

Tőszókinszünk etimológiai rétegenkénti hangstatisztikája*

1. Vegyük azt a hatezer-egynéhány vitathatatlan (teljesen motiválatlan) magyar tőszót, amelyet az ÉrtSz. önálló szócikkkel rendelkező tőszóként nyilvántart, tehát az olyanokat, mint *lát, fű, kék* stb. (E tőszavakról valamivel bővebben l. NytudÉrt. 58: 520—521.) Osszuk őket tíz etimológiai csoportra a SzófSz. nyújtotta minősítésnek megfelelően: nincs benne a SzófSz.-ban (rövidítve: „0”), fgr., magyar belső keletkezésű stb., „egyéb” (erről bővebben l. alább; rövidítve: „8”), „bizonytalan, ismeretlen stb.” (rövidítve: „9”). E csoportokon belül végezzünk hangstatisztikát — immár ne csupán a végző hangokra vonatkozóan (ezt l. NyK. 70: 200 kk.), hanem minden egyes szó minden egyes hangját felölelően.

Bevezetőül némi magyarázatra a „hang” fogalma szorul. Bizonyos célszerűségi-felhasználhatósági okokból kiindulva felmérésünk nem mindenütt és nem pontosan hangokat ábrázol, hanem az egyes hangoknak megfelelő helyesírási jeleket. Így külön szerepelnek nálunk az *a, á, b, bb* (= hosszú „b”), *c, cc, cs, ccs, d, dd, dz, ddz, dzs, ddzs* stb. . . . *z, zz, zs, zzs* szimbólumok, vagyis tulajdonképpen minden egyes olyan szimbólum, amelynek egy-egy fonéma felel meg mai felfogásunk szerint. Van azonban legalább egy kétséges pár: a *ch, cch* (ez utóbbi a magyar költészetben előfordul a *Bacchus* tulajdonnév miatt) — nem tudjuk, külön fonéma lesz-e ez, avagy nem, ezért inkább külön kezeltük őket (a *cch* tőszavainkban nem fordult elő). Aki ezzel nem ért egyet, az az ide vonatkozó (ez alatt a szimbólum alatt hozott) adatokat adja hozzá a *h, hh* alattiakhoz (NB: abban sem vagyunk biztosak, hogy lesz-e a magyar rendszerben végül is hosszú „h” fonéma — de azért inkább ezt is külön kezeltük). Van egy írásos elem, mely nyilvántartásainkban külön szerepel, itt azonban a táblázaton eltűnt: ez a „w” jel, melynek adatait a „v”-éi alá soroltuk be. Az ÉrtSz. anyagában egyetlen egyszer előfordul a „q” írásos jel, ez egyben tőszó is. Ha ennek ellenkezőjére nincs külön utalás, mégis teljesen figyelmen kívül hagytuk, mert nem lehet tudni, hogy mely hangok felelnek meg neki: „k”? „kv”? „kú”? „métermázsa”? Hasonlóan nehéz a helyzet az *y* írásos jelet tartalmazó négy szó esetében: ezt a jelet, bár nem felel meg neki önálló hang (fonéma), táblázatunkon külön kezeljük és eredményeit az összeredménybe beszámítjuk; a *brandy* miatt a magán-

* Jelen munka kiinduló adatait a KLTE Odra 1013 gépén nyertük. A szerző őszinte köszönetét fejezi ki JÉKEL PÁL tudományos munkatársnak, aki a hangfelismerő és statisztikázó programot összeállította, Sarkadi Antalné operátornak, aki a vizsgálandó anyagot lelyukasztotta és a KLTE Számológépközpont minden dolgozójának, akik a program sikeres futtatását lehetővé tették.

hangzók számításakor eggyel (ti. egy „i” hanggal) növeljük ezek mennyiségét, a *yard* miatt eggyel (ti. egy „j”-vel) növeljük ugyanakkor a mássalhangzókat is; az *y* és *nylon* szavaknak az *y* jelölte hangelemét ugyanakkor a mássalhangzó—magánhangzó szembeállításakor nem vettük figyelembe (az *y* címszónál a „q”-éhoz hasonló nehézségek merülnek fel, a *nylon* címszónál pedig nem tudom, fel kellett volna-e vennem egy „ej” vagy „áj” diftongust — ez egészen egyedül álló lenne mai rendszerünkben). Meghagytuk végül együtt az *x* jelölte hangcsoportot: még mindig könnyebb innen kivánság szerint a *k* és a *sz* (helyenként esetleg a *g* és a *z*?) hangok adataihoz hozzáadni őket, mint ha eleve oda lettek volna beolvasztva, és onnan kellene őket össze-szednie annak, akit speciálisan az *x* jelölte elem érdek. Ezen hozzáadások és elhagyások után maradt összesen 70 egységünk: köztük szerepelnek tehát külön-külön mind a rövid, mind a hosszú magán- és mássalhangzók; külön egység itt az *x*; a mondott felemás módon szerepel az *y*; nem szerepel a *q*; a *v* alatt szerepel a *w*. Ha a továbbiakban külön kikötés nélkül „hang”-ot mondunk, akkor ezen e 70 egység valamelyikét kell értenünk.

Közleményünk célja az, hogy megadja az ÉrtSz.-ban szereplő tőszókinésünk egészének, valamint egyes etimológiai rétegeinek az ilyen értelemben vett hangstatisztikáját: a pusztá számokhoz némi magyarázatokat is fűzve, az egyes etimológiai rétegeket e szempontból jellemezni próbálva. (Épp e magyarázó-jellemző kísérletek bátorítottak fel bennünket arra, hogy VÉRTES EDIT átfogó munkájának megjelenése után húsz évvel újra felvegyük lényegében ugyanazt a fonalat.¹) Azt lehetne ellenvetni: kár ezzel ma már a SzófSz. alapján foglalkozni, amikor hamarosan megjelenik a TESz. harmadik, befejező kötete is. Ez az ellenvetés azonban véleményünk szerint több szempontból jogosulatlan. Egyrészt: a „hamarosan” még nem ma van, az ember joggal kíváncsi ezekre az eredményekre minél előbb, még ha ezek az eredmények közelítőek is. Másrészt: korábban kimutattuk, hogy ez a közelítés a fontosabb s régebbi etimológiai rétegek esetében a TESz. megjelenése után sem fog érezhetően változni; egész etimológiai összképünk arányai várhatóan változni fognak a nemzetközi vándorszavak és az átlátszó német etimológiájuk „szaporodásával” (tulajdonképpen: első összegező etimológiai számbavételével — vö. NyK. 71: 129 kk.). És végül: valóban, hamarosan megjelenik a TESz. záró kötete is, és akkor majd rendelkezésünkre fog állni egy, a TESz. alapján készített hasonló újabb összesítés. Ám természetesen naivság lenne azt hinni, hogy a TESz. megjelentével megszűnnek az etimológiai kutatások: épp ellenkezőleg, máris látható folyóirataink hasábjain, de eleve várható is volt, hogy e munka, valamint a Magyar szókészlet finnugor elemei, hatalmas lökést adott a további etimológiai kutatásoknak. (Valami hasonló történt a matematika egyes ágaival a számológépek megjelenését követően.) Tehát mire megjelenik a TESz., máris elavult lesz — sőt, remélhetjük, hogy még rohamosabban fog avulni, mint a SzófSz. a maga megjelenését közvetlenül követő, — finoman szólva — az etimológiai (s egyéb) kutatásoknak nem éppen kedvező évek során. Egyszerűen arról van szó, hogy, függetlenül az „avulástól”, érdemes, úgy látszik, összegezőleg lezárni etimológiai kutatásainknak a SzófSz. fémjelezte korszakát (ezt a „lezárást” közismert módon már VERMES STEFÁNIA megkezdte a maga idejében, vö. NyK. 51: 435), mint-

¹VÉRTES EDIT: Adalékok a magyar nyelv hangtani szerkezetéhez. — NyK. 54: 96—141 (1952); 55: 138—180 (1953), és kivált: 56: 215—266 (1954).

egy ezzel (is) fogadni a TESz. utolsó kötetének megjelentét, hogy azután a TESz. eredményeire is összegezve visszapillantsunk; esetleg összevetéseket eszközöljünk a SzófSz. és a TESz. nyújtotta kép között stb. — Ennek a cél-
nak és a használt forrásoknak felel meg feldolgozásunk pontossága (inkább: pontatlansága?) is. Nyilván eléggé viszonylagos anyagunk körülhatárolása: „az ÉrtSz.-ban foglalt kétségtelen tőszavak”, több okból is. Először: csak a mai állapotot tekintve is él még az ÉrtSz.-on kívül is nem egy tőszó — ha pontosan nem is tudjuk, mennyi (tájnyelv, szaknyelvek stb.). Továbbá: vajon az *arc* vitathatatlan tőszó-e (ma már nyilván az), vajon a *gítár* és a *citera* két külön tőszó-e (nyilván így igaz), vajon a *szinkronciklofazon* is a jelzett vitathatatlan magyar tőszavak közé tartozik-e (nem oda tartoz-
n é k : szerencsére nem szerepel az ÉrtSz.-ban) stb.? Annyit mindenesetre mondhatunk, hogy hatezer-egynéhány tőszót vizsgáltunk át (a pontos szám éppenséggel megvan a NytudÉrt. i. h. — de ne nézze meg az olvasó, nincs szüksége akkora pontosságra), ez közel harmincezer hangot jelentett (a pontos szám megvan itt az 1. táblázat végén). Természetes, hogy ez nem a magyar tőszavak összessége, mint ahogy az egyes etimológiai rétegek sem véglegesek, sőt még csak nem is tükrözik egészen pontosan a SzófSz. képviselte helyzetet, saját apparátusunk pontatlanságai (az emberi beavatkozás) miatt. Helyeztük úgy az olvasó, hogy az ÉrtSz.—SzófSz. forráspárból itt egy jó nagy mintát kap — ez a minta megközelíti a forráspár teljes anyagát, arányai is rend-
kívül közel állnak a forráspárban meglévő objektív arányokhoz. Bizonytalan-
nak talán a század százalékokat kell tekintenünk az 1. táblázaton; tulajdon-
képpen valószínűleg még ez is túlzott és csupán látszólagos pontosság a pre-
misszák említett gyengesége miatt. (A TESz. eredményei már pontosabbak
lesznek, de ott a premisszák is erősebbek.)

2. Ezek után közreadom az 1. táblázatot. Úgy kell olvasni, mint egyéb hasonló táblázatokat: A „0”, „8” és „9” jelentésére a bal szélső oszlopban
vö. fentebb, 1. alatt. Az egyes sorokban, illetőleg oszlopokban k ö z é p e n
elhelyezett álló kurrens számok abszolút értékek. Például: fgr. tőszavaink
összességében 175-ször fordult elő az *a* hang, 72-szer az *á*, 24-szer a *b* stb.;
illetőleg, egyetlen oszlop mentén haladva: az *a* 1431-szer fordult elő a SzófSz.-
ban nem szereplő tőszavak között, 175-ször a fgr. tőszavak között, 226-szor
a magyar belső keletkezésűek között stb. Az abszolút számok fölött jobbra
írt dőlt szám százalékos érték; 100,00 = az adott etimológiai rétegben elő-
forduló összes hang. Tehát, példánknl maradva: az összes fgr. tőszóban
előforduló összes hang 8,19%-a *a*, 3,37%-a *á*, 1,12%-a *b* stb. Az abszolút
számok alatt balra elhelyezett dőlt számok esetében viszont 100,00 = az adott
hang összes előfordulása a tőszókincs egészen belül. Tehát: az összes elő-
forduló *a*-k 46,75%-a a SzófSz.-ban nem szereplő tőszavakban áll, 5,72%-a
a fgr-ok között, 7,38%-a a magyar belsők között stb. A táblázat „szélein”
(a legelső sorban és a legutolsó oszlopban) a vizsgált tőszókincs e g é s z é r e
vonatkozó adatok találhatóak. Eszerint tehát a vizsgált tőszókincs egész hang-
állományának 10,40%-a volt *a*, 3,90%-a *á*, 2,60%-a *b* stb.; másrészt a vizs-
gált hangállomány 45,97%-át adták a SzófSz.-ban nem szereplő elemek,
7,26%-át a fgr-ok, 7,76%-át a magyar belsők stb. (Ezek az arányok eltérnek
a NytudÉrt. 58: 521 alatt közölt táblázatétól: azok az adatok a s z a v a k
számára vonatkoznak, míg az itt közöltek a h a n g o k é r a. Például: tőszó-
kincsünk 10,10%-a fgr., minden tizedik s z ó tőszókincsünkben fgr. — ám

	a	á	b	bb	c
Nincs benne a SzófSz-ban: e	10,57 1431 46,75	3,52 477 41,55	2,64 357 46,67	0,01 1 33,33	1,52 206 45,58
fgr	8,19 175 5,72	3,37 72 6,27	1,12 24 3,14	0,00 0 0,00	0,09 2 0,44
magyar	9,89 226 7,38	3,24 74 6,45	2,45 56 7,32	0,00 0 0,00	2,45 56 12,39
török	6,70 82 2,68	4,50 55 4,79	4,42 54 7,06	0,00 0 0,00	0,08 1 0,22
szláv	15,12 447 14,60	4,57 135 11,76	2,87 85 11,11	0,03 1 33,33	2,23 66 14,60
német	8,04 137 4,48	6,22 106 9,23	2,17 37 4,84	0,00 0 0,00	2,82 48 10,62
latin – görög	14,71 149 4,87	5,33 54 4,70	1,78 18 2,35	0,00 0 0,00	1,88 19 4,20
neolatin	13,22 55 1,80	5,53 23 2,00	2,16 9 1,18	0,00 0 0,00	1,92 8 1,77
egyéb	6,48 7 0,23	6,48 7 0,61	1,85 2 0,26	0,93 1 33,33	0,00 0 0,00
bizonytalan	8,66 352 11,50	3,57 145 12,63	3,03 123 16,08	0,00 0 0,00	1,13 46 10,18
Össz/hang	3061 10,40	1148 3,90	765 2,60	3 0,01	452 1,54

