

Adatok tőszókincsünk etimológiai rétegeinek végződés szerinti eloszlásához

Az alábbiakban, a címnek megfelelően, csupán néhány adatot szeretnénk szolgáltatni ahhoz, hogy tőszókincsünk egyes fontosabb etimológiai rétegeire milyen végső magán- és mássalhangzók jellemzőek (és megfordítva: hogy szóvégi magán- és mássalhangzóink milyen arányban állnak finnugor, török stb. eredetű tőszavaink végén). Ezekhez az adatokhoz most mindössze egy-két kommentárt fűzünk. Hogy az itt szereplő, tisztos mennyiségű szám félre ne vezesse az olvasót, mindjárt bevezetőben le kívánjuk szögezni: egyáltalán nem a nyelvstatisztikát kívánjuk velük szolgálni, hanem az etimológiai és mai szinkrón leíró kutatásokat; a számok ezen belül egyszerűen csak azért vannak, hogy lássuk, milyen anyag fölött is rendelkezünk tulajdonképpen.

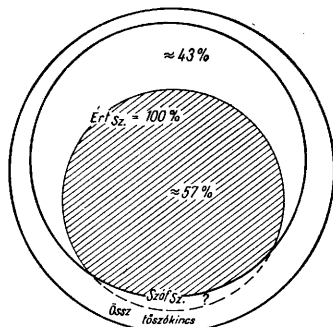
1. A vizsgált anyagra vonatkozóan a következőket mondhatjuk. Jó lenne, ha kijelenthetnénk: átvizsgáltuk az egész ma élő magyar tőszókincset, és arra vonatkoznak az alábbi adatok. Ez távolról sincs így. Az az érzésünk, s ezt itt nem kívánjuk bizonyítani (valószínűleg nehezen is lehetne egzaktan bizonyítani), hogy nem csupán valamely nyelv összes szavát nem lehet felsorolni, hanem — legalábbis számos nyelv, így a magyar esetében is — a tőszavakat sem. A tőszavakra vonatkozóan azonban, legalábbis a magyar filológiában, kiváló fogódzóink vannak, így azokat sokkal pontosabban le tudjuk írni, mint az egész szókincset.

Ezek a fogódzók: az ÉrtSz. és — tekintettel arra, hogy épp az eredettel kapcsolatos kérdésre hozunk adatokat — a SzófSz. Feldolgoztunk minden olyan tőszót, amely szerepel az ÉrtSz.-ban is, meg a SzófSz.-ban is, és csak ezeket a tőszavakat dolgoztuk fel. Amint erre más helyen (vö. A magyar nyelv története és rendszere. Nytud.Ért. 58: 518—22) már rámutattunk, azt elég pontosan tudjuk, hogy az ÉrtSz.-ban szereplő szűkebb értelemben vett tőszavaknak mintegy 57 százaléka szerepel a SzófSz.-ban is; a többi 43 százalék az utóbbi forrásban nem szerepel. Arra viszont nincs adatunk, hány olyan tövet tartalmaz a SzófSz., mely hiányzik az ÉrtSz.-ból. Feltesszük — nem teljesen alap nélkül —, hogy ilyen elem jóval kevesebb van. Ezt a helyzetet kívántuk durván érzékeltetni az 1. ábrán: eszerint a továbbiak során az ábrán vonalkázott résszel, a két forrás közös részével fogunk foglalkozni.

Hozzá kell ehhez tennünk, hogy az itt közlendő adatok nem tartalmazzák a nem egészen vitathatatlan tőszavakat (például az *anti-*, *bibli(o)-*, *mon(o)-* stb. kezdéseket, az *-ancia/-encia* végűeket, az *újra-féle* alakulatokat: erre vonatkozóan is vö. fent idézett előadásomat, id. kiadvány 519—20). Az adatok még nem véglegesek; a várható eltérések azonban olyan miní-

málisak lehetnek csupán, hogy sem az összképen, sem annak arányain már nem változtathatnak.

2. A címben ígért adatokat teljes egészükben tartalmazza az 1. sz. táblázat — e tény bejelentésével tulajdonképpen be is fejezhetnénk jelen



1. ábra. Az ÉrtSz.-ban és a SzófSz.-ban foglalt magyar tőszavak. (100% = az ÉrtSz.-ban szereplő vitathatatlan tőszavak. „Össz tőszókinés” — az egész ma élő magyar tőszókinés, melyet pontosan nem ismerünk.)

közleményünket, az olvasót a továbbiakban e táblázatra bízva. Ám tartunk tőle, hogy e táblázat, bár minden olyan adatot tartalmaz, amelyről az alábbiakban szó lesz, túlságosan tömören s így közvetlen filológiai felhasználásra kevésbé alkalmasan találja ezt az anyagot. Ezért először néhány megjegyzést fűzünk hozzá.¹

Kezdjük a sűrítve közölt adatok jobb szélső oszlopával. Abból kiderül, hogy az a bizonyos közös rész az ÉrtSz. és a SzófSz. tartalmazta tőszavak között összesen 3459 egyedet foglal magába (ez egyszermind az ÉrtSz. közölte szűkebb értelemben vett tőszavak 57 százaléka, l. fentebb). Ennek a mennyiségnek 17,75 százaléka, vagyis 614 szó — finnugor eredetű, 14,37 százaléka (497 szó) magyar belső fejlődés eredménye, és így tovább.

