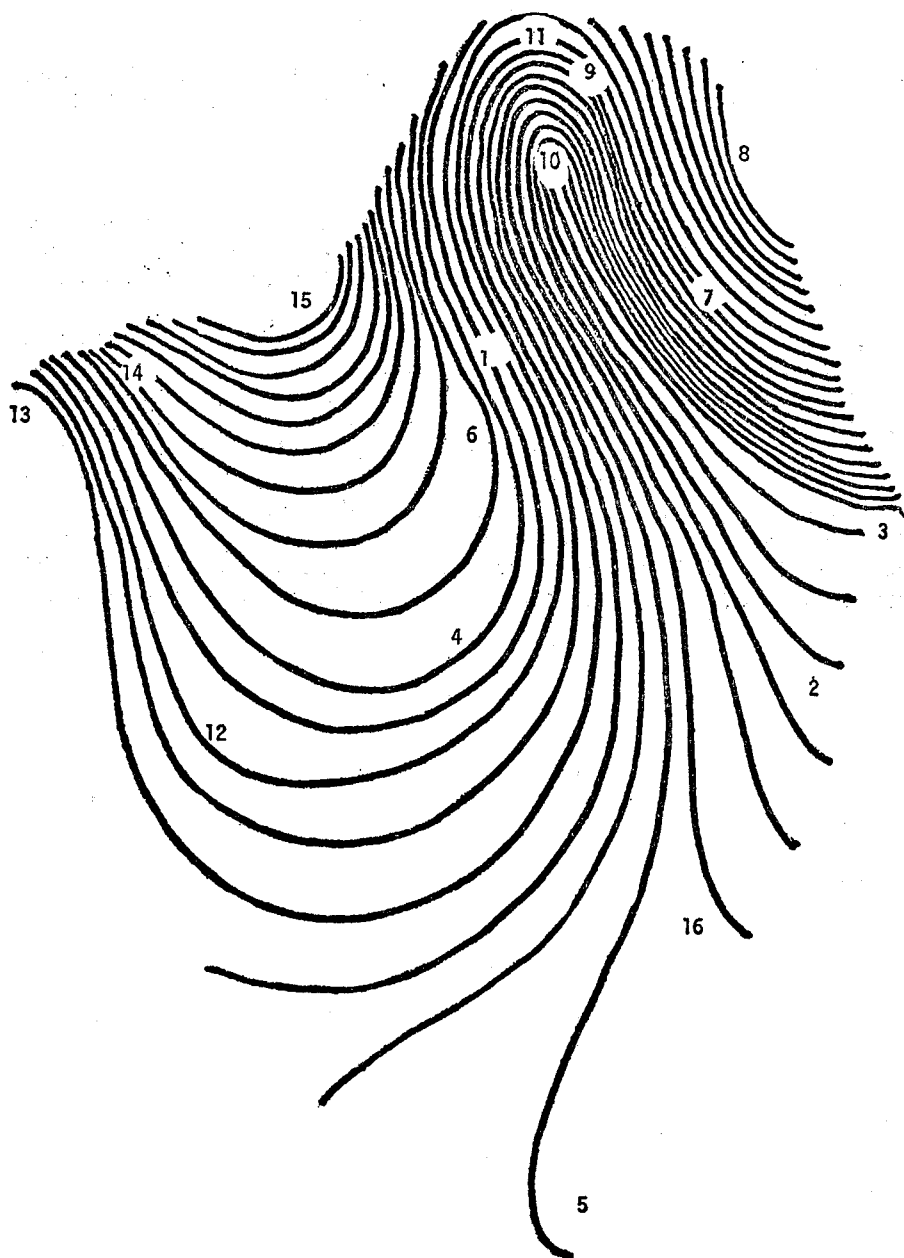
HU = vrijednosti mjera Hemingovih udaljenosti

## Résumé

### APPLICATION DES ALGORITHMES TAXONIMIQUES SUR LES VARIABLES NUMÉRIQUES DANS LES RECHERCHES LINGUISTIQUES D'UNE MICRORÉGION

Partant du fait qu'il y a une analogie entre les différentes caractéristiques biologiques et linguistiques sur un même territoire, les auteurs ont considéré nécessaire d'entamer les recherches anthropolinguistiques, ou plus précisément ethno-linguistiques.

A cette fin a été choisie une microrégion appartenant au territoire istrien (en totalité 16 villages) et on a recueilli une collection comportant 106 mots pour lesquels on a pensé qu'il sont très réfractaires au changement. »La distance linguistique« entre les villages a été déterminée au moyen des paramètres d'Hermin. C'est ainsi qu'on a obtenu un groupement des villages au degré maximum de similitude dans la recherche des mots mentionnés mais ce groupement ne correspond pas toujours au groupement géographique dans la réalité. Le travail en question n'a qu'un caractère tout expérimental, pourtant les auteurs sont convaincus de la possibilité d'une application des algorithmes non-numériques, taxonomiques dans les recherches linguistiques de la microévolution.



*Sl. 1. Izoglose dobivene ucrtavanjem Hemingovih mjera udaljenosti između ispitivanih naselja u Istri*