táblázat

ec	cs	cca	d	dd	dz	dda
0,09 12 70,59	1,03 140 31,75	0,07 9 75,00	2,43 329 46,53	0,01 1 25,00	0,01 2 33,33	0,00 0 0,00
0,00 0 0,00	1,17 25 5,67	0,00 0 0,00	1,82 39 5,52	0,05 1 25,00	0,05 1 16,67	0,00 0 0,00
0,04 1 5,88	2,01 46 10,43	0,09 2 16,67	2,54 58 8,20	0,00 0 0,00	0,04 1 16,67	0,00 0 0,00
0,00 0 0,00	3,27 40 9,07	0,00 0 0,00	1,80 22 3,11	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00
0,00 0 0,00	2,30 68 15,42	0,00 0 0,00	2,84 84 11,88	0,00 0 0,00	0,03 1 16,67	0,00 0 0,00
0,12 2 11,76	0,47 8 1,81	0,06 1 8,33	1,88 32 4,53	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00
0,10 1 5,88	0,59 6 1,36	0,00 0 0,00	2,17 22 3,11	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00
0,24 1 5,88	2,88 12 2,72	0,00 0 0,00	1,20 5 7,07	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00
0,00 0 0,00	0,00 0 0,00	0,00 0 0,00	4,63 5 7,07	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00
0,00 0 0,00	2,36 96 21,77	0,00 0 0,00	2,73 111 15,70	0,05 2 50,00	0,02 1 16,67	0,00 0 0,00
17 0,06	441 1,50	12 0,04	707 2,40	4 0,01	6 0,02	0 0,00

	dzs	ddzs	e	é	t
e	0,10 13 72,22	0,00 0 0,00	5,87 794 43,89	1,31 177 30,84	1,64 222 52,86
fgr	0,00 0 0,00	0,00 0 0,00	8,19 175 9,67	3,42 73 12,72	3,23 69 16,43
magyar	0,00 0 0,00	0,00 0 0,00	7,09 162 8,96	1,97 45 7,84	1,40 32 7,62
török	0,25 3 16,67	0,00 0 0,00	5,15 63 3,48	2,21 27 4,70	0,16 12 0,48
szláv	0,00 0 0,00	0,00 0 0,00	5,95 176 9,73	1,72 51 8,89	0,14 4 0,95
német	0,00 0 0,00	0,00 0 0,00	4,11 70 3,87	3,70 63 10,98	2,23 38 9,05
latin-görög	0,00 0 0,00	0,00 0 0,00	5,13 52 2,87	1,38 14 2,44	1,38 14 3,33
neolatin	0,24 1 5,56	0,00 0 0,00	3,37 14 0,77	2,64 11 1,92	1,68 7 1,67
egyéb	0,00 0 0,00	0,00 0 0,00	9,26 10 0,55	4,63 5 0,87	0,00 0 0,00
bizonytalan	0,02 1 5,56	0,00 0 0,00	7,21 293 16,20	2,66 108 18,82	0,79 32 7,62
Össz/hang	18 0,06	0 0,00	1809 6,14	574 1,95	420 1,43

táblázat

ff	g	gg	gy	gey	h	hh
0,09 12 80,00	2,14 289 41,29	0,01 1 14,29	0,18 24 14,81	0,01 1 33,33	1,45 196 41,00	0,00 0 0,00
0,00 0 0,00	2,34 50 7,14	0,09 2 28,57	1,31 28 17,28	0,05 1 33,33	3,60 77 16,11	0,00 0 0,00
0,13 3 20,00	2,06 47 6,71	0,00 0 0,00	0,88 20 12,35	0,00 0 0,00	2,58 59 12,34	0,00 0 0,00
0,00 0 0,00	1,96 24 3,43	0,00 0 0,00	1,96 24 14,81	0,00 0 0,00	0,74 9 1,88	0,00 0 0,00
0,00 0 0,00	2,43 72 10,29	0,00 0 0,00	0,14 4 2,47	0,00 0 0,00	1,12 33 6,90	0,00 0 0,00
0,00 0 0,00	2,46 42 6,00	0,00 0 0,00	0,12 2 1,24	0,00 0 0,00	1,41 24 5,02	0,00 0 0,00
0,00 0 0,00	2,07 21 3,00	0,10 1 14,29	0,59 6 3,70	0,00 0 0,00	0,39 4 0,84	0,00 0 0,00
0,00 0 0,00	2,64 11 1,57	0,00 0 0,00	0,48 2 1,24	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00
0,00 0 0,00	3,70 4 0,57	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00	0,93 1 0,21	0,00 0 0,00
0,00 0 0,00	3,45 140 20,00	0,07 3 42,86	1,28 52 32,10	0,02 1 33,33	1,85 75 15,69	0,00 0 0,00
15 0,05	700 2,38	7 0,02	162 0,55	3 0,01	478 1,62	0 0,00

	ch	i	t	j	jj
o	0,12 16 100,00	6,54 885 58,34	0,51 69 44,52	0,98 133 43,46	0,02 3 50,00
fgr	0,00 0 0,00	3,32 71 4,68	1,17 25 1,61	1,82 39 12,75	0,09 2 33,33
magyar	0,00 0 0,00	5,25 120 7,91	0,44 10 6,45	1,01 23 7,52	0,04 1 16,67
török	0,00 0 0,00	3,60 44 2,90	0,57 7 4,52	0,65 8 2,61	0,00 0 0,00
szláv	0,00 0 0,00	2,98 88 5,80	0,17 5 3,23	1,22 36 11,76	0,00 0 0,00
német	0,00 0 0,00	4,46 76 5,01	0,23 4 2,58	1,06 18 5,88	0,00 1 0,00
latin--görög	0,00 0 0,00	4,34 44 2,90	0,20 2 1,29	0,49 5 1,63	0,00 0 0,00
neolatin	0,00 0 0,00	4,57 19 1,25	0,48 2 1,29	0,72 3 0,98	0,00 00 0,00
egyéb	0,00 0 0,00	1,85 2 0,13	1,85 2 1,29	1,85 2 0,65	0,00 0 0,00
bizonytalan	0,00 0 0,00	4,13 168 11,07	0,71 29 18,71	0,96 39 12,75	0,00 0 0,00
Össz/hang	16 0,05	1517 5,15	155 0,53	306 1,04	6 0,02

táblázat

k	kk		ll	ly	lly	m
5,08 688 41,60	0,27 36 73,47	4,63 627 49,25	0,58 79 54,86	0,28 38 19,90	0,00 0 0,00	3,59 486 48,36
5,76 123 7,44	0,00 0 0,00	5,66 121 9,51	0,80 17 11,81	0,89 19 9,95	0,00 0 0,00	2,90 62 6,17
5,21 119 7,19	0,13 3 6,12	3,85 88 6,91	0,31 7 4,86	0,61 14 7,33	0,00 0 0,00	4,16 95 9,45
9,16 112 6,77	0,08 1 2,04	2,78 34 2,67	0,33 4 2,78	0,82 10 5,24	0,00 0 0,00	1,96 24 2,39
7,07 209 12,64	0,00 0 0,00	3,72 110 8,64	0,20 6 4,17	0,91 27 14,14	0,03 1 50,00	2,67 79 7,86
4,63 79 4,78	0,35 6 12,24	4,46 76 5,97	1,00 17 11,81	1,64 28 14,66	0,00 0 0,00	3,23 55 5,47
4,15 42 2,54	0,00 0 0,00	4,94 50 3,93	0,20 2 1,39	0,99 10 5,24	0,10 1 50,00	5,13 52 5,17
3,61 15 0,91	0,00 0 0,00	4,57 19 1,49	0,24 1 0,69	0,72 3 1,57	0,00 0 0,00	6,01 25 2,49
3,70 4 0,24	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00	5,56 6 0,60
6,47 263 15,90	0,07 3 6,12	3,68 148 11,63	0,27 11 7,64	1,03 42 21,99	0,00 0 0,00	2,98 121 12,04
1654 5,62	49 0,17	1273 4,32	144 0,49	191 0,65	2 0,01	1005 3,41

	mm	n	nn	ny	nny
o	0,06 8 100,00	5,24 709 55,52	0,10 13 50,00	0,38 52 18,84	0,01 2 25,00
fgr	0,00 0 0,00	2,48 53 4,15	0,09 2 7,69	2,01 43 15,58	0,19 4 50,00
magyar	0,00 0 0,00	3,59 82 6,42	0,09 2 7,69	1,49 34 12,32	0,04 1 12,50
török	0,00 0 0,00	3,19 39 3,05	0,00 0 0,00	1,80 22 7,97	0,00 0 0,00
szláv	0,00 0 0,00	3,72 110 8,61	0,10 3 11,54	1,42 42 15,22	0,03 1 12,50
német	0,00 0 0,00	4,69 80 6,26	0,06 1 3,85	0,65 11 3,99	0,00 0 0,00
latin - görög	0,00 0 0,00	5,03 51 3,99	0,39 4 15,38	0,20 2 0,72	0,00 0 0,00
neolatin	0,00 0 0,00	4,81 20 1,57	0,00 0 0,00	1,68 7 2,54	0,00 0 0,00
egyéb	0,00 0 0,00	2,78 3 0,23	0,00 0 0,00	0,93 1 0,36	0,00 0 0,00
bizonytalan	0,00 0 0,00	3,20 130 10,18	0,02 1 3,85	1,53 62 22,46	0,00 0 0,00
Össz/hang	8 0,03	1277 4,34	26 0,09	276 0,96	8 0,03

táblázat

o	ó	ö	ő	p	pp	q
5,02 679 45,75	2,11 285 54,08	0,48 65 21,38	0,27 37 27,82	3,42 463 49,05	0,13 18 69,23	0,00 0 0,00
3,98 85 5,73	1,31 28 5,31	1,82 39 12,83	1,17 25 18,80	1,36 29 3,07	0,09 2 7,69	0,00 0 0,00
5,21 119 8,02	1,27 29 5,50	1,49 34 11,18	0,70 16 12,03	3,37 77 8,16	0,13 3 11,54	0,00 0 0,00
5,15 63 4,25	2,13 26 4,93	3,68 45 14,80	1,55 19 14,29	1,80 22 2,33	0,00 0 0,00	0,00 0 0,00
5,99 177 11,93	1,25 37 7,02	1,01 30 9,87	0,10 3 2,26	4,26 126 13,35	0,03 1 3,85	0,00 0 0,00
4,99 85 5,73	1,64 28 5,31	1,11 19 6,25	0,18 3 2,26	4,05 69 7,31	0,12 2 7,69	0,00 0 0,00
5,43 55 3,71	2,17 22 4,17	0,20 2 0,66	0,00 0 0,00	4,24 43 4,56	0,00 0 0,00	0,00 0 0,00
6,01 25 1,68	1,44 6 1,14	0,00 0 0,00	0,00 0 0,00	4,09 17 1,80	0,00 0 0,00	0,00 0 0,00
2,78 3 0,20	0,93 1 0,19	0,00 0 0,00	0,93 1 0,75	0,93 1 0,11	0,00 0 0,00	0,00 0 0,00
4,75 193 13,01	1,60 65 12,33	1,72 70 23,03	0,71 29 21,80	2,39 97 10,28	0,00 0 0,00	0,00 0 0,00
1484 5,04	527 1,79	304 1,03	133 0,45	944 3,21	26 0,09	0 0,00

	r	rr	s	ss	sz
o	7,22 977 44,53	0,03 4 50,00	2,02 274 43,29	0,10 13 72,22	2,56 347 46,83
fg	5,38 115 5,24	0,14 3 37,50	2,20 47 7,42	0,05 1 5,56	3,51 75 10,12
magyar	7,13 163 7,43	0,04 1 12,50	1,66 38 6,00	0,04 1 5,56	2,54 58 7,83
török	9,32 114 5,20	0,00 0 0,00	2,45 30 4,74	0,00 0 0,00	2,70 33 4,45
szláv	7,17 212 9,66	0,00 0 0,00	1,42 42 6,64	0,00 0 0,00	3,42 101 13,63
német	10,44 178 8,11	0,00 0 0,00	3,17 54 8,53	0,12 2 11,11	0,94 16 2,16
latin – görög	8,29 84 3,83	0,00 0 0,00	3,95 40 6,32	0,10 1 5,56	0,39 4 0,54
neolatin	8,65 36 1,64	0,00 0 0,00	2,40 10 1,58	0,00 0 0,00	1,20 5 0,67
egyéb	12,04 13 0,59	0,00 0 0,00	2,78 3 0,47	0,00 0 0,00	2,78 3 0,40
bizonytalan	7,43 302 13,76	0,00 0 0,00	2,34 95 15,01	0,00 0 0,00	2,44 99 13,36
Össz/hang	2194 7,45	8 0,03	633 2,15	18 0,06	741 2,52

táblázat

szz	t	tt	ty	tty	u	ú
0,19 26 66,67	5,26 712 50,07	0,61 82 90,99	0,21 28 37,84	0,03 4 23,53	3,07 416 59,60	0,30 41 28,28
0,23 5 12,82	4,16 89 6,26	0,09 2 2,18	0,05 1 1,35	0,05 1 5,88	1,08 23 3,30	1,45 31 21,38
0,09 2 5,13	3,54 81 5,70	0,18 4 4,40	0,39 9 12,16	0,26 6 35,29	1,79 41 5,87	0,61 14 9,66
0,08 1 2,56	4,91 60 4,22	0,00 0 0,00	0,16 2 2,70	0,08 1 5,88	1,39 17 2,44	1,64 20 13,79
0,00 0 0,00	3,75 111 7,81	0,00 0 0,00	0,24 7 9,46	0,14 4 23,53	1,76 52 7,45	0,17 5 3,45
0,12 2 5,13	5,51 94 6,61	0,12 2 2,18	0,12 2 2,70	0,00 0 0,00	2,11 36 5,16	0,23 4 2,76
0,00 0 0,00	5,33 54 3,80	0,10 1 1,10	0,59 6 8,11	0,00 0 0,00	2,57 26 3,72	0,00 0 0,00
0,00 0 0,00	6,01 25 1,76	0,00 0 0,00	0,48 2 2,70	0,00 0 0,00	1,92 8 1,15	0,24 1 0,69
0,93 1 2,56	5,56 6 0,42	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00	0,00 0 0,00
0,05 2 5,13	4,68 190 13,36	0,00 0 0,00	0,42 17 22,97	0,02 1 5,88	1,94 79 11,32	0,71 29 20,00
39 0,13	1422 4,83	91 0,31	74 0,25	17 0,06	698 2,37	145 0,49