Ugyanilyen könnyen kezelhető a táblázat alsó két sora is. Itt például kiderül, hogy a vizsgált tőszó mennyiségből 614 egyed (17,75 százalék — teljesen véletlen egybeesés az előbbi adattal) *-a*-ra végződik, 3 *-á*-ra (0,09%), 36 *-b*-re (1,04%), és így tovább. Mindeddig tehát száz százalékként az egész vizsgált tőszókinés (az említett 3459 egyed) szerepelt.

A táblázat belsejében (tehát az A—ZS és „kötőjel” feliratú oszlopokban és az egyes etimológiai rétegek neveit viselő sorokban; ezt a nagy mezőt vastagabb vonallal kerítettem be a táblázaton) változik a helyzet. E belső területen ugyanis minden egyes s o r b a n a megfelelő etimológiai réteg szó mennyiségét tekintjük száz százaléknak; az erre vonatkoztatott, végződés szerinti rész százalékát az e végződést kifejező abszolút szám fölött, attól jobbra hozzuk (megjegyzendő: a táblázaton az abszolút számok mindig álló, a százalékok mindig dőlt típpussal vannak szedve). Például: *-a*-ra végződő finnugor eredetű tőszavunk volt 28, ez a finnugor tőszavak 4,56 százaléka; *-á*-ra végződő finn-

¹ Az 1. sz. táblázat adatainak összeállítását, a százalékok kiszámítását, valamint az eredmények átrendezését a 2. és a 3. ábra számára a szerző Mária és Ferenc nevű gyermeke végezte.

ugor eredetű tőszavunk volt 2, finnugor tőszavaink 0,33 százaléka, és így tovább. — Az abszolút számok alatt, tőlük kissé balra elhelyezett százaléértékek viszont o s z l o p b a n olvasandók, egymás alatt. Így, hogy továbbra is finnugor elemeinknél maradjunk: az említett 28 *-a* végű finnugor elem az összes *-a* végűek 4,56 százaléka, a 2 *-á* végű az összes *-á* végű kétharmada (66,67%), és így tovább. Még egy példa: *-b* végű szavaink 25 százaléka finnugor eredetű, 19,44 százaléka — magyar belső fejlődés eredménye, 5,56 százaléka — török, 16,67 százaléka szláv, 8,33 százaléka német, 2,78 százaléka latin—görög, 22,22 százaléka bizonytalan eredetű.

Vagyis: a sorokban (jobbra fent) elhelyezett százalékok megmondják, hogy az egyes eredet szerinti csoportok hány százaléka *a*, *b*, ... *kötőjel* végű (egyszerűbben így is mondhattam volna: *a_i* végű, ahol *a* — a magyar ábécé egy tetszőleges betűje, beleértve a kötőjelet is), míg az oszlopokban (balra lent) elhelyezett százalékok arra a kérdésre adnak választ, hogy adott *a_i* végződés mellett annak hány százaléka finnugor, magyar, ... bizonytalan eredetű (egyszerűbben így is mondhattam volna: *et_k* eredetű, ahol *et* az egyes fontosabb etimológiai csoportok valamelyike).

Azt hiszem, ezek után már senkit sem fognak visszariasztani az 1. táblázat számai: külön-külön mindegyikük nagyon jól érthető. Csak épp az a baj, hogy külön-külön: együttesen annyian vannak, hogy a kutató — ha csak nem már eleve valamely rejtett gondolattal fordul a táblához, nem valamely eleve megfogalmazott kérdésre keres benne választ — elveszhet közöttük. Ezen a további pontokban kívánunk segíteni, egyes összefüggésekre néhány ábrával is rávilágítva. (Talán mondanom sem kell, hogy az ábrák csak a durva arányokat érzékeltetik — ám a pontos szám adatok, abszolút értékben és százalékban, megvannak a táblázaton; a „durvaság” éppen arra jó, hogy a nagy vonalakat lássuk, az apró részletek helyett.)

A további feldolgozás előtt azonban ide kívánczok néhány részmegegjegyzés. — A fentiekben azt mondtuk: kötőjelre „végződő” szó. Amint erre már többször rámutattunk: ezek a szavak tulajdonképpen „valamelyik végükön kötőjelet tartalmazó elemek” — ez a végük lehet éppen az elejük is. (A többi betű esetén természetesen már nem tréfáltunk ilyen módon: az *-a* mindig szóvégző *-a-t*, a *-b* mindig szóvégző *-b-t* jelent stb.) De oly kevesen vannak ezek az elemek a minősített tőszókincsben, hogy fel is sorolhatjuk őket: *-né* (fgr); *-szám* (török); *-forma*, *-mód* (latin). E felsorolással egyúttal azt is elértük, hogy, ha úgy tetszik, a kutató ezeket az elemeket besorolhatja valódi végződés szerinti helyükre. Itt nem kívánok kitérni arra, miért volt célszerű számunkra ahhoz a megoldáshoz folyamodni, hogy e kötőjeles „szavak” (ti. önálló szócikkkel rendelkező címszavak az ÉrtSz.-ban) egy helyre kerüljenek — az idézőjellel utaltam is döntésünk legfőbb okára.