	ü	ű	v, w	vv	x
ø	0,43 58 38,93	0,10 14 20,00	1,39 189 38,89	0,00 0 0,00	0,18 24 100,00
fgr	0,94 20 13,42	0,61 13 18,57	3,09 66 13,58	0,00 0 0,00	0,00 0 0,00
magyar	0,53 12 8,05	0,22 15 7,14	2,14 49 10,08	0,00 0 0,00	0,00 0 0,00
török	0,98 12 8,05	1,55 19 27,14	1,06 13 2,67	0,00 0 0,00	0,00 0 0,00
szláv	0,17 5 3,36	0,00 0 0,00	1,45 43 8,85	0,00 0 0,00	0,00 0 0,00
német	0,12 2 1,34	0,00 0 0,00	0,94 16 3,29	0,00 0 0,00	0,00 0 0,00
latin - görög	0,20 2 1,34	0,00 0 0,00	1,09 11 2,26	0,00 0 0,00	0,00 0 0,00
neolatin	0,48 2 1,34	0,00 0 0,00	0,48 2 0,41	0,00 0 0,00	0,00 0 0,00
egyéb	0,93 1 0,67	0,93 1 1,43	5,56 6 1,23	0,00 0 0,00	0,00 0 0,00
bizonytalan	0,86 35 23,49	0,44 18 25,71	2,24 91 18,72	0,00 10 0,00	0,00 0 0,00
Össz/hang	149 0,51	70 0,24	486 1,65	0 0,00	24 0,08

táblázat

y	z	zz	zs	zsz	össz/etim.	%
0,03 4 100,00	1,16 157 46,31	0,01 2 50,00	0,55 75 50,00	0,00 0 0,00	13532	45,97
0,00 0 0,00	1,82 39 11,50	0,05 1 25,00	0,09 2 1,33	0,00 0 0,00	2137	7,26
0,00 0 0,00	1,27 29 8,55	0,00 0 0,00	0,31 7 4,67	0,00 0 0,00	2285	7,76
0,00 0 0,00	1,06 13 3,83	0,00 0 0,00	0,16 2 1,33	0,00 0 0,00	1223	4,15
0,00 0 0,00	1,18 35 10,32	0,00 0 0,00	0,74 22 14,67	0,00 0 0,00	2957	10,04
0,00 0 0,00	0,82 14 4,13	0,00 0 0,00	0,94 16 10,67	0,00 0 0,00	1705	5,79
0,00 0 0,00	0,20 2 0,59	0,00 0 0,00	1,38 14 9,33	0,00 0 0,00	1013	3,44
0,00 0 0,00	0,24 1 0,29	0,00 0 0,00	0,72 3 2,00	0,00 0 0,00	416	1,41
0,00 0 0,00	4,63 5 1,47	0,00 0 0,00	0,93 1 0,67	0,00 0 0,00	108	0,37
0,00 0 0,00	1,08 44 12,98	0,02 1 25,00	0,20 8 5,33	0,00 0 0,00	4063	13,80
4 0,01	339 1,15	4 0,01	150 0,51	0 0,00	29439	100,00

ezen ősi elemek természetesen rövidebbek, egyenként általában kevesebb hangból állnak, mint fiatalabb lexemáink, ezért az itt közölt 7,26 %-os arány a hangok vonatkozásában.)

A táblázat felvet néhány többé-kevésbé technikai jellegű kérdést, ezekre szeretnék az alábbiakban röviden kitérni.

Az egyes „vegyes” etimológiai rétegekkel kapcsolatosan a következő a helyzet. A „0”-sokkal részletesen foglalkoztunk (NyK. 71: 129 kk.): majdnem fele részben nemzetközi vándorszavak meg átlátszó német kölcsönzések ezek. — Korábban (NyK. 70: 205) utaltunk arra, hogy az „egyéb” kategóriából, összetételének tarkaságánál fogva, semmiféle következtetést nem vonhatunk le. Egészítsük ki itt az akkor mondottakat a következőkkel. Az ide vont 25 szó jó részét a SzófSz. által egységesen „iráni”-ként minősített elemek alkotják (ahol a TESz. I—II. kötete lényegében ugyanígy nyilatkozik, ott zárójelben semmi sem áll a szó mögött): *bűz*, *büszke* (TESz. e kettőről, illetőleg a *bűz* közös töről: vitatott, 1. iráni, 2. magyar belső; megjegyzendő, már a SzófSz. is így nyilatkozik: „t alán iráni”, mi azonban az ÉrtSz. gépi feldolgozásakor az olyan s ahhoz hasonló módosító megjegyzéseket, mint „talán”, „valószínűleg”, „esetleg” stb. nem tudtuk figyelembe venni, s gyakran úgy tekintettük, mintha ott sem állnának — a TESz. gépi feldolgozását viszont már eleve úgy készítettük elő, hogy mind a történeti [első előfordulás stb.], mind az etimológiai adatok mellett egy külön, tíz érték valamelyikét felvehető oszlop áll az efféle finom filológiai megjegyzések rögzítésére, vö. MNy. 65: 148, 149), *egész* (SzófSz.: „talán”, TESz. vitatott: talán iráni, közelebről alán), *vám*, *vár²*, *város*, *vásár*; SzófSz. „alán”: *asszony*, *nemez*, *verem*, *vért*; SzófSz. „árja”: *tíz*. Ez ősi elemek mellett vannak azután itt jóval frissebb „egyéb” kölcsönzések: cigány (*dádé*, *rajkó*), lapp (*mintá*), héber (*rabbi*), angol (*sport*, *sztrájk*), arab (*tózsér*). — A „bizonytalan”-ok nagy része ősi elem, eredete épp ezért vitatott. — Végül, valaki talán vegyesnek tekinti (vagy legalábbis nehezen megfoghatónak) a magyar belső keletkezésű elemeket: részben felsoroltuk korábban (NyK. 70: 205—206).

A biztos etimológiájúak között is szemet szúrhat egy-két dolog. Inkább csupán példaként arra, hogy milyen kérdések vetődhetnek itt fel és milyen jellegű lehet a megoldásuk, utalok a következőre. A fgr. eredetű tőszavaink között két ízben is szerepel a *c* hang. Hogyan lehet ez? Íme: *arc*, *hurcol* (ez utóbbi mint származék a *hord* tőből — *hurc*- tő nem lévén, az *-ol* elemet nem foghattuk fel képzőnek, így ez a szó bekerült a tőszavak közé). És a két *zs*, ugyancsak a fgr. eredetűek között? Íme: *darázs* (SzófSz.: *v a l ó s z í n ű l e g* fgr. eredetű; TESz.: „Ismeretlen eredetű. ... A szóvégi hang eredetileg *s* lehetett...”), *morzsa* (SzófSz.: ... végső elemzésben a ... *mar* ige származéke; TESz.: vitatott: 1. < *mar¹*; 2. hangut.-hangfestő).

3. Az egyes etimológiai rétegek jellemzését ezek után több oldalról is meg lehet közelíteni.

a) Az anyag felkínálkozó, számos lehetséges nyelvészeti csoportosításából ragadjunk ki néhányat.

Az összes hang hány százaléka magánhangzó? Erre a kérdésre ad választ a 2. táblázat. Az eredmény első pillanatra talán kissé mehökkentő: honnan a latin-görög és kivált a szláv eredetű elemek mai alakjának ez a nagy — a finnugorokénál és a magyar belsőknél mindenestre nagyobb — magánhangzósága? Kissé képletesen, de talán nem hami-

san szólva azt kell mondanunk, hogy nyelvünket épp a magánhangzó— mássalhangzó-arány tekintetében, a dús mássalhangzócsoportok vonatkozásában vehemens és tömeges támadás érte a Kárpát-medencébe kerültekor (a korábbi, török, iráni stb. kölcsönzések e szempontból nem tették ennyire próbára nyelvünk — nyelveink — ősének rendszerét); erre nyelvünk egész-

2. táblázat

Sorszám	Etimológiai réteg	Valamennyi hang hány százaléka magánhangzó
1	latin – görög	41,658
2	szláv	40,954
3	török	40,229
4	finnugor	40,009
5	neolatin	39,904
6	bizonytalan	39,700
7	magyar	39,694
8	német	37,126

séges rendszere túlzott reakcióval válaszolt, és így mára elértük azt, hogy a szláv eredetű elemek a maguk összességében aránylag több magánhangzót tartalmaznak, mint szókincsünk ősbib rétegei. Bármilyen szépen hangozzék is azonban a „túlzott reakció”, csupán biomorf hasonlat marad, ha a folyamat konkrét nyelvi lejátszódását nem vesszük figyelembe. E folyamat okai eléggé ismertek (az én figyelmemet erre BENKŐ LORÁND és KÁLMÁN BÉLA hívta fel személyes beszélgetés során). Az ómagyar kor hajnalára magunkkal hozott elemek, függetlenül attól, hogy ősi örökségként őriztük-e meg őket, vagy valamilyen ősi kölcsönzés eredményeképpen, jelentős részükben CVCV felépítésűek voltak; ezekről az alakvégi V közismert módon hamar lekopott. Míg tehát a CVCV felépítés korában a mássalhangzó—magánhangzó arány durván szólva 50—50 százalékos volt, addig a CVC korra a magánhangzók aránya ezen elemeink között 33 százalékra esett vissza. A szláv kölcsönzések hosszabbak voltak, meg más felépítésűek is: a *mezsgye*-félékben a C : V arány eleve 60 : 40 százalék volt (tehát jobb, mint az ősbib elemekben); másokban némi átalakítás után ilyenné lett (*asztag*, *asztal* stb.); olykor még egy-egy magánhangzó kidobására is futotta ezen a jó arányon belül (*málna*, *medence*). Különösen erős volt nyelvünk reakciója szókezdő mássalhangzócsoport esetében, kivált, ha eleve hosszabb (kéttagú) elemről volt szó, vö. *berkenye*, *borona*¹, *borona*², *borostyán*¹, *cseresznye* stb. (C : V = 57 : 43). Alább, a 4. pontban látni fogjuk, hogy bizonyos korreláció van a szavak szótári alakjának hossza és magánhangzóssága között: rövidebb szóban aránylag több mássalhangzó-megfér, és megfordítva: a hosszabb alak eleve nagyobb magánhangzó-arányt indikál. Itt, a tulajdonképpen tőszavak etimológiai rétegeiben ez a szabályszerűség még nem érvényesül kellő zavartalansággal. Íme, az egyes etimológiai rétegekre jellemző átlagos szóhossz (a zárójelben közölt első szám — hangokban, a második — szótagokban mért hossz): 1. lat.-gör. (5,63, 2,34), 2. neolat. (5,47, 2,18), 3. szláv (5,21, 2,13), 4. német (5,17, 1,92), 5. magyar (4,60, 1,82), 6. bizonytalan (4,58, 1,82), 7. török (4,43, 1,78), 8. fgr. (3,50,

1,40). Jól látható, hogy ez a sorrend nem egyezik meg a 2. táblázaton adódóval: a neolatin elemek hosszabbak ugyan, mégis elég kevés bennük a magánhangzó; másrészt a törökök meg a finnugorok rövidek ugyan, de ehhez képest sok bennük a magánhangzó. Érdekes a helyzet a német jövevényszavakkal, ezt a következő módon tudjuk érzékeltetni. Alább fogunk közölni adatokat egész tőszókincsünk magánhangzóságára, átlagos hosszára stb. vonatkozóan: azokat az adatokat úgy fogjuk megkapni, hogy az egyes tőszavakban tapasztaltak számtani közepét vesszük. Egy ilyen középértékre természetes módon erősen hat az, hogy milyen fajta szóból volt több az anyagban: esetünkben például mindenféle átlag a „SzófSz.-ban nem szereplő” és a „SzófSz.-ban bizonytalanak minősített” elemekhez fog közelebb állni, tekintettel arra, hogy e két rétegből volt a legtöbb (az egész hanganyag közel 60 százaléka). Most azonban tekintsünk el az egyes rétegek nagyságától, szerepeltessünk minden egyes réteget egyenlő súllyal. Akkor a 2. táblázat, illetőleg az itt felsorolt értékek alapján azt kapjuk, hogy a vizsgált rétegek átlagos szóhossza 4,82 hang/szó vagy 1,92 szótag/szó, a magánhangzók az összes hang 39,909%-át teszik ki. Nos, látható, hogy a német réteg elemei szótagokban mérve éppen átlagos hosszúságúak, ám a magánhangzók aránya bennük a legkisebb, tehát jóval az átlag alatt van (éppen ezért természetesen hangokban mért hosszuk némileg magasabb is az átlagosnál). Külön vizsgálat tárgyát kellene képeznie annak, hogy míg nyelvünk rendszere, mint láttuk, oly sikeresen felvette a harcot a szláv mássalhangzó-torlódásokkal, miért s hogyan blokkolt le a beözönlő német kölcsönzésekkel szemben. Vajon arról van-e szó, hogy a német kölcsönzések szerencsésebb szótagstruktúrájuk (például: nem tartalmaznak annyi *s z ó k e z d ő* mássalhangzó-csoportot, mint a szlávok)? Vagy: a német elemek tömegükben már később jöttek, amikor nyelvünk rendszerének korábbi vehemens ellenakciója lecsitult, kezdett hozzáidomulni egyes mássalhangzó-csoportokhoz? Vagy: a német mássalhangzó-csoportok elfogadhatóbbak voltak nyelvünk rendszere számára, mint a szlávok? Vagy: társadalmi okok játszottak közre (durván szólva: a szlávok elnyomottak voltak, a németek a legrosszabb esetben viszonylag független iparosok)?

A magánhangzók hány százaléka mély és hány magas? Erre a kérdésre a választ a 3. táblázat tartalmazza, mégpedig kétféle módon: (i) az *i, í* (és a *brandy y*-vel jelölt eleme) nélkül; (ii) az *i*-kkel együtt. Történetileg nyilván az (i) ad hívebb képet, mert a (ii) esetében óhatatlanul az összes *i* a „magasak” közé kerül. („Óhatatlanul”, ha nem akarunk a mai szinkrón anyagba történeti elemeket becsempészni, és így az *ír¹* ige *í*-jét valahogy az *ír²* fn-mn [nemzetségnév] *í*-jétől megkülönböztetni.) A sorrendet éppen ezért az (i)-nek megfelelően alakítottuk ki. Egyébként látható, hogy nincs elvi eltérés az (i) és a (ii) szerinti eredmények között. Durván szólva: tőszókincsünk legősibb rétegeiben még közel egyenlő arányban szerepelnek a mély és a magas magánhangzók; ahogy azután haladunk korunk felé előre az időben, a fiatalabb s fiatalabb rétegekben egyre inkább többségbe kerülnek a mély magánhangzók. A valóságban még sarkítottabb a helyzet: tekintettel arra, hogy az *-izmus*-t, az *-istá*-t magyar képzőnek vettük, az e véggel ellátott elemek nem szerepelnek a tőszavak között; sőt, minden olyan eset is, ahol a szóvégen felismerhető gyakori idegen szóvég áll (mint: *-or*, *-tor*, *-us*, *-um* stb.) kikerült a szigorú értelemben vett tőszavak köréből (ezekről l. alább, 4.), tehát a 3. táblázat ezeket az adatokat sem tartalmazza, jóllehet világos, hogy ezek mind vagy majdnem mind a mély magánhangzók arányát

emelték volna. Jellemző a „bizonytalan” etimológiájúak helyzete: rögtön legősibb elemeink után következnek e szempontból — főntebb említettük, hogy valóban ősi elemek teszik ki a zömüket. Ugyancsak megnyugtató a fentebb szintén érintett „0”-sok fiatal, a németek és a szlávok közötti, helye, valamint a latin-görögök és a neolatinok „közelsége” egymáshoz. — Egyébként, bár ez távolról sem hangstatisztikai kérdés, hozhatok adatokat arra vonatkozóan, hogy a szigorú értelemben vett tőszavak zömét kitevő névszók (főnevek, főnév-melléknevek, melléknév-főnevek stb.) hogyan illeszkednek. Mély hangrendű toldalékokat kap (100 = az adott etimológiájú szavak): 1. fgr. 55, 2. török 59, 3. bizonytalan 63, 4. magyar 66, 5. német 76, 6. szláv 82, 7. neolatin 83, 8. latin-görög 92. Ez a sorrend teljesen megfelel a 3. táblázatbelinek: az *i*-k (és, egyes esetekben: a zárt *ē*-k meg az *é*-k) a hangállományban nem zavarták meg a tulajdonképpeni (a történetnek megfelelő, ma: a ragozási rendszerben tükröződő) szóstatisztikai képet — ahol a hangállomány egészében több a mély magánhangzó, ott az illeszkedés szemszögéből is több lexikális elem minősült mélynek. Az is jól látható, hogy az időben előre haladva viszonylag kevesebb mély magánhangzó a hangállományban biztosítani tudja egyre több lexéma mély toldalékolását: a fgr. állományban például még 55% a mélyek aránya — toldalékolás szempontjából is 55% kap mély toldalékot, ugyanígy a török elemek között (59—59%) —, ugyanakkor a latin-görög állományban a magánhangzóknak 81%-a mély, a szavak 92%-a mély toldalékolású. A mély magánhangzók ezen nagyobb aktivitása (potenciális vagy alaktani súlya) természetesen következik ma érvényesülő illeszkedési törvényünkből.

3. táblázat

Sorszám	Etimológiai réteg	A magánhangzók hány %-a mély	
		(i)	(ii)
1	finnugor	54,55	48,42
2	török	58,71	52,71
3	bizonytalan	60,95	53,50
4	magyar	64,74	55,46
5	német	71,61	62,56
6	„0”	71,41	61,32
7	szláv	76,30	70,44
8—9	latin—görög	81,38	72,51
	neolatin	81,38	71,08

Az 1. táblázat adatai módot adnak arra, hogy nyelvészetileg érdekes egyéb csoportokat is kiragadjunk belőle és azok egyedeinek eloszlását vizsgáljuk. Megnéztük például a kvantitás kérdést. A mássalhangzók körében nem találtunk túlságosan érdekes képet: a mássalhangzók 1—4%-a bizonyult hosszúnak. Az egyes rétegek sorrendje a hosszú mássalhangzók aránya szerint (100,00 = az összes mássalhangzó az adott rétegben): 1. fgr. 3,4, 2. német 3,3, 3. magyar 2,7, 4. latin-görög 1,9, 5. szláv 1,0, 6. török 1,0, 7. neolatin 1,0. A vizsgált tőszóállomány összes mássalhangzójának mindössze 2,9%-át alkották a hosszú mássalhangzók, azt kell tehát mondanunk, hogy tőszókincseink egészében felettébb ritkák ezek az elemek. Ha mégis szótári állományunk vagy valamely szöveg vizsgálata során viszonylag több hosszú

4. táblázat

Etimológiai réteg	Mély		Össz mély	Magas		Össz magas	Össz	
	hosszú	rövid		hosszú	rövid		hosszú	rövid
„0” 5429	24,12 803 14,79	75,88 2526 46,53	3329	14,14 297 5,47	85,86 1803 33,21	2100	1100 20,26	4329 79,74
fgr. 855	31,64 131 15,32	68,36 283 33,10	414	30,84 136 15,91	69,16 305 35,67	441	267 31,23	588 68,77
magyar 907	23,26 117 12,90	76,74 386 42,56	503	18,81 76 8,38	81,19 328 36,16	404	193 21,28	714 73,72
török 499	38,40 101 20,24	61,60 162 32,46	263	30,51 72 14,43	69,49 164 32,87	236	173 34,67	326 65,33
szláv 1211	20,75 177 14,62	79,25 676 55,82	853	16,48 59 4,87	83,52 299 24,69	358	236 19,49	975 80,51
német 633	34,85 138 21,80	65,15 258 40,76	396	29,54 70 11,06	70,46 167 26,38	237	208 32,86	425 67,14
lat. – gör. 422	24,84 76 18,01	75,16 230 54,50	306	13,79 16 3,79	86,21 100 23,70	116	92 21,80	330 73,20
neolatin 166	25,42 30 18,07	74,58 88 53,01	118	27,08 13 7,83	72,92 35 21,08	48	43 25,90	123 74,10
egyéb 40	44,44 8 20,00	55,56 10 25,00	18	40,91 9 22,50	59,09 13 32,50	22	17 42,50	23 57,50
„9” 1613	27,69 239 14,82	72,31 624 38,69	863	24,53 184 11,41	75,47 566 35,09	750	423 26,22	1190 73,78

mássalhangzóra bukkanunk, akkor ezt az adott rétegre (szócsoporra, stílusra stb.) erősen jellemzőnek kell majd tekintenünk. — A magánhangzók körében a helyzetet összegezően a 4. táblázat rögzíti: a jobbra felül elhelyezett, dőlt szedésű százalékos értékek itt az adott etimológiai réteg adott hangrendű magánhangzóinak összességére vonatkoznak, míg a balra lent elhelyezett értékek esetében 100 = az adott etimológiai réteg összes magánhangzója

abszolút értékben emlékeztetőül megadva az etimológiai réteg megnevezése alatt). Némileg áttekinthetőbb és értékelhetőbb képet kapunk, ha az egyes etimológiai rétegeket sorrendbe állítjuk aszerint, hogy magánhangzó-állományuk hány százaléka hosszú: 1. török 35, 2. német 33, 3. fgr. 31, 4. neolatin 26, 5. latin-görög 22, 6. magyar 21, 7. szláv 19. Ehhez hasonló sorrendet kapunk akkor is, ha külön az egyes etimológiai csoportok mély magánhangzóit tekintjük száznak, ebből hosszú: 1. török 38, 2. német 35, 3. fgr. 32, 4. neolatin 25, 5. latin-görög 25, 6. magyar 23, 7. szláv 21; illetőleg az egyes csoportok magas magánhangzóit, ebből hosszú: 1. fgr. 31, 2. török 31, 3. német 30, 4. neolatin 27, 5. magyar 19, 6. szláv 16, 7. latin-görög 14. A török—német—ünnugor eredetű szavak mindenképpen olyan csoportot alkotnak, ahol durván szólva minden harmadik magánhangzó hosszú (ezen belül: a hosszúak között viszonylag több, a magasak között viszonylag kevesebb a hosszúak aránya), másrészt közel vannak egymáshoz a latin-görög, magyar és szláv eredetűek: itt csupán minden negyedik-ötödik magánhangzó hosszú, sőt a magasak között csupán minden hatodik-nyolcadik. Majdnem minden egyes magánhangzó-párra jellemző minden egyes rétegben, hogy a megfelelő rövid magánhangzó gyakoribb, mint a hosszú. Kivétel: az *u* és az *ú*, valamint az *ü* és az *ű* a török csoporton belül; az *u* és az *ú* a fgr.-on belül: e párok esetében a hosszú magánhangzó gyakoribb, mint a rövid (pontos számokat l. I. táblázaton).

Végső soron véleményünk szerint érdemes lenne az itt közölt adatokat megkülönböztető jegyekre (featurekre) bontva vizsgálni: a magánhangzóság—mássalhangzóság, a palatalitás—velaritás, hosszúság—rövidség ebben az összefüggésben csupán kiragadott példák voltak egy rendszeres feature-elemzésből. Lehet, hogy egyes featurek kvantitatív-diakrón vizsgálatára egészen triviális eredményeket hoz (nem is beszélhetünk egészen diakróniáról, hiszen a szavakat *m a i m a g y a r* hangállományuk szerint vizsgáljuk); másutt talán meglepőbbeket. Lehet, hogy eredményeink nagyobb része, talán éppenséggel mindegyike triviális a magyar nyelvtörténet ismerőinek; számunkra a szláv eredetűek mai alakjának viszonylag magas magánhangzó-aránya, valamint szókincsünk fokozatos mélyülése váratlan volt. (A mélyülésnek messzemenő rendszerbeli következményei kezdtek jelentkezni nézetünk szerint — ezzel másutt kívánunk foglalkozni.)

b) A másik fő kutatási—feldolgozási lehetőség: tekintsünk el attól, hogy itt hangokról (fonémákról, feature-mátrixokról stb.) van szó és rendszerezük a kapott eredményeket tisztán mennyiségi szempontból, bizonyos elemi matematikai műveletek segítségével.

Tekintsük meg az 1. ábrát. Ott az 1. táblázat egyes sorait átrendeztük előfordulási gyakoriságok szerint, és ezen átrendezésnek az „elejét” (a leggyakoribb néhányszor tíz elemet) érzékeltettük szokásos módon körgrafikonainkon. A ritkább elemeket tisztán technikai szempontból hagytuk el: azokat már nem lehetne jól ábrázolni, mi pedig ezekkel az ábrákkal lényeges dolgot akarunk illusztrálni. Figyeljük meg, mennyire egyenletesen oszlanak meg az egyes elemek az egyes csoportokban. Minél több elem esik egy adott hányadot — egy negyedre, felet, 87%-ot stb. — szimbolizáló területre (vagy, ami ugyanez: minél kevesebb elem marad az „ábrázolhatatlanul kicsik” fertályában), annál egyenletesebbnek kell tekintenünk az eloszlást. Az 5. táblázat a rajzoknál pontosabban (csak nem olyan szemléletesen) érzékelteti, mit akarunk mondani: azon a táblázaton már eleve az eloszlás egyenletessé-

gének sorrendjében állítottuk fel az egyes rétegeket, a legegyszerűsebbel kezdve és haladva a kevésbé egyszerűen megosztlók felé. Ismét egy közelítő kronológiai sorrendet kaptunk az egyes rétegekre vonatkozóan: elég közel állnak egymáshoz egyrészt a fgr., magyar, török elemek, mint egyszerűbben megosztlók, másrészt a szláv, neolatin, latin—görög lexémák, mint kevésbé egyszerűen megosztlók. Ismét szépen megtalálják a maguk helyét a viszonylag ősi „bizonytalanok” (1-a), meg az újabb „0”-sok (3-a). Próbáljuk meg interpretálni a kapott eredményt. Mit is jelent az, hangokról szól-

5. táblázat

Sorszám	Etimológiai réteg	Hány különböző hang teszi ki az összeállomány			
		25	50	75	87
		százalékát?			
1	finnugor	3,5	9,5	18,0	25,0
2	magyar	3,2	8,5	17,5	23,5
3	török	3,0	8,2	17,0	23,0
4	német	3,1	8,1	15,1	21,0
5	szláv	2,4	7,0	14,5	20,0
6	neolatin	2,2	7,2	13,5	19,0
7	latin—görög	2,2	7,2	12,5	18,5
(1—a	bizonytalan	3,2	9,2	18,0	24,0
3—a	„0”	3,1	8,0	15,0	21,5)