Következő megjegyzésünk magukat az eredet szerinti csoportokat érinti. Köztudott, hogy a SzófSz. gondos és hozzáértő szerzője összesen nem kilencféle eredet-minősítést használ, hanem ennél jóval többet. Hiszen számos esetben olyan kifejezéseket találunk nála, mint „valószínűleg”, „talán”, „minden bizonnyal”, „nem elfogadható” stb. Ahogy azt kedves elődünk is tette e folyóirat hasábjain (VERMES: NyK LI, 435.), mi (személy szerint: JAKAB LÁSZLÓ, aki a gépi feldolgozás számára az etimológiát kódolta) lényegesen egyszerűsítettünk ehhez képest. A „talán”-t legtöbbször biztosra vettük (!) és csak a „nem”, „nem fogadható el” stb. kifejezéseket vettük figyelembe olyan módon, hogy ahol e szavak álltak, ott az etimológiát „bizonytalan”-ra vettük

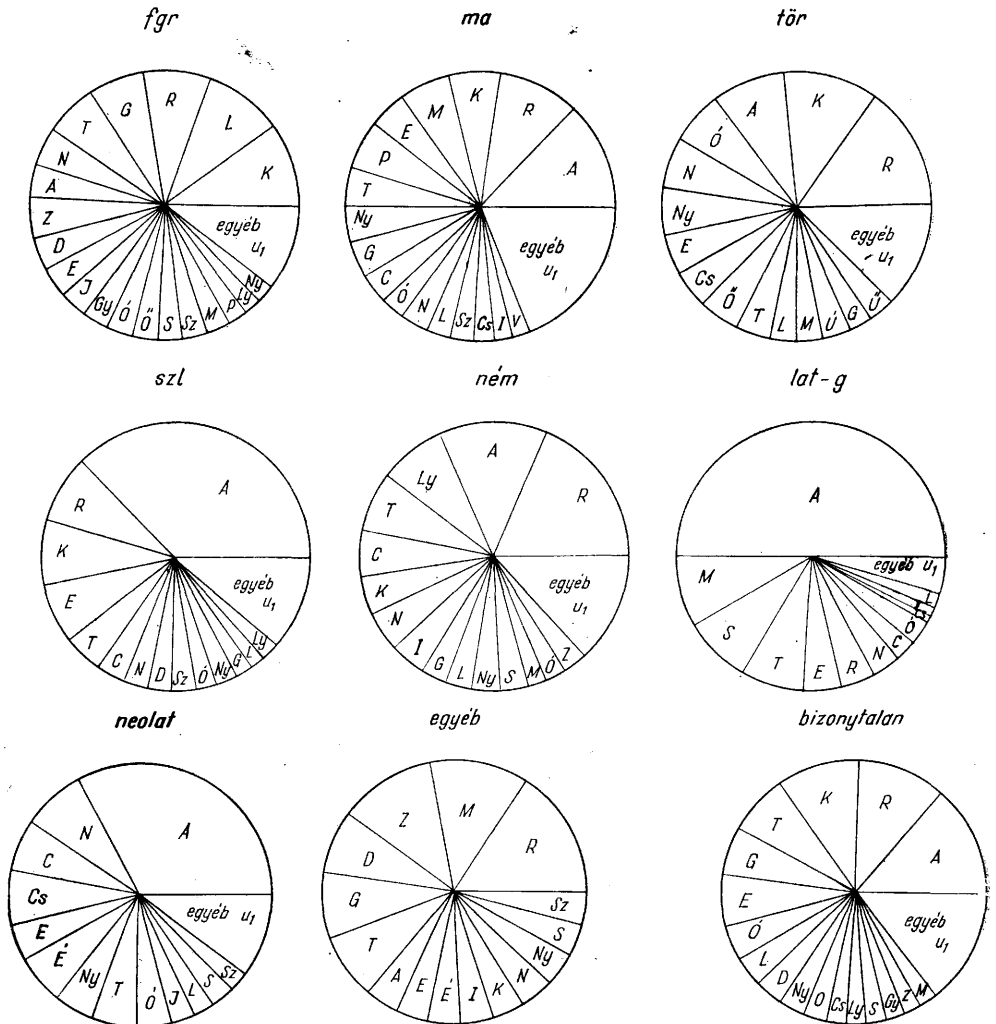
(illetőleg adtuk a pozitív etimológiát, ha ilyen volt a szócikkben). Tudjuk, hogy ezért sokan haragudni fognak ránk, sőt talán egyesek feldolgozásunknak ezt a részét használhatatlannak is fogják minősíteni. Azt azonban még ennél is jobban tudjuk, hogy másképp, árnyaltabban munkánk jelen szakaszában nem járhattunk el, a kérdés tehát így állt: vagy egyáltalán ne legyen etimológia, vagy csak ilyen elnagyoltan legyen. Bár, ismétlem, lehet, akadnak kollégák, akik erre azt mondják: akkor inkább ne is lett volna — azokat csak arra kérhetjük, ám ne használják ezt az egyetlen egy oszlopot feldolgozásunkban s leljék annál több örömeiket a fennmaradó hetvenhét kártyaoszlopon tárolt információkban. Reméljük, az 1. táblázat meggyőzi egyébként őket is arról, hogy olyan kiváló minőségű forrás esetén, mint a SzófSz., még ez az elnagyolt etimológiai jelölés is bőven megérte a fáradságot. — Az etimológiai rétegek egy további (korok szerinti) elmélyítéséről, pontosabbá tételéről, az alábbiakban lesz még szó.