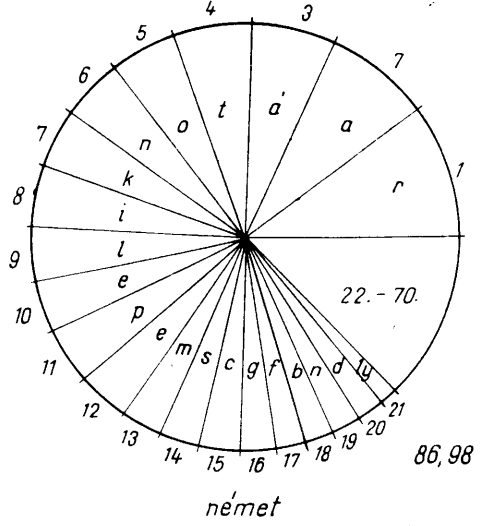
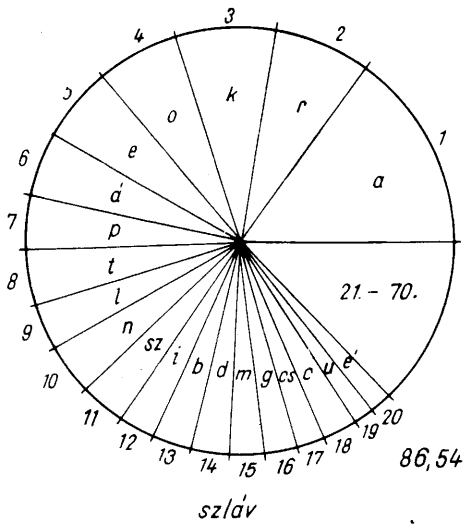
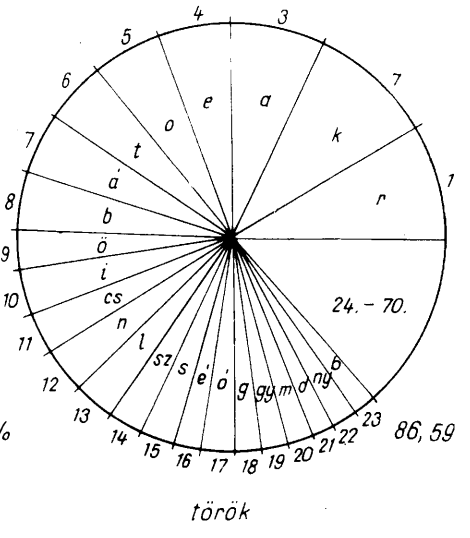
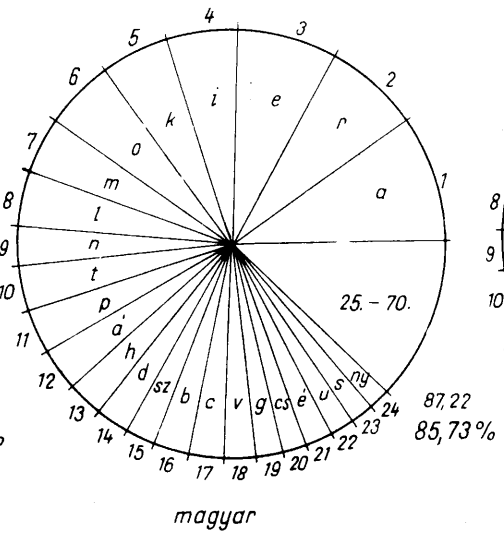
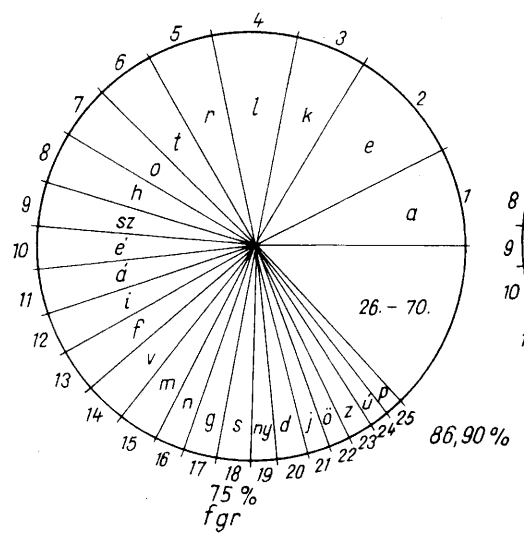
ván, hogy „egyszerűsebb eloszlás”. Nevezhetnénk ezt (szótárt vizsgálva) „otthonosságnak” is: minél egyszerűsebben eloszlanak valamely réteg lexémái a rendelkezésre álló hangállomány nagy részén, annál jobban kihasználják azt az állományt, annál jobban rajta vannak azon az állományon, annál inkább „otthon” vannak benne. Nos, nem meglepő az eredmény: tőszókincsünkben a hangállományhoz képest a mondott értelemben legotthonosabban a fgr., török, magyar belső elemek szerepelnek; legkevésbé otthonosan pedig az említett fiatalabb kölcsönzött rétegek elemei. (Szövegeket vizsgálva kissé másként kell interpretálnunk az eloszlás egyszerűségét: minél egyszerűsebb az eloszlás, annál színesebb, hangokban gazdagabb a szöveg — minél kevésbé egyszerűen, annál egyszínűbb, a szó bizonyos etimológiai értelemben „monotonabb”. Ez távolról sem kíván értékítélet lenni: a szóértékszervezési céljának megfelelően, ösztönösen vagy kevésbé ösztönösen, választhat „szürkébb” anyagot és „színesebbet”.)

Az 5. táblázat kissé nehézkesnek tűnhet az egyes csoportok eloszlása egyszerűségének érzékeltetésére. Próbáljuk meg valahogy egyetlen, az egyes csoportokra jellemző, értékkel kifejezni ugyanezt, vagy legalábbis valami ehhez hasonlót.² Ebből a célból állítsunk össze minden egyes etimológiai csoport részére egy-egy azonos felépítésű táblázatot, a következő módon. A hangokat nem ábécé-rendben soroljuk fel, hanem csökkenő gyakoriságuk

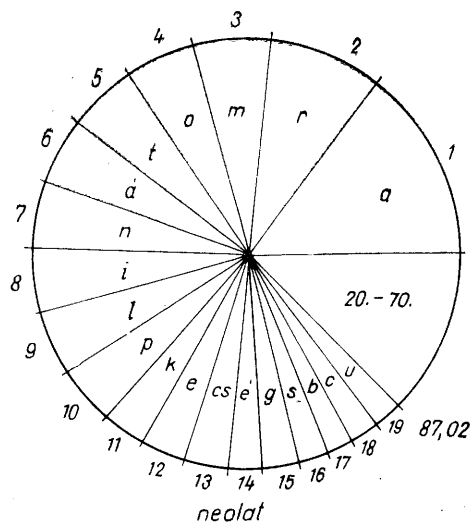
² Más célból ugyanezt a módszert alkalmazza V. A. MOSKOVICZ következő könyvében: Статистика и семантика. Изд-о «Наука», Москва. 1969. — 24 k. — Arra a körülményre, hogy MOSKOVICZ módszerét itt alkalmazhatjuk, NAGY FERENC tudományos kutató hívta fel a figyelmemet. Ugyanő számos további jó ötlettel szolgált az anyag feldolgozása során, ezért őszinte köszönetemet kívánom kifejezni.

sorrendjében, minden egyes hang mellé odaírva, hányszor fordult elő. Ha ezt a munkát elvégeztük, akkor például a fgr. réteg számára egy olyan felsorolást kapunk, mint amilyent a 6. táblázat mutat. (Az 1. ábra körgrafikonjai mögött is eleve egy-egy ilyen felsorolás húzódik meg.)

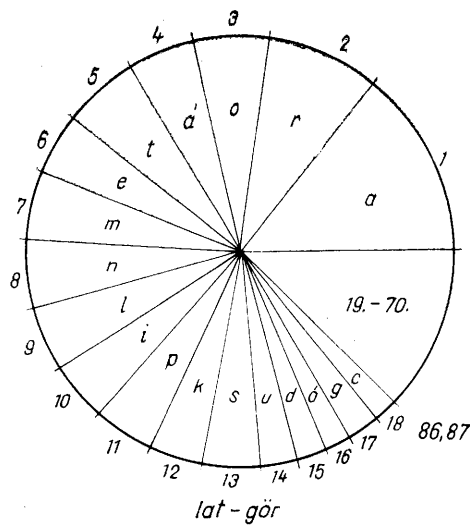
Ezenkívül még azt láthatjuk a 6. táblázaton, hogy a csökkenő gyakoriságban egymást követő hangokat (hangok csoportjait) bizonyos szakaszokra osztottuk. Tettük ezt a következő módon: a *bb*-től kezdődően leválasztottuk a nullások (egyetlen egyszer sem előfordulók) csoportját. A fennmaradókat, tehát a legalább egyszer előfordulókat, kilenc további egyenlő csoportra osztottuk úgy, hogy a beosztás alapjául az előfordulásokat vettük. A legkevesebbszer, vagyis egyszer előfordulók (*dd*, *dz* stb.) és a leggyakrabban, vagyis 175-ször előfordulók (*a*, *e*) között a távolság (gyakoriság szerint mérve) 175, ezt osztjuk kilenc egyenlő részre, akkor egy-egy osztály 19,45 „széles” vagy „magas” lesz. Más szóval, ha a 0-sok osztályát, mint utolsó osztályt, tekintjük a X. osztálynak, akkor az ezt megelőző, IX. osztály 0,01—19,46-ig fog tartani (vagyis, hangokban: a *zz*-től fel egészen a *ly*-ig: ez utóbbi 19-szer fordul elő, tehát még éppen a 19,46-os határon belül van), a VIII. osztály 19,47-től 38,91-ig (figyeljük meg a határt az *ú* és a *z* között: ez utóbbi a maga 39-es előfordulásával már túlesik a VIII. osztály felső határán, a VII.-be), a VII. 38,92-től 58,36-ig, VI.: 58,37—77,81, V.: 77,82—97,26, IV.: 97,27—116,71, III.: 116,72—136,16, II.: 136,17—155,61 (ez az osztály egyetlen hanggal sem volt képviselve a fgr.-ok között), I.: 155,62—175,06 (ide esik be az *a* és az *e*). Adjunk rendre bizonyos súlyt az egyes osztályoknak: az I.-é legyen, mondjuk, 1,0, a II.-é 0,9, III. 0,8 . . . IX. 0,2, X. 0,1. És végül adjunk súlyt (számértéket) az egész rétegnek, a rá jellemző eloszlásnak úgy, hogy az I.—X. osztály mindegyikébe tartozó elemek számát megszorozzuk az adott osztály 0,1—1,0-ig terjedő súlyával, e szorzatokat összeadjuk, s az így kapott eredményt tekintjük az adott réteg eloszlási súlyának (természetesen szerencsésebb nevet is lehetne ennek a jellemzőnek találni), jelöljük *v*-vel. A v_{fgr} ezek szerint így alakul: az I. osztályban volt két elem, $2 \times 1,0 = 2,0$, II. — 0,0; III. szintén két elem (*k* és *l*), $2 \times 0,8 = 1,6$, IV. egy elem (*r*): $1 \times 0,7 = 0,7$. . . IX. 20 elem, $20 \times 0,2 = 4,0$, X. 17 elem, $17 \times 0,1 = 1,7$. A szorzatokat összeadva végül azt kaptuk, hogy $v_{fgr} = 21,40$. Ez így önmagában semmit sem mond, ám ha rendre kiszámítjuk a többi rétegre is ezt az értéket és a *v*-k csökkenő nagyságrendjébe állítjuk az egyes rétegeket, ezt kapjuk: 1. fgr. 21,40, 2. magyar 19,80, 3. török 18,90, 4. német 16,80, 5. latin—görög 16,00, 6. neolatin 15,90, 7. szláv 15,80. Világos, hogy a *v* értéke elméletileg 70 és 7 között mozoghat: az előbbi, maximális értéket akkor venné fel, ha mind a 70 hang egyenlő, legnagyobb gyakorisággal állna (mint a fgr.-ok között csupán kettő tette: az *a* és az *e*); az utóbbi, minimális értéket akkor, ha minden egyes elem 0-szor állna, vagyis egyik vizsgált hang sem fordulna elő egyszer sem az adott rétegben, az adott réteg ideálisan idegen az adott hangokhoz. (Ez az idegenség olyan nagy fokú, hogy a valóságban elő sem fordulhat: nem lehet olyan etimológiai réteg, amely csak nemlétező hangokból áll.) Minél nagyobb a *v* értéke, annál több viszonylag gyakoribb hangot tartalmazott az adott réteg, annál több hangot kihasznált az adottakból — és megfordítva: minél kisebb a *v* értéke, annál jobban megközelíti az „ideálisan idegen” (valóságban nem is létező) réteget. Nos, így tekintve ezek a számok már valóban kifejeztek valamit: valami olyat, amit az 5. táblázat kissé körülényesebben fogott meg. A tíz osztályra bontás teljesen önkényes volt —



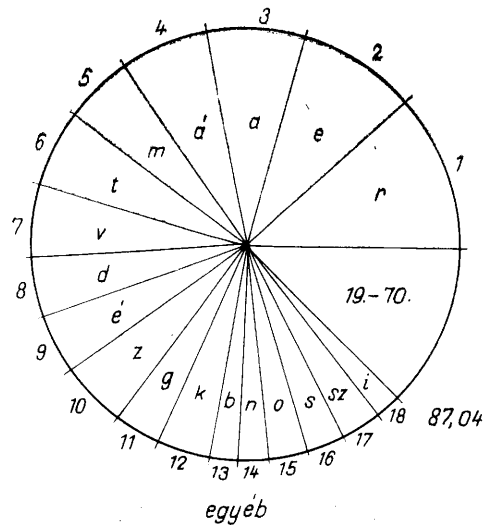
1. ábra



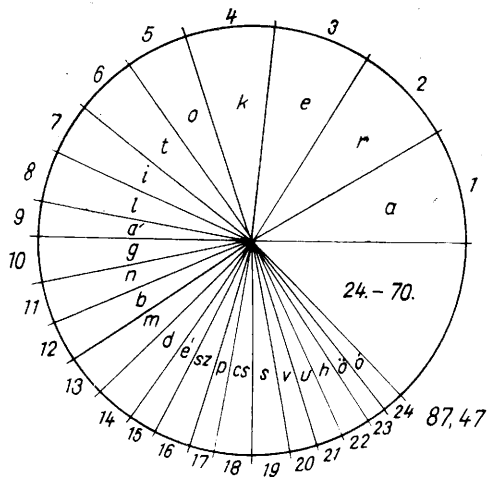
neolat



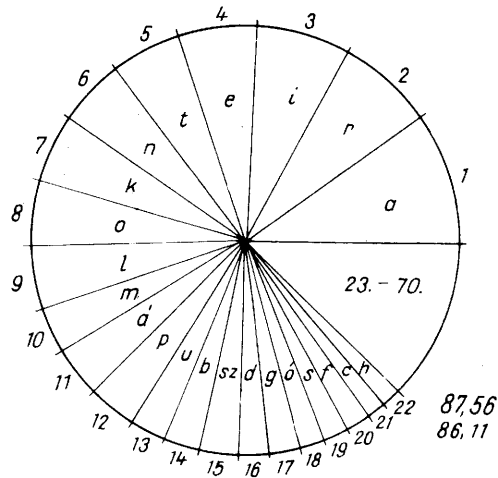
lat-gör



egyéb



bizonytalan



nincs benne a Szófsz.-ban

1. ábra (folyt.)

lehetett volna kisebb, tehát finomabb osztályokat is létrehozni, ám már ezzel a durva felosztással is elértük célunkat. A „súlyozás” mértéke (1,0, 0,9, 0,8 stb.) szintén önkényes volt. Tegyük még hozzá, hogy általános esetben elegánsabb, ha nem a v -vel magával, hanem $\frac{v}{N}$ -nel dolgozunk, ahol N a számí-

tásba vett elemek száma (esetünkben: 70). A $\frac{v}{N}$ értéke néhány réteg szá-

mára: fgr. — 0,306, magyar — 0,283 ... szláv — 0,226. A $\frac{v}{N}$ „értelmesebb” mint maga a v : megmondja, hogy az adott réteg hangjainak melyik osztály-

6. táblázat

Hang	Gyako- riság	Hang	Gyako- riság	Hang	Gyako- riság	Hang	Gyako- riság
a	175	d	39	c	2	hh	0
e	175	j	39	gg	2	ch	0
		ö	39	jj	2	cch	0
k	123	z	39	nn	2	kk	0
l	121			pp	2	lly	0
r	115	ú	31	tt	2	mm	0
		p	29	zs	2	vv	0
t	89	gy	28	dd	1	x	0
o	85	ó	28	dz	1	y	0
		es	25	ggy	1	zsz	0
h	77	í	25	as	1		
sz	75	ó	25	ty	1		
é	73	b	24	tty	1		
á	72	u	23	zz	1		
i	71	ü	20				
f	69			bb	0		
v	66	ly	19	cc	0		
m	62	ll	17	ccs	0		
		ű	13	ddz	0		
n	53	ssz	5	dzs	0		
g	50	nny	3	ddzs	0		
s	47	rr	3	ff	0		
ny	43						

ban van mintegy a súlypontja. Így a fgr. 0,3-as érték azt jelenti, hogy e rétegen belül a súlypont a VIII. osztályban van (annak az értéke volt 0,3), míg a szlávé majdnem egy teljes osztállyal lejjebb, a IX. osztályban (0,2-es érték).