És végül: számos célból nem csupán a legvégső betűkre (illetőleg: fonémákra — mint látható, a szóvégi helyzetben előforduló fonémák közül csupán a *-dz-t* nem vettük fel külön, valamint a hosszú mássalhangzókat), hanem a szóvégző fonémakettősökre, fonémahármasokra, sőt fonémanégyesekre is szükség lenne (illetőleg szükség lehet). Nem tudom azonban, el tudja-e képzelni az olvasó, az 1. táblázat ismeretében is, mit jelent e kívánsága. Nekünk e tábláink megvannak, mind a tőszókincsre, mind az ÉrtSz. egész szókincsére vonatkozóan, így elég pontosan megmondhatom: ez csupán a tőszókincsre vonatkozóan mintegy tíz nyomtatott ívet jelent — amelyik nap ezt a papírmennyiséget megkapjuk valamely kiadótól, boldog örömmel adjuk át közlésre az anyagot.² (Addig pedig szeretettel várjuk az érdeklődőket, hogy a nálunk levő egyelőre egyetlen példányt használják.)

3. Rendezük át az 1. táblázat egyes sorait olyan módon, hogy az eredet szerinti csoportokon belül az egyes végződésök csökkenő nagyságrendjükben szerepeljenek: előbb az abban az etimológiai csoportban leggyakoribb (arra a rétegre legjellemzőbb) végzések, majd a ritkábbak — és így tovább. Az ilyen átrendezés haszna, a mechanikus ábécé-rendhez képest, nyilvánvaló, így annak indokolására kár is szót vesztegetnünk. (De hogy miért nem mind-

² Pontosabban azt kellene mondanunk: 5–20 ív terjedelmű, attól függően, hogy milyen és mennyi kombinációban közöljük a betűkettősöket, hármásokat (a tőszavak esetén a betűnégyesek általában már nem mutatkoztak érdekesekeknek), hogy csak a tőszókincs egészére hozzuk-e ezeket az adatokat minden további bontás nélkül, vagy eredet szerinti csoportokra vonatkozóan is, esetleg szófaji bontásban is (önállóan vagy a szófaji bontást az eredettel kombinálva stb.). Látszólag elég lenne az egészet csak egy valamely kombinációban közölni, ám a helyzet az, hogy egyetlen kombináción belül hiába szerepel minden szükséges adat, bizonyos mennyiségen túl — márpedig a jelzett anyag ezt a határt messze túllépi — áttekinthetlenné, kezelhetlenné válik. A bizonyos szemszögből közös vonásokat felmutató elemekre vonatkozó adatok összegyűjtése, az adatok csoportosítása stb. rendkívüli jelentőségű. Az ÉrtSz.-ban végső soron — az etimológiát kivéve — minden adat együtt van: „csak” csoportosítani kellett az ÉrtSz. címszavait szófajok szerint; „csak” a végüktől számítva kellett őket ábécé-rendbe rakni; „csak” az azonos stílusminősítéssel ellátott elemeket kellett egy-egy csoportba gyűjteni, és így tovább. Hasonló a helyzet a fonéma-(betű-) kettesekkel, hármásokkal stb. is; vö. még a 3. pontban ismertetett átrendezést, mint egy lehetséges kombinációt csupán az utolsó betűre vonatkozóan. Két, három, négy betű esetén az ilyen kombinációk lehetősége rohamosan nő, az ezeknek megfelelő rendezések manuálisan épp úgy elvégezhetlenek, mint az említett példák a szófaji stb. csoportosításra vonatkozóan (vagy épp úgy elvégezhetőek ugyan emberi erővel is, de erre mégsem vállalkozott ez idáig senki emberfia. csak a gép — bár igaz, az is az ember fia).

járt így közöltük az 1. táblázaton, az is könnyen belátható: akkor az egész táblázat szétesett volna, lévén, hogy minden egyes etimológiai csoportban más és más a végző fonémák sorrendje, az egyetlen táblázat helyett tehát kilencet kellett volna adnunk. És továbbá: az ábécé-rendben könnyebben megtalálható minden egyes keresett végződés, ám próbáljon valaki egy-egy konkrét végződést megtalálni a nagyság szerinti rendezés esetén: persze meg fog úgy is mindent találni, csak hosszabb-rövidebb keresés után.) E rendezés eredményét (legalábbis, ami a gyakoribb végződéseket illeti — a ritkébbak már természetesen nem voltak jelölhetők) grafikusán a 2. ábra mutatja, szokásos körgrafikonok formájában (a körök nagysága megfelel azoknak a százalékoknak, amelyeket az egyes formák az 1. táblázat szerint az egyes etimológiai rétegekben elfoglalnak).



2. ábra. A különböző szövegző magán- és mássalhangzók eloszlása etimológiai rétegenként. Jelmagyarázat: u_1 szövegző magán- vagy mássalhangzó.

Mutassunk rá az ábra alapján először egy triviális igazságra. Látható, hogy, bár a vitathatatlan tőszavak sorából az *-ancia*, *-(e)um*, *-us* végű latin szavakat kizártuk, ha valaki semmit sem tudna a latin nyelvről, csak a magyar nyelv latin eredetű kölcsönszavait ismerné, akkor is látna legalább annyit, hogy az az alak, amelyben a latin szavak (nyilván főleg főnevek, a tőigék zömének finnugor-magyar elemnek kell lennie: vö. Nyr. 91 : 52) átkerültek, néhány határozott végződést mutat; hasonló szabályosság tapasztalható a szóvég eloszlását illetően a szláv és részben a neolatin kölcsözéseket illetően; elég élesen szembenállnak ezzel a finnugor-magyar-török szóvégek, sokkal egyenletesebb eloszlásukkal („tehát itt meg feltehetően null-végződés van bizonyos bázis-alakokban — amelyekben ti. a szótár hozza a szavakat — és az ez előtt álló tővégre vonatkozóan nincsenek olyan szigorú megkötések, mint a latin, szláv stb. végzödésekre” — állapítaná meg ismét az, aki viszont nyelvünk finnugor, magyar, török elemeiről csak annyit tudna, amennyit az ábrán lát).

Az „Egyéb” feliratú csoportból természetesen már semmi efféle következtetést nem vonhatunk le, itt a dolog természeténél fogva különféle nyelvekből átkerült kölcsönszavak kerültek egybe: a szolid százalékok mögött különben is igen kis tételszámok húzódnak meg, ezekből aligha lehetne valamiféle konklúziót levonni. Így például szúrhat e csoportban a *-z* végzet előkelő helye, mely mögött azonban csupán három szó rejtőzik: *büz* „talán iráni eredetű”, *nemez* „az alánból”, *tíz* „összefügg a *z*-voty. *das* '10' számnévvvel, azonban a hangtani megfelelés...nem szabályos s így alkalmasint mind a *z*-voty., mind a *m*. egymástól független kölcsönzés valamely árja nyelvből” (az idézetek a SzófSz. megfelelő szócikkéből). — Viszonylag nagy mennyiségek húzódnak meg viszont a „Bizonytalan” csoport százalékaikai mögött. Így itt joggal tűnhet fel e csoport végzödései eloszlásának „nem flektáló” jellege (vö. a fentebb mondottakkal), valamint „magyaros” kezdete: mind a magyar belső keletkezésű, mind a bizonytalan eredetű tőszavaink legjellegzetesebb szóvégei, mint látható az ábrán, az *-a*, *-r*, *-k*.

Am itt mindjárt feltűnhet a magyar nyelvtörténet ismerőinek a gyanúsán sok *-a* végű magyar belső keletkezésű szó. Valóban, ha felütjük a megfelelő tételes listát (vagyis azt a listát, ahol nem csupán a számok vannak, hanem tételesen fel vannak sorolva az egyes egyedek is), azt látjuk, hogy e szavak jó része gyanús, *-ca*, *-ka* vagy ehhez hasonló szóvéget tartalmaz, melyet azonban esetükben nem tudunk képzőnek minősíteni s így a tőszavak közé kerültek; hangutánzó, de úgy, hogy szláv mintára keletkeztek; elvonások, elhomályosult összetételek stb. Vagyis: ezek az *-a* végűek, úgy látszik, valóban belső keletkezésűek ugyan, de ez a belső keletkezés az esetek többségében — jól ismert külső hatás közvetlen vagy közvetett eredményeképpen jött létre. Más részük még fiatalabb, a nyelvújítás korában jelent meg, ám azt a képzőszerű elemet, amely az *-a*-t tartalmazza, mégsem tudtuk képzőnek minősíteni s így kerültek ide. De nincsenek is olyan sokan ezek az érdekes elemek, hogy itt fel ne sorolhatnám őket, a szokásos végmutató sorrendben. Íme: *baba*, *liba*, *tuba*¹, *maca*, *cica*, *álca*, *koca*, *poca*, *orca*, *ruca*, *dada*, *handabanda*, *dalárda*, *szófa*,* *guriga*, *balga*, *eszterga*, *vizsga*, *nyafka*,