Mint mindenféle matematikai modell, ez is alkalmazható a legkülönbébb területeken: a szemantikától kezdve (ahonnan mi is kölcsönöztük, vö. I. sz. lábjegyzet) például egészen az antroponimikáig, amikor az N (a vizsgált elemek száma) mondjuk az *István, József, Lajos* stb. előforduló keresztnévek mennyisége és a feladat az, hogy valamely terület és/vagy kor névadási eloszlását jellemezzük úgy, hogy az egységesen szembeállítható legyen más területek és/vagy korok névadási gyakorlatával (a nevek eloszlásával). Ez a jellemzés természetesen sommás lesz: a módszer nem érzékeny arra, hogy konkrétan *József* volt-e a leggyakoribb keresztnév, vagy *Abelard*; egy ilyen jellemzés segítségével várhatóan inkább bizonyos nyelven túli okok lennének kitapinthatók, mint például valamely közösség konformizmusa—atomizáltsága a névadásban; diakrón síkon: valamely közösség elszigeteltsége—nyíltsága, hagyománytiszteletének foka stb. Ugyanilyen sommás jellemzést kaptunk itt a hangok tekintetében is.

A hangokat azonban egyéníthetjük is, például a következő módon. Számozzuk meg az egyes hangokat gyakoriságuk sorrendjében és nevezzük ezt a sorszámot rangnak (r). Így például a 6. táblázaton az r_a (= 'az a hang rangja') = 1., $r_e = 2.$, $r_k = 3.$. . . $r_y = 69.$, $r_{zss} = 70.$ (Bár itt egyedítettünk, valójában megint eltekintettünk a hangok természeti mivoltától, tehát még a matematikai módszereken belül maradtunk: „rangja”, tehát gyakorisági sorszáma lehet például a keresztnév-anyag minden egyes típusának; valamely több jelentésű szó egyes jelentéseinek; valamely szónak a gyakorisági szótárban stb. — arról már nem is beszélve, hogy a nyelven kívül az élet mennyi más területén lehet értelme bizonyos rangsorolásnak, s akkor minden olyan területen alkalmazhatók az itt következő vagy ahhoz hasonló módszerek.) Vegyük számba, hogy a fgr., magyar, török stb. listák mindegyikén az egyes hangok milyen rangot kaptak, állapítsuk meg az adott hang számára a legmagasabb rangot (r_{\max}) és a legkisebbet (r_{\min}), majd képezzük a két érték különbségét, melyet d -vel fogunk jelölni. A 7. táblázaton közöljük az egyes d értékeket és a hozzájuk tartozó elemeket (azokat a hangokat, ahol a d ekkora értékű volt; technikai okoktól itt és minden további, a rangra alapuló számításnál összesen 72 elemmel dolgozunk: szerepeltetjük a q -t is, meg a w -t is; a 8-as — „egyéb” — etimológiájú elemeket itt számításán kívül hagytuk, mert e kis számú s heterogén csoport erősen torzította az eredményeket). A táblázatot a következő módon kell érteni: A hosszú zss elem minden egyes etimológiai rétegben azonos rangot kapott, ezért esetében az $r_{\max} - r_{\min} = d = 0.$ (Egyik etimológiai rétegben sem fordult elő, ezért mindenütt ugyanazt a 72.-es rangot kapta.) Nem volt egyetlen olyan hang sem, amelynél $d = 1.$ Ha az „egyéb” réteget nem számítjuk, a 15—24 d -s maximális távolság jellemző: e távolság valamelyike lép fel összesen 34 esetben, tehát majdnem az esetek felében. Amely hangok ennél lényegesen kisebb d értékeket mutatnak fel, azokról azt mondhatjuk, hogy tőszó-kincśünk minden egyes etimológiai rétegére szinte egyaránt jellemzőek, egy-egy etimológiai réteg szemszögéből tehát nem szignifikánsak. Ilyen az a (ném.: 2., tör.: 3., az összes többi rétegben 1.), az r (tör., ném.: 1., fgr.: 5., a többiben: 2.), az o (már kissé szórtaabb sorrenddel: lat.—gör.: 3.; szl., neol.: 4.; tör., ném.: „9”: 5.; ma.: 6.; fgr.: 7.; „0”: 8.), az e, k, i, n stb. Sajátos a

7. táblázat

d	hangok	d	hangok	d	hangok
0	zsz	16	tty, y	31	kk
1	—	17	ccs, dz, jj, m	32	—
2	a	18	ff, ggy, ch, cch, v	33	—
3	—	19	gy, hh	34	ő
4	r, vv	20	pp	35	—
5	o	21	cs, ly	.	.
6	í, t	22	cc, dd, dzs, sz, zz	39	h, tt
7	—	23	b, bb, c, ddz, ddsz, ssz	40	ű
8	á	24	f, gg, nny, ty, zs	41	ú
9	d, g, i, l	25	—	42	—
10	e, k, n, q, s	26	rr, ss	.	.
11	ó	27	—	.	.
12	j, ll, w	28	x	48	ö
13	ü	29	lly	.	.
14	é, mm, u, z	30	nn	.	.
15	ny, p				

helyzet a 0-s előfordulású *zsz, vv* hangokkal: ők éppen teljes hiányukkal jellemzők egyaránt minden rétegre. Talán még érdekesebb nyelvészeti szempontból a másik vég, a nagy *d*-k tartománya. Itt nyilván polarizált helyzetről kell beszélnünk: egy vagy több rétegben valamely hang viszonylag gyakori (kis sorszámú), a többiben viszonylag ritka — így adódik a nagy különbség. Valóban: az *ő* 48-as maximális rangkülönbségét az adja, hogy *e* hang igen gyakori a török rétegben (a 9., vö. *bögöly, bölcs, bölcső, bölény, csök, csökik, gödény, görény, gyöngy, gyötör, gyümölcs*: összesen 45 *ő*, vö. 1. táblázat; ugyanonnan az is kiderül, hogy míg az összes hangnak mindössze 4,15%-a fordul elő a különféle török eredetű lexémákban, addig az összes *ő*-nek majd négyszer ennyi, közel 15%-a), ám egyszer sem fordul elő a neolatin elemek között. Az *ú* a fgr. *s* a török rétegre jellemző (egyenként a 24. helyen) — teljesen hiányzik viszont a latin—görögök közül. Némileg más a helyzet az *ű*-vel: viszonylag gyakori a törökök között (a 26.), közepes helyen áll a fgr. és magyar csoportban — ám egyetlen sincs belőle szláv, német, latin—görög, neolatin jövevényeink között. A *h* a 8. a fgr.-ok sorában (ott több mint kétszer annyi, mint a fgr.-ok átlagos súlya a hangok között: *ha*⁽¹⁾, *hab, had, hogy*⁽¹⁾, *hagyma, haj*¹, *hajó, hajt*¹, *hal*¹, *hal*² stb., ritkán szó belsejében: *felhő* — itt épp a *h* eredete tisztázatlan, vö. TESz. *felleg* a. — és szóvégen: *juh, méh*¹); ugyanakkor teljesen hiányzik a neolatinok közül. Még néhány viszonylag nagy *d*-s hang, utána zárójelben + jellel az a réteg, amelyre jellemző, — jellel, amelyben ritka: *tt* (+, „0”, —, „9”), *ő* (+tör., —lat.—gör.), *kk* (+ném., —fgr.), *nn* (+lat.—gör., —tör.), *lly* (+lat.—gör., szl., —, „0”), *x* (+, „0”, —[az összes többi réteg]), *rr* (+fgr., —, „9”), *ss* (+ném., lat.—gör., —, „9”).

A 4. részben látni fogjuk, hogy az egyes hangok megkaphatják sorszámukat (rangjukat) a tőszókincs egészében elfoglalt helyük (gyakoriságuk) szerint, ezek az értékek, mint sorszámok, természetesen egész számok lesznek 1.—72.-ig. Ám eljárhatunk úgy is, hogy az egyes hangoknak az egyes rétegekben elért rangjait vesszük és ennek átlagát írjuk mellé (jelölve: \bar{r}). Az \bar{r} már nem okvetlenül egész szám. Így tőszókincsünk leggyakoribb hangjának, az

a -nak esetében $\bar{r} = 1,33$ (ha az „egyéb” réteget továbbra is elhagyjuk; ha azt számításba vesszük, akkor 1,5), mert rangja hét rétegben 1., egyben 2. (ném.), egyben 3. (tör., ugyancsak 3. az „egyéb” kategóriában is), akkor

$$\bar{r}_a = \frac{(7 \times 1) + (1 \times 2) + (1 \times 3)}{9} = \frac{12}{9} = 1,33 \quad (1)$$

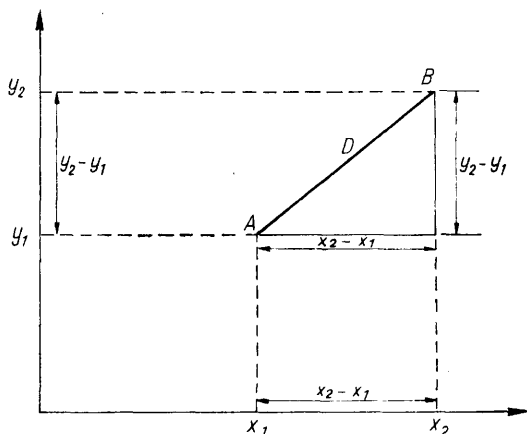
hasonló módon a második leggyakoribb hang, az r átlagos rangja 2,11 (ha az „egyéb” kategóriát is vesszük, akkor véletlenül éppen 2,0) és így tovább. A 4. alatt tárgyalandó rang megmondja, hogy az adott hang ténylegesen hányadik helyet foglalt el mai tőszókincseinkben. Ám mai tőszókincseinkön a különféle etimológiai rétegek különféle arányban osztoznak: legtöbb bennük az olyan, ami benne sincs a SzófSz.-ban, vagy ismeretlen—vitatott eredetű. Ezek a természetes, egész sorszámok tehát az e két rétegben kialakult sorrendhez fognak leginkább hasonlítani, ha akarjuk, ha nem — e két réteg tagjai a maguk nagy számával mintegy „elnyomják” a többiek szavazatát. Az \bar{r} értéke viszont éppen abban van, hogy érzéketlen az egyes rétegek számosságára: a 108 neolatin szavakban kialakult hang sorrendje épp olyan súllyal számít bele az átlagba, mint a 13 532 „0”-s hang alapján kialakult sorrend. Az \bar{r} azt mutatja, milyen lenne a hangok sorrendje, ha minden egyes etimológiai réteget ugyanannyi hang képviselne szókincseinkben.

A rangokra épülő eddigi legszebb konstrukció véleményünk szerint a következő. Alkossunk egy pontthalmazból egy hetvenkét dimenziós teret. Az egyes dimenziókat jelöljük rendre a következő szimbólumokkal: a, \acute{a}, b, bb stb. . . . z, zz, zs, zzs . E jelöléssel egyben rámutattunk, mit kívánunk megfeleltetni az egyes dimenzióknak. Egy-egy pontnak viszont feleltessünk meg egy-egy etimológiai réteget (csoportot): minden egyes pontot egy-egy szám-hetvenkettes határoz meg, ahol is a számértékeket természetes módon az egyes dimenziókban mért rang adja meg. Így például a „fgr.”-ként interpretált pont koordinátái (vö. a 6. táblázat adataival): $a = 1, \acute{a} = 11, b = 31, bb = 54, c = 40, cc = 55, cs = 28, ccs = 56, d = 20, dd = 47, dz = 48, ddz = 57, dzs = 58, ddzs = 59, e = 2, \acute{e} = 10, f = 13, ff = 60, g = 17, gg = 41, gy = 26, ggy = 49, h = 8, hh = 61, ch = 62, cch = 63, i = 12, \acute{i} = 29, j = 21, jj = 42, k = 7, kk = 38, l = 9, ll = 29, ly = 36, lly = 69, m = 10, mm = 52, n = 6, nn = 47, ny = 19, nny = 38, o = 7, \acute{o} = 27, \acute{o} = 22, \acute{ó} = 30, p = 25, pp = 44, q = 67, r = 5, rr = 39, s = 18, ss = 50, sz = 9, ssz = 37, t = 6, tt = 45, ty = 51, tty = 52, u = 32, \acute{u} = 24, \acute{u} = 33, \acute{u} = 36, v = 14, vv = 68, w = 69, x = 70, y = 71, z = 23, zz = 53, zs = 46, zzs = 72. (A számok tehát, ismételjük meg, azt fejezik ki, hogy az adott csoporton belül h á n y a d i k, csökkenő gyakoriságok sorrendjében, az adott hang: a fgr.-ok között az a az első, az \acute{a} a tizenegyedik, a b a harmincegyedik és így tovább.) Minden további etimológiai réteget is ugyanígy meghatározunk egy-egy szám-hetvenkettessel — ahhoz hasonlóan, ahogy a szokásos síklapon egy számpárossal jellemzünk egy pontot (például: $x = 2, y = 3$), ahogy a térben egy számhármassal ($x = 3, y = -5,6, z = 99$) stb. Könnyűszerrel és egészen elemi ismereteket felhasználva ki tudjuk számítani mármost két pont távolságát a két-dimenziós koordináta-rendszerben (l. 2. ábra): az A és B pont közötti távolságot felfoghatjuk úgy, mint egy derékszögű háromszög átfogóját. Az egyik befogó hossza: $x_2 - x_1$, a másiké: $y_2 - y_1$. Tekin-$

tettel arra, hogy a Pitagorasz-tétel értelmében a derékszögű háromszög átfogójának négyzete egyenlő a két befogó négyzetének az összegével, ha az A és B közti távolságot D -vel jelöljük, ezt is írhatjuk: $D^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$, ebből pedig

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2)$$

Ha még egy harmadik dimenzió is van, akkor a két pont három dimenzióban mért távolságának a négyzete egyenlő a három dimenzió mindegyikében felvett értékek különbsége négyzeteinek az összegével — és így tovább. A kapott eredmény számszerűen kifejezi a két pont közötti távolságot az