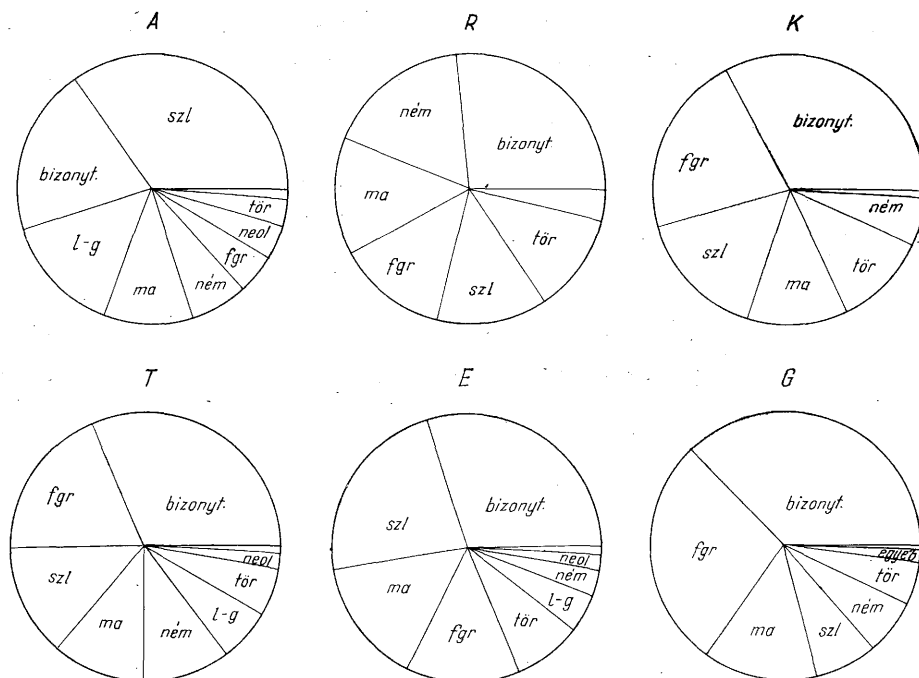
* A felsorolásban *-gal megjelölt szavakkal kapcsolatosan a következőket kell mondanunk: A *sajna*, *cafra*, *hógulya* szavak nem szerepelnek ugyan a SzófSz.-ban, mi azonban mindegyiküket úgy tekintettük, mint olyan származékot, melynek az alapszóval való összefüggése nyilvánvaló a laikus előtt is s ezért nem szerepel külön a SzófSz.-ban

barka⁽¹⁾, *irka*, *firka*, *fruska*, *muszka*, *csúszka*, *kelekóla*, *mama*, *ima*, *szakma*, *adoma*, *lakoma*, *kozma*, *sajna*,* *bálna*, *harsona*, *tárna*, *torna*, *kapa*, *lámpa*, *zúzmara*, *csatara*, *ábra*, *condra*, *cafra*,* *cifra*, *makra*, *zongora*, *bura*, *tusa*¹, *séta*, *vágta*, *vita*, *csalafinta*, *hahota*, *zagyva*, *hebehurgya*, *pocsolya*, *hógulya*,* *banya*, *koponya*, *tivornya*, *kelekótya*, *szittyá*, *kószá*, *kusza*. Azt is mondhatnánk talán, hogy ezek az *-a* végű elemek a legkevésbé „magyarok” a magyar belső keletkezésű elemek között.

Ezzel az *-a* véggel szemben elég markánsan ott állnak ma különféle mássalhangzóra végződő finnugor tőszavaink, valamint a ma szintén nagyobb részt mássalhangzóra (valamint *-a*-ra és *-ó*-ra) végződő török elemeink. (Per-