2. ábra

adott többdimenziós térben. Nem kell itt metrikus értelemben vett távolságra gondolni, mégis kezelhető értékeket kapunk, amelyeket értelmezhetünk. A kérdést ezek után úgy tehetjük fel: vajon milyen lesz a távolság az egyes etimológiai rétegek között? A részletes választ a 8. táblázat tartalmazza, melyről két pont közötti távolságot éppen úgy le lehet olvasni, mint a közönséges, például városok egymás közötti távolságát feltüntető táblázatokról. (A vegyes „egyéb” réteget figyelmen kívül hagytam. A szükséges számításokat részben manuálisan — megfelelő táblázatok, asztalszámológép stb. — segítségével végeztem, ezért a benne előforduló hibák vár, hatóan egy nagyságrenddel nagyobbak lesznek, mint az egyéb értékek esetében. melyeket közvetlenül elektronikus gépekről kaptunk.) Így, mondjuk, a fgr csoport távolsága a SzófSz.-ban nem szereplő, modernebb elemektől 93,34, a magyar belső keletkezésűekétől viszont csupán 70,61, a törökökétől már 85,56, a szlávokétól 93,08 stb. Talán plasztikusabbá válik a táblázat, ha etimológiai csoportonként úgy rendezzük, hogy a hozzá közel állókat közelebb, a tőle távolabb állókat távolabb soroljuk fel. Íme: a **fgr.** réteghez a legközelebb van a magyar, majd sorra így következnek a távolodó csoportok: „9”, török, szláv (ezek az átlagosnál közelebb voltak a kiindulástól vett fgr. réteghez); „0”, német, latin—görög, neolatin (ezek voltak az átlagosnál távolabb — a továbbiakban az átlagostól inneni s túli határt csupán egy pontosvesszővel

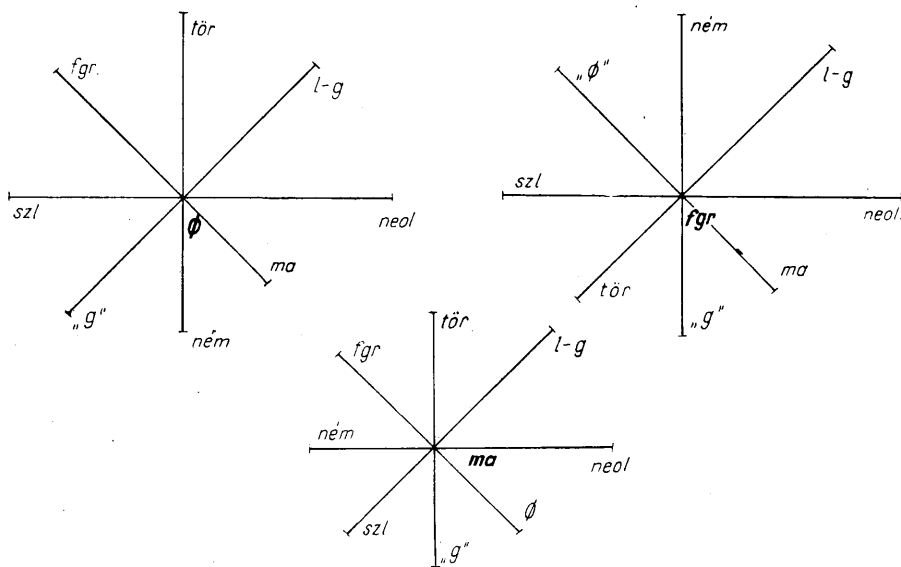
fogom jelezni; a pontos számértékek természetesen leolvashatók a 8. táblázat-ról); **magyar:** „0”, szláv, német, fgr., török; latin—görög, neolatin; **török:** „9”, magyar, szláv; fgr., német, „0”, neolatin, latin—görög; **szláv:** „9”, magyar, neolatin, latin—görög, német; török, „0”, fgr.; **német:** magyar, latin—görög, „0”, szláv; „9”, neolatin, török, fgr.; **latin—görög:** neolatin, német, szláv, magyar; „9”, „0”, török, fgr.; **neolatin:** latin—görög, szláv, német, „9”; magyar, török, „0”, fgr. Az eredmények egyáltalán nem mondanak ellent

8. táblázat

„0”	fgr	magyar	török	szláv	német	latin—görög	neolatin	„9”	
0	93,34	63,46	97,63	92,42	72,33	99,58	110,65	88,03	„0”
	0	70,61	85,56	93,08	93,57	112,88	119,48	77,55	fgr
		0	72,54	66,13	66,58	87,70	96,08	65,61	magyar
			0	79,92	90,41	108,04	97,83	97,63	török
				0	75,00	74,08	71,97	63,26	szláv
					0	71,82	83,61	80,25	német
						0	65,65	91,58	lat. —görög
							0	88,07	neolatin
								0	„9”

alapvető várakozásunknak; ennek alapján, ahol eltérés van a várttól, azon érdemes elgondolkodni: vajon nem az itt felhozott számoknak van-e igazuk, és nem a mi a priori (mérés nélküli) elképzelésünk a hamis. Nagyon szép s megnyugtató például a sorrend, ami a latin—görög vagy a neolatin után kialakul: valami effélet vártunk volna is (bár nem tudom, bizonyított tény-e, hogy a közeli rokon nyelvek hozzánk átkerült, tehát bizonyos értelemben véletlenszerűen megválogatott szavai mai magyar alakjukban ennyire hasonlítanak egymáshoz a maguk összességében — ha eddig ez nem volt bizonyítva, akkor egyelőre fogadjuk ezt az empirikus eredményt bizonyíték helyett). Ám mit szóljunk például ahhoz, hogy a német kölcsönszavak egészéhez a magyar belső keletkezésűek állnak a mondott értelemben a legközelebb? Itt abban a szerencsés helyzetben vagyok, hogy utalhatok a „magyar belső keletkezésű” szavak egy tekintélyes csoportjára tett korábbi megállapításomra, amikor még csak „szemmel” vizsgáltam azokat az elemeket, s alkalmasint nem is sejtettem, hogy ennyire atomjaira bontva is látni fogom őket. Eszerint a „magyar” -a végűek egy része gyanús, -ca, -ka stb. szövéget tartalmaz, vannak köztük szláv mintára keletkezett hangutánzó, elvonások, elhomályosult összetételek, jól ismert külső hatásra jöttek létre stb. (vö. NyK. 70: 205). Nos, úgy látszik, ami érvényes volt az -a végűekre, az nem kis-mértékben áll erre az egész csoportra. Jellemzők az e magyar csoport körül kialakult távolságok: a fgr.-okhoz valóban a magyarok vannak a legközelebb; ugyanakkor a „magyarok” minden más réteghez a legközelebb vannak, mintegy központban állnak úgy, hogy belőlük vezet a legrövidebb út valamennyi többi csoport felé, ha az egy pontból kiinduló utak összességének átlagos hosszát tekintjük. Nézzük meg a „magyar” csoportot követő felsoro-

láts s a nekik megfelelő számokat: csak a latin—görög és a neolatin van kissé távolabb a magyaroktól, a többi mind elég szorosan körülötte áll. Ilyen alapon nem kell különösnek tartanunk, hogy a magyarból kiindulva a fgr. csak a negyedik legközelebbi állomás (láttuk, hogy fordítva nem ez a helyzet), s hogy a németből kiindulva éppenséggel a magyar a legközelebbi „rokon”. Érdekes, hogy e szempontból a másik póluson — a finnugor csoport áll: a belőle kiinduló utak átlagos hossza a legnagyobb. A „minden más csoport-hoz közeli” magyar belső és a „minden más csoporttól távoli” finnugorok között foglal helyet a többi réteg: az átlagosnál rövidebb össz-úthosszal a „9”, a szláv és a német; az átlagosnál hosszabbal — a török, a latin—görög, a SzófSz.-ban nem szerplő és a neolatin. Egyébként, ha már ennyire biztosan mozgunk a távolságok között, megnézhetjük, mi áll közelebb s mi távolabb a két „vegyes” csoporthoz. Íme: „Ø”: magyar, német, „9”: szláv, fgr., török, latin—görög, neolatin; „9”: török, szláv, magyar; fgr., német, „Ø”, neolatin, latin—görög. Valóban, mintha a „Ø”-sok kissé „fiatalabbak” lennének, a „9”-esek pedig régebbiek — mint korábban említettük, más forrásból is ez a kép alakul ki e két rétegről. A 3. ábra ezeket a távolságokat érzékelteti három réteg szemszögéből („Ø”, fgr., magyar).



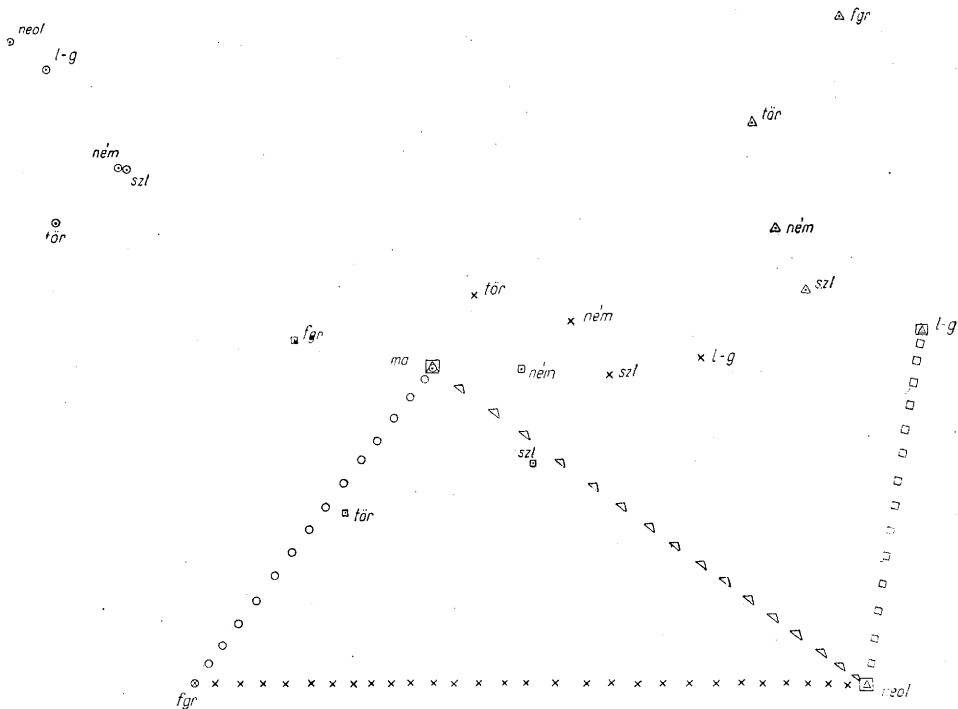
3. ábra

Az egyes rétegek kölcsönös közelségét, illetőleg távolságát másként is lehet érzékeltetni. Kölcsönösen legközelebb vannak egymáshoz a következő párok: „Ø” és magyar, latin—görög és neolatin, török és „9”. Egyoldalú közelség esete áll fenn, amennyiben az elsőként felsorolt és vastagon szedett nyelvhez a második legközelebb van, de e másodikhoz valamely más nyelv áll legközelebb: **fgr.** — magyar, **szláv** — „9”, **német** — magyar. Kölcsönösen legtávolabb vannak egymástól: fgr.—neolatin. Egyoldalúan a vastagon szedett nyelvtől legtávolabb helyezkedik el a másodikként közölt nyelv: „Ø” —

neolatin, magyar — neolatin, török — latin—görög, szláv — fgr., német — fgr., latin—görög — fgr., „9” — latin—görög.

Joggal vetődik fel egy olyan kérdés is, hogy egyáltalán mely pontok vannak a legközelebb egymáshoz (abszolút értékben, nem viszonylag) és melyek a legtávolabb. Íme, a legközelebbi két pont: török—„9” = 51,58, majd szláv—„9” = 63,26, magyar—„9” = 65,61 és csak ezt követi a latin—görög—neolatin = 65,65. Kár, hogy a „9”-es kategória maga kissé szétfolyó (vagy éppen ezért van annyi mindenhez közel? egyébként, mint fentebb láttuk, a magyar után rögtön a „9” következik az összetak rövidege tekintetében); ha ezt leszámítjuk, akkor valóban az áll legközelebb, amit vártunk. Másrészt az egymástól legtávolabbi pontokat még könnyebb rögtön kiemelni a táblázatból: fgr.—neolatin, majd fgr.—latin—görög, „0” —neolatin és török—latin—görög (ezek voltak azok a pontok, amelyek egymástól több mint 100 egységre helyezkedtek el).

Tekintsük meg végül a 4. ábrát. Ezen kiemeltünk négy etimológiai réteg-párt (az ábra alsó sorában: „fgr.—neol.”, bal oldalt: „fgr.—



4. ábra. Mag y a r á z a t: A körökkel (o o o) jelzett szakasz a fgr. és a magyar, a háromszögekkel (Δ Δ Δ) — a magyar és a neol., az ikszekkel (x x x) a fgr és a neol., a négyzetekkel (□ □ □) a neol. és a lat.—gör. rétegek közötti távolsággal arányos. Az e szimbólumokkal azonos szimbólumok az egyes rétegeknek e két-két rétegtől való távolságával arányosak, így tehát egymás között nem, csupán az alappárral hasonlíthatók össze. Például: az ábra bal felső sarkában elhelyezkedő, körrel jelzett tör., szl., ném., l-g., neol. rétegek távolsága itt a fgr.—ma. rétegpártól értendő; az ezzel mintegy átellenben elhelyezkedő, háromszögjelek a ma.—neol. rétegpártól való távolsággal arányosak; stb.

ma.”, jobb középén: „ma.—neol.”, jobb szélén: „l.-g.—neol.”). A „fgr.—ma.” rétegpár szimbolizálta szakasztól balra fent elhelyezkedő köröcskéknek az e szakasztól mért távolsága érzékelteti a megfelelő etimológiai rétegek távolságát a kiemelt rétegpártól; a „ma.—neol.” rétegpárhoz a kis háromszögekkel szimbolizált rétegek tartoznak; a „fgr.—neol.”-hoz az x-szel jelöltek, végül a „neol.—l.-g.”-höz az ábra középrészén és attól valamivel jobbra elhelyezkedő kis négyzetek. Az ábráról rögtön szemléletessé válik, hogy például a „fgr.—ma.” rétegpártól a török, a szláv és a német szinte egyenlő távolságra van (ezen belül a török mégis inkább a fgr.-hoz közelebb, az utóbbi kettő — a magyarhoz); jóval messzebb (egymáshoz viszont közel) helyezkedik el tőlük a latin-görög és a neolatin réteg: mindketten a magyarhoz észrevehetően közelebb vannak, mint a fgr.-hoz. Figyeljük meg a távolságok eloszlásának két típusát: a fgr.—magyar párhoz tartozó távolságok egyre nőnek (a török és a neolatin között itt 57 egység a különbség), hasonlóan a neolatin—latin-göröghöz tartozók (a maximális különbség, a fgr. és a szláv között: 86 egység) — ez a helyzet, ha rokon nyelveket emeltünk ki egy párba: ők egymáshoz a térben eléggé közel vannak, viszont a többi réteg tőlük távolabb; másrészt a fgr.—neolatin és a magyar—neolatin párhoz tartozó távolságokat (maximális különbségek: 18, 36), melyek kevésbé különböznek egymástól: nem rokon nyelvek kerültek párba.