(vö. BÁRCZI GÉZA „Tájékoztató”-jával: SzófSz X — a *sajna* és a *cafra* ezen belül származék létrehozásért került a tőszavak közé, mert elvonás útján keletkezett, nem képzővel vagy összetétel útján; a *hógulya*, mert második elemét szinkrón szempontból nem tudjuk hová tenni: ez utóbbi szót éppen ezért talán jobb is lett volna nem ide, hanem az *újra* típusú tőszavak közé tenni). A *szófa* 'díszes kerevet, pamlag' valószínűleg tévedésből került ide, mert egyrészt ellenőrzésem során nem találtam meg a SzófSz.-ban, másrészt aligha magyar belső fejlődés eredménye. Itt mégis közlöm, hogy az olvasó fogalmat alkot-hasson magának arról, mit értettünk fentebb azon, hogy a közölt adatok még nem teljesen véglegesek, az ellenőrzések során kijavított hibák azonban sem az összképen, sem annak arányain nem változtathatnak. A *szófa* kijavítása önmagában azt fogja eredményezni, hogy nem 3459, hanem csupán 3458 olyan tőszavunk van az ÉrtSz.-ban, melyet a SzófSz. eredet szerint minősít; eggyel kevesebb lesz úgyszintén az *-a*-ra végződő és a magyar belső fejlődés útján létrejött minősített tőszavak száma is. Amint könnyen ellenőrizhető, mindezek a változások a százalékos értékekben, melyekkel tulajdonképpen dolgozunk nyelvészeti gondolatmeneteink során, egyáltalán nem, vagy csak alig okoznak változást (csupán a század százalékokban — ezek az értékek gondolatmeneteinkben ismét csak nem szoktak szerepelni). A dolog további érdekessége, hogy jó munka esetén az 1. táblázathoz hasonló összeállítások önjavító jellegűek. Amennyiben ugyanis (és ez éppen a jó munka eredménye) nem valamely rendszer szerint tartalmaznak hibákat, tehát nem volt helytelen az az alapkoncepció, melyet — például — az etimológia kódolásánál elfogadtunk, akkor a *szófa*-hoz hasonló hibák rendszerentlenül, véletlenszerűen jelentkeznek s ennek eredményeképpen minél tovább javítjuk a táblát, annál közelebb kerülünk egy ideális és a benne most feltüntetethez közeli eredményhez. Például: a *szófa* miatt eggyel csökkent a végösszeg, de egyszer csak előadódik majd egy szó, melyet innen elhagytunk; ez már aligha lesz *-a* végű és magyar, de a végösszeg ettől függetlenül helyreáll, vagyis újból 3459 lesz. Egy másik esetben kiderülhet, hogy valamint — tévesen — finnugornak minősítettünk, jóllehet magyar: ezzel a finnugor elemek száma csökkent eggyel az itt feltüntetethez képest, de ismét „helyreállt” a magyar elemek itt közölt száma, és így tovább. Általánosságban azt mondhatjuk, hogy az 1. táblázat eredményei a tizedszázalékokig (legrosszabb esetben az egyes nagyságrendű százalékokig) pontosak. Munkatársaink jellemének és képzettségének, az alkalmazott gépek jellegének, a munka egész menetének (például: épp az etimológiát kivéve minden egyes adatot kétszer kódoltunk le egymástól függetlenül, majd a kódokat összeolvasással ellenőriztük; minden egyes adatot a vállalatnál kétszer lyukasztottak le egymástól függetlenül és a második — ún. kontroll-lyukasztás — ellenőrizte az elsőt; több tucat szempontból kontrolláltattuk a géppel magával a lelyukasztott anyagot stb.) ismeretében azt is bátorkodunk állítani, hogy ennél valamivel pontosabb eredmények csak elektronikus gépeken és az emberi tényező sokkal nagyobb kikapcsolásával (sokkal teljesebb automatizálással) érhetők el. (Nem kételkedünk abban, hogy egy nálunk képzetesebb kollektíva, nálunk lelkiismeretesebb munkával pontosabb eredményeket ért volna el ugyanazokon a gépeken is: ám ez a pontosságbeli különbség, okunk van feltételezni, ekkora anyag esetén alig lenne érzékelhető: az elektronikus gépek bekapcsolása viszont már érzékelhető változást okozna.) Ám, ismételjük, azt tapasztaltuk, hogy az ennél nagyobb pontosság nem szükséges tartalmas nyelvészeti gondolatmenetek esetén (a „tartalmas” nem értékítélet, hanem a gondolatmenet jellegére utal). Vö. például ezzel kapcsolatosan a 4. pont alatt közölt hipotézist: ennek érvényén aligha változtat az, hogy eggyel kevesebb lett az *-a* végű magyar tőszavak száma s hogy ugyanakkor — mondjuk — nőtt eggyel a honfoglalás utáni *-a* végű török kölcsönszavaké stb.

sze, a „török” elemekkel baj van: akárcsak annak idején VERMES STEFÁNIA, mi is egy kalap alá vettük e címmel a honfoglaláskor előtti és a későbbi török jövevényszavakat — a mindössze 279 töröknek minősített elem szétválogatása azonban manuális úton is könnyen megoldható: mindenesetre jóval könnyebben, mint ha a „török” elemek egyáltalán nem lennének kiválogatva.) Ezt a



3. ábra. Az egyes etimológiai rétegek elemeinek részesedése a gyakoribb töszóvégekben. (A száz százalékból hiányzó néhány százalékot az *-a*, *-t* és az *-e* esetén az „egyéb”, a többi végződés esetén az egyéb, valamint a latin–görög és a neolatin elemek teszik ki; pontos adatokat l. az 1. táblázaton.)

(mai alakjában) zömmel konsonantikus tövérendszerért érdemes kiemelni: elég élesen szembenáll vele a szláv és latin *-a*-végűség. Ugyanakkor azt sem érdektelen megállapítani, hogy a máig fennmaradt finnugor–török elemek mai alakjukban veláris dominanciát mutatnak (több köztük a veláris, mint a palatális hangrendű).

Hasonló módon rendezhetjük az 1. táblázat egyes oszlopaait is. Ilyen rendezés után szembetűnővé válhat, hogy egy adott α_i végződés esetén a vizsgált töszókincsen belül mely etimológiai rétegek elemei foglalnak el nagy helyet, melyekéi — kisebbet. A 3. ábra egy ilyen rendezés eredményét ábrázolja az első hat leggyakoribb végződésre vonatkozóan (e hat fonéma valamelyikére végződik a vizsgált töszókincsek mintegy a fele). A „Bizonytalan” minősítések kissé elhomályosítják első pillanatra a képet: minden fonéma alatt ők szerepelnek az első helyen, kivéve az *-a* végződést. Ez utóbbi esetben jól látható, hogy nem csupán szláv kölcsönszavaink között volt viszonylag igen

sok *-a* végű, hanem megfordítva is áll a dolog: *-a* végű tőszavaink között (sőt, egész *-a* végű szókincsükben is, tekintettel arra, hogy az egész szótári szókincsben az *-a* végűek aránya már jelentősen csökken: vö. Nyr. 91 : 347) messze a legjelentősebb helyet épp a szláv kölcsönszavak foglalják el. *-a*-ra végződő vizsgált tőszavainknak majdnem a felét a szláv, a bizonytalan és a latin eredetű elemek alkotják; ezt követik a fent felsorolt magyar belső keletkezésűek.