Minták összehasonlításakor azok n -dimenziós térben mért távolságának összehasonlítása szintén igen elterjedt módszer az élet számos területén; nyelvészetben (megintcsak a szemantikában) már legalább másfél évtizeddel ezelőtt is éltek vele.³ El lehet képzelni, hogy az egyes minták között a távolság még kisebb is, mint itt, ha eleve olyan mintákat hasonlítunk össze, amelyek feltételezésünk szerint egy és ugyanazon automatáról (gépsorról stb.) jöttek le: például egy és ugyanazon költő egyenként eléggé hosszú életműdarabjait. A mi itteni mintáink is egy gépcsaládról jöttek le végső soron (ti. mind a mai magyar tőszókincs tagjai voltak, mai magyar hangalakjukban; nagyon meggyőző véleményünk szerint e szempontból az a körülmény, hogy ezen belül épp a tulajdonképpeni magyar belső keletkezésűek voltak a legközelebb az összes többi réteghez, ők voltak mintegy a tömegközéppontban); ám eleve feltételeztük, hogy e gépcsaládon belül annak különféle tagjaitól származnak, ahol is e tagok között eleve természetes volt bizonyos kisebb (fgr.—magyar, latin-görög—neolatin stb.) vagy nagyobb (fgr.—neolatin, török—neolatin stb.) távolságokat feltételezni. Láttuk, hogy ezek a távolságok már ezen a szinten, a hangok szintjén is, eléggé szépen bejöttek.

4. Mit mondhatunk ezek után összegező módon, immár az egyes etimológiai rétegektől függetlenül, hatezer-egynéhány valódi tőszavunk képviselte hangjaink összességéről, ezen szavak egész hangállományáról?

Hogy erre a kérdésre válaszolhassunk, lépünk egy kissé ki teljesen motiválatlan, tulajdonképpeni tőszavaink (a továbbiakban ezeket fogjuk a *m o r f* tőszavaknak nevezni) köréből. Állítsuk velük szembe azokat a kétes tőszavakat, amelyek valamely gyakori idegen eredetű előtagot hordoznak (mint *mikroszkóp*, *bibliotéka* stb.): nevezzük ezeket *E - m o r f* tőszavaknak („előmotivált”) — volt ilyen forrásunkban kétszáz-egynéhány. És vegyünk

³ CHARLES E. OSGOOD—GEORGE J. SUCI—PERCY H. TANNENBAUM, *The Measurement of Meaning*. University of Illinois Press. Urbana, 1957. — 91 kk.

fel még egy népesebb kétes tőszócsoportot azokét, amelyek esetében a szó végén van valamely gyakori idegen eredetű elem (mint *pliszé, imitáció, kli-ens* stb.): nevezzük ezeket U - m o r f tőszavaknak. Világos, hogy miért kétes az E-morf és az U-morf tőszavak tőszói státusa: nehezen lehetne magyar szempontból azt állítani, hogy a *mikroszkóp* — összetett szó, ám ugyanakkor a naiv magyar beszélő is érez valami gyanúsát (ti. szokatlanul gyakran s hasonló jelentéssel ismétlődöt) ezekben az elemekben éppen úgy, mint a gyakori idegen végűekben, amelyeket mégsem lehet magyar képzős alakulatoknak felfogni. (E felosztást, a pontos számszerű adatokkal együtt l. Nytud-Ért. 58: 520—521.) E két kontroll-csoport segítségével már mondhatunk valamit maguknak az amorf tőszavaknak az összességéről is.

Az átlagos szóhosszokat és egyben a magánhangzók arányát az összes hanghoz a 9. táblázat tartalmazza. Jól látható az, amit fentebb egyszer már érintettünk: korreláció van már így a szótári alakban is az átlagos szóhossz és a magánhangzósság között: minél hosszabb a szó, aránylag annál több magánhangzót kell tartalmaznia. (Ezért is került együvé ezen a táblázaton

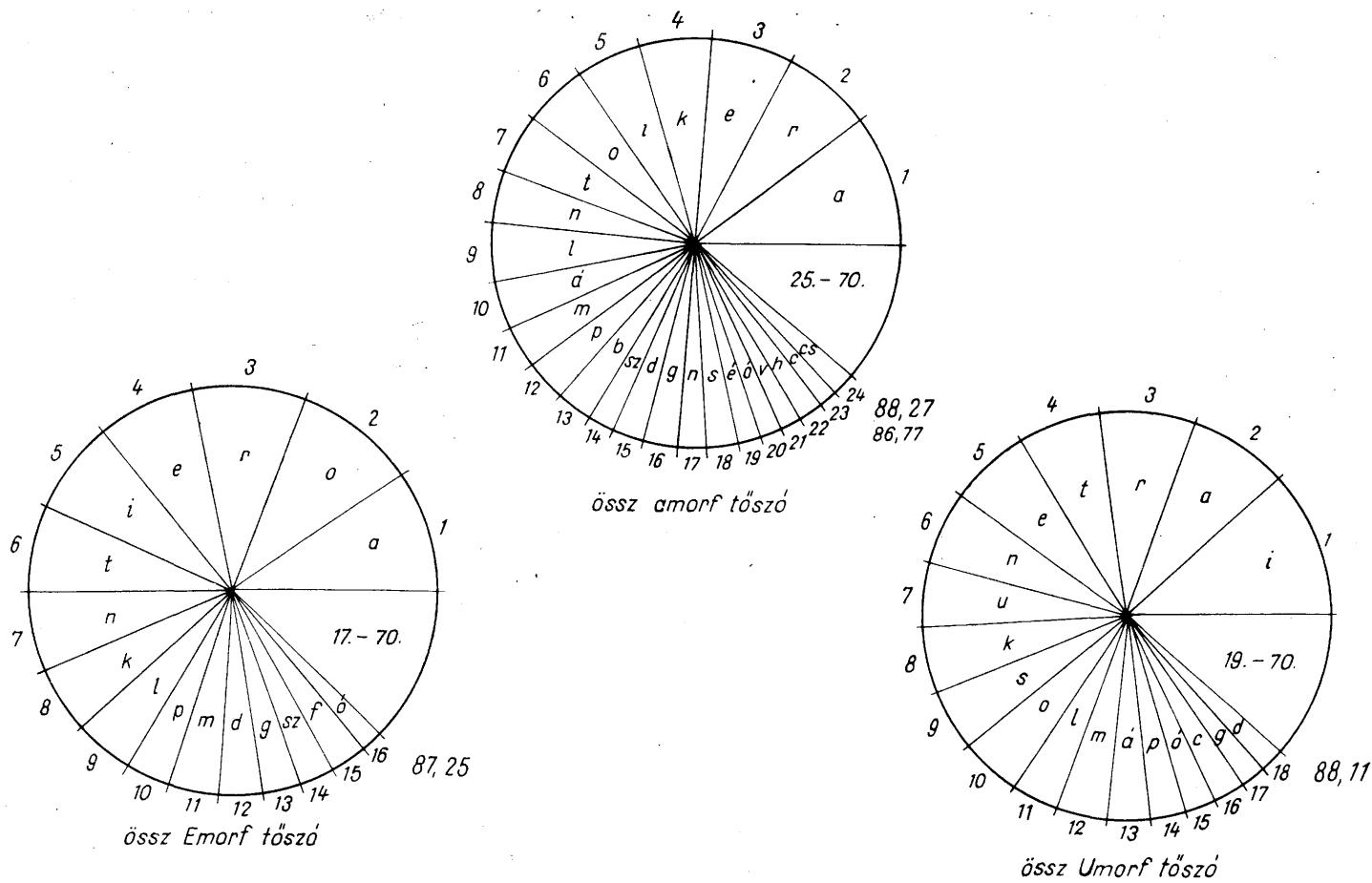
9. táblázat

Átlagos szóhossz és magánhangzósság

Tőszótípus	Átlagos szóhossz		Az összes hang hány százaléka magánhangzó?
	hangokban	szótágokban	
amorf	4,851	1,940	40,000
E-morf	7,925	3,236	40,837
U-morf	8,240	3,665	44,477

két olyan, első pillanatra diszparátnak tűnő, jelenség, mint a hossz és a magánhangzósság.) Itt megint nincs közvetlen szerepe a szó származásának: az amorf tőszavak között természetesen, a fentebb látott arányban, vannak ősi elemek; az E-morfok és U-morfok között a dolog természeténél fogva egyetlen fgr., magyar belső, török elemet sem találunk. Nincs közvetlen köze a magánhangzósságnak a szó korához sem: az E-morf és U-morf tőszavak természetesen mind egészen fiatalok, vagyis olyan korban jelentek meg nyelvünkben, amikor az már a feltevések szerint eléggé hozzászózott a mássalhangzó-csoportokhoz.

Tekintsünk végre az 5. ábrára. A felső középső kör érzékelteti a helyzetet az amorf tőszavak összességében: látjuk, hogy a leggyakoribb hangok az *a, r, e*; látjuk, hogy 23—24 hang adja ki az egész állomány 87—88%-át. Ennél kevésbé „otthonosak” a magyar hangrendszeren az E-morfok: a 16 leggyakoribb hang osztozik az állomány 87%-án, valamivel jobban eloszlanak mai hangkészletünkön az U-morfok. Mint fentebb, úgy most is mindezt pontosabban is érzékeltethetjük számokkal (l. 10. táblázat). Az amorf szavaknak ez a szembenállása az U- és E-morfokkal egyben megerősíti annak a korábbi elképzelésünknek a helyességét, amely szerint az eloszlás egyenletesebb voltát kell éppen mint „otthonosságot”, „meghonosodottságot” interpretálni.



5. ábra

10. táblázat

Tőszótypus	Hány különböző hang teszi ki az összes állomány			
	25	50	75	87
százalékát?				
amorf	3,2	8,2	16,5	23,0
U-morf	2,8	7,0	12,5	17,4
E-morf	2,8	6,0	11,5	16,0

Ezzel a vizsgálatunkkal párhuzamosan költői (s egyéb) szövegeket is feldolgoztunk hasonló módon. Az itt közölt adatok tehát nem csupán ön-magukban érdekesek számunkra, valamint egy esetleges további összehasonlítás céljából a TESz. utáni állapottal, hanem azért is, mert összevethetők magyar szövegek hasonló adataival. Szövegekben, ahol egyes gyakori elemek igen gyakran ismétlődnek; ahol egyes szavak gyakori toldalékos alakjaikban jelennek meg és így tovább — részben más törvényszerűségek uralkodnak, mint a szótárban. Szövegek ilyen szintű statisztikai feldolgozásában számos elődünk is volt, akiknek az eredményeihez hasonlíthatjuk a mieinket. Ez azonban már másik, bár ezzel szorosan kapcsolatos téma.

PAPP FERENC

Звуковая статистика разных этимологических пластов корневых слов венгерского языка

1. Было обработано примерно шесть тысяч корневых слов венгерского языка, взятых в их современной форме, расклассифицированных предварительно с помощью САМ по этимологическим пластам, так что в результате получена звуковая статистика для каждого из этимологических пластов, установленных на основе этимологического словаря Г. Барци.

2. В табл. 1 даны сводные количественные результаты: по столбцам — звуки, по строкам — этимологические пласты.

3. Некоторые частные результаты и дальнейшая обработка этих данных: В табл. 2 показано, какой процент всех звуков занимают гласные по разным этимологическим пластам. В табл. 3 показано, какой процент гласных (данного этимологического пласта) является велярным. В табл. 4 показано, какой процент гласных является длинным, а какой — кратким. В рис. 1 изображается нагруженность отдельных звуков в разных этимологических пластах. Табл. 5 содержит некоторые сводные результаты этих же данных.

Вводится понятие ранга: это — порядковый номер данного звука в данной этимологической группе при условии, что звуки перечислены и перенумерованы в порядке убывающей частоты. Устанавливается разность $d = r_{\max} - r_{\min}$ для каждого звука, помещенного, вообще говоря, под разными рангами в разных этимологических пластах. В табл. 7 перечислены звуки в порядке возрастающих разностей d . Особого внимания заслуживают звуки с большими d . По формуле (1) подсчитывается средний ранг (\bar{r}) для каждого звука (в формуле (1) для примера подсчитан \bar{r} для a , \bar{r}_a).

На понятии рангов строится модель 72-мерного пространства для совокупности этимологических пластов. Разные измерения соответствуют разным звукам; величина в каждом из измерений соответствует рангу данного звука в данной этимологической группе — тогда каждой точке в этом пространстве соответствует по этимологическому пласту,

рассматриваемому с точки зрения рангов звуков, составляющих венгерские слова в их современной форме, входящие в данный пласт. Расстояние в двумерном пространстве вычисляется по формуле (2) (ср. рис. 2); расстояние в 72-мерном пространстве подсчитывается по этому же принципу. В табл. 8 показаны расстояния между этимологическими пластами. Результаты этой таблицы можно перегруппировать в порядке убывающей величины (см. рис. 3: к финно-угорским, например, ближе всего расположены венгерские образования, дальше всех — романские; или к группе «не фигурирующих в этимологическом словаре» — О — ближе всего венгерские, дальше всех — романские и т. д.). В рис. 4 изображаются расстояния к разным парам этимологических групп: к паре финно-угорский—венгерский, финно-угорский—романский и т. п.

4. В табл. 8 наряду с бесспорными корневыми словам рассматриваются и те, у которых в начале выступает часто встречаемый иностранный элемент, или у которых в конце такой же элемент. Дается средняя длина в звуках, в слогах; процент гласных ко всем группам данной группы. (Видно: чем длиннее слово, тем больше в нем пропорция гласных.) В табл. 10 (рис. 5) даны пропорции наиболее часто встречаемых звуков в этих же трех группах корневых слов.

Ф. Папп