4. Az *-a*-val kapcsolatos megjegyzéseket egyáltalán nem véletlenül tettük. Hosszabb ideje foglalkoztatott már bennünket, egyéb kérdések mellett, az egyes számú harmadik (és vele kapcsolatos néhány más) birtokos személyragos alak kérdése, a *csillagzatja* ~ *csillagzata*, *üvege* ~ *üvegje*-féle diakrón változások, szinkrón ingadozások és párhuzamos alakok problémája. E sorok írásakor még nem dolgoztuk ki és nem ellenőriztük minden részletében feltevéssünket, melyet jelenleg a következőképpen fogalmazhatnánk meg: *Bizonyos tővégi mássalhangzók* (eredetileg: tővégi mássalhangzó + sorvadó mhzó), után a nyelvemlékes kor elejétől nyomon kísérelhető módon azért jelent meg a mondott alakban a *-j*-elem, hogy a mássalhangzóra (eredetileg: mássalhangzó + sorvadó magánhangzóra) végződő tövek birtokos bázisalakja megkülönböztethető legyen a Kárpát-medencében ránk özönlő (főleg szláv eredetű), *-a* végű nem birtokos bázisoktól. A *-j*-elemnek azóta is bázismegkülönböztető szerepe van, nagy általánosságban szólva tehát ott jelenik meg, ahol a bázisvég nem világos és hasonló fonetikai helyzetben nem jelenik meg ott, ahol a bázisvég világos (vö. pl. *türelme*, de *filmje*; *rektora*, de *párja* — jöllehet a *rektor* fiatalabb jövevény, mint a *pár* —; *tagja*, de *boldogsága* stb.: a birtokos személyrag előtt valamilyen módon megváltozott tő éppen úgy, mint a képző vagy a gyakori idegen szóvég egyértelművé teszi a bázisvéget). Feltevéssünket csak megerősítették azok az adatok, amelyek a szóvégi fonémakettősökre, fonémahármasokra vonatkoznak (tehát a tővégi mássalhangzókapcsolatokra, illetőleg az *-a*-t megelőző mássalhangzókapcsolatokra, különösen szláv jövevényyszavainkban). Ezeknek az adatoknak a közlése azonban messze túlnőne jelen közleményünk kitűzött határain éppen úgy, mint arra vonatkozó elképzeléseink, hogy ha egyszer a magyarban sikerült erre a tendenciára, a bázismegkülönböztető tendenciára, rábukkannunk, akkor annak működnie kell más, genetikailag rokon vagy tipológiailag hasonló nyelvekben is, ahol több bázis van.³

Mindez azonban egyelőre, ismételjük, hipotézis — lehet, még sokáig az is marad, csupán tényekkel jobban körülbástyázva, részleteiben publikáláshoz méltóan kidolgozva; végső soron, bármilyen fájdalmas lenne számunkra, az is előfordulhat, hogy el kell vetnünk, mint nem helytállót. Hogy itt mégis, úgyszólván idő előtt, exponáltuk, annak oka távolról sem csupán a friss felfedezés (vagy vélt felfedezés) öröme, mely e sorok írásakor még kétségtelenül áthat e kérdéssel kapcsolatosan. A bázismegkülönböztető tendencia felemelését csupán példának szántuk. Példának arra, hogy a látszólag száraz számok, melyek elrettentő tömegét tartalmazza az 1. táblázat, milyen messzire elvezethetnek önmaguktól nyelvünk belseje, a benne uralkodó egyáltalán nem statisztikai jellegű, hanem a nyelvre mint kódra jellemző strukturális törvény-

³ A kézirat lezárása óta elhangzott a szerző előadása a Magyar Nyelvtudományi Társaság magyar szakosztályában a bázismegkülönböztető tendenciáról 1967. dec. 5-én.

szerűségek felé. Sok mindent elvégeztethetünk gépekkel; az elvégezhetőt velük is k e l l elvégeztetnünk, hogy minden erőnket épp az ilyen jellegű összefüggések feltárására fordíthassuk: ez az ember feladata.

PAPP FERENC

Относительно распределения этимологических пластов венгерских корневых слов по окончаниям в словарной форме

На основе обработки на счетно-аналитических машинах венгерского толкового словаря (A Magyar Nyelv Értelmező Szótára. I—VII. 1959—1962), дополненного данными об этимологии слов на основе венгерского этимологического словаря, составленного акад. Г. Барци (Magyar Szófejtő Szótár. Budapest, 1941), автор рассматривает вопрос о том, какие конечные буквы (соответственно — фонемы) характерны для разных этимологических пластов венгерских корневых слов, т. е. слов несложных и непроезводных. В табл. 1 даны цифровые результаты, на рис. 1 схематично изображается исследуемый материал (это — не все корневые слова венгерского языка, а общая часть двух множеств: множеств корневых слов в толковом словаре и в этимологическом словаре), на рис. 2 — распределение букв (соотв.: фонем) по этимологическим пластам, на рис. 3 — распределение разных этимологических пластов по конечным буквам (соотв. — фонемам). (Легкое сопоставление букв фонемам возможно вследствие характера современного венгерского письма.)

Ф. Пани