

A magyar szókincs lyukkártya-rendszerű feldolgozása

Nyelvtudományunk műhelyében egy elkészítési módját tekintve szokatlan munkán dolgozunk most jónéhányan — nyelvészek, matematikusok, statisztikusok, gépi szakemberek és diákok. Talán a kissé különös készítési mód teszi, hogy a munka célja sem világos sokak előtt, pedig az már, mint látni fogjuk, egyáltalán nem szokatlan. Tevékenységünk néhány elvi kérdéséről szólunk másutt. Itt első, kísérleti jellegű és értékű eredményeinkről szeretnénk beszámolni — ezek az eredmények, úgy gondoljuk, megfoghatóvá fogják tenni egész tevékenységünket, érthetővé fogják tenni, hogy mit is akarunk elérni viszonylag rövid idő alatt (előreláthatólag 1965 végéig).

1. Munkánk célja az, hogy a magyar szótári szókészlet egészére vonatkozóan bizonyos mennyiségi és minőségi mutatókat kapjunk. Kiindulási alapul a nemrég megjelent ÉrtSz.-t vettük. Amikor tehát a továbbiakban „szókészletünk”-ről beszélünk, ezen szókincsünknek az ebben a szótárban rögzített törzsanyagát kell értenünk. Később természetesen semmi akadályja sem lesz annak, hogy szélesebb alpra fektessük kutatásunkat, más szótárak anyagának bevonásával. (Gondolunk itt különösen kiváló, az ÉrtSz.-nál jóval több címszót tartalmazó kétnyelvű szótáraink magyar—idegen köteteire.)

Ugyanígy viszonylag egyszerűen elmondható az is, hogy milyen kérdések tekintetében várjuk azokat a bizonyos „mennyiségi és minőségi mutatókat”. Olyan kérdésekben, amelyek a szavakra mint szótári egységekre (tehát nem mint a szövegek egyes, jól elhatárolható darabjaira) jellemzőek, érdekesek lehetnek. Mi jelenleg a következő jellemzőket vettük fel: összetett—nem összetett szó, (igazi és ál-)homonimák, szófaj, jelentések száma, stílusminősítés, főnév—melléknév—határozószó esetén a tőtípus (*hajó* — mindig változatlan, *alma/almá-* — nyúló, *nyár/nyar-* — rövidülő; összesen 12 ilyen tőtípust sorolunk fel), ugyane szófajok esetén a ragok—képzők alakjainak meghatározása, ige esetén az erős vonzat, végül minden szó esetén — az eredet (finnugor, önálló, török stb. — összesen 9 csoportot vettünk fel Bárczi SzófSz.-a alapján). Mármost e kérdések bármely kombinációjára vonatkozóan megkaphatjuk a kívánt mennyiségi mutatót, vagyis hogy hány adott jellemzővel rendelkező szavunk van. Minőségi mutatón pedig azt értjük, hogy — ha ez szükséges — listászerűen fel is soroljuk ezeket az egyedeket. Hozzunk itt csak egyetlen, igen egyszerű példát. Feltesszük azt a kérdést, hogy szókészletünk szófajonként hogyan oszlik meg. Ennek mennyiségi mutatója: van ennyi meg ennyi ige, ennyi meg ennyi főnévünk stb. A minőségi mutató nem más, mint e szavak szófajonkénti felsorolása egy-egy listán. Igen egyszerű, minden kombináció nélküli kérdést tettünk fel. Ám rendszerünkben épp ilyen értelmes egy olyan, a gyakorlatban aligha előforduló, de könnyen konstruálható kérdésre adandó válasz, mint amilyen például a következő: Hány olyan szavunk van és melyek azok, amelyek a

következő követelményeknek tesznek eleget: *b*-re végződnek, három vagy annál több betűből állnak, kettőnél több jelentésük van, finnugor eredetűek és az ÉrtSz. szerkesztőinek megítélése szerint „durva” stílushoz tartoznak. Ha nem is éppen ez a kérdés (melyre, lehet, az lesz a válasz, hogy ilyen szavunk nincsen), de ehhez hasonlóak a nyelv-művelő munkában éppen úgy, mint a nyelvoktatás során, a szófejtések esetén, egyszerűen magyar fogalmazáskor, fordításkor stb. nyilvánvalóan előfordulhatnak. Csupán merészeknek kell lennünk és minden olyan kérdést, melyet eddig a megoldhatatlanság (pontosabban: az igen fárasztó megoldás) miatt visszafojtottunk magunkban, minden olyan kérdést, amely így bukkant fel bennünk: De jó volna most tudnom, hogy például hány olyan szavunk van és melyek azok, amelyek . . . (és itt következnek egy a fentihez hasonló követelménysor) — ki kell bátran mondanunk, és a választ rá (ha éppen szabad a gép) legfőbb három órán belül megkaphatjuk. Egyes „közérdekűbb” kérdéseket igyekszünk előre látni, s azokat részint könyvalakban, részint kéziratban kiadni. Éppen néhány ilyesféle kérdést és választ kívánunk az alábbiakban bemutatni. Bevezetőül említettük, hogy még csupán kísérleti mennyiséget dolgoztunk fel — az egész anyagnak csupán egy negyvenedrészét. Ha tehát egyes kérdések véglegeseknek tekinthetők is, a rájuk most kapott válaszok igencsak tájékoztató jellegűek: az egész anyag 2,5%-a alapján igen óvatosan szabad következtetnünk magára az egész anyagra, kiváltképp, ha figyelembe vesszük, hogy e minta összeállításakor tudatosan elhanyagoltuk azokat a szigorú követelményeket, melyeket a statisztika tudománya a „teszt” elé állít.¹ Kísérleti anyagunk ugyanis elsősorban nem a teszt szerepét kívánta betölteni, csupán apparátusunk főpróbája volt az, a gépirónóktól kezdve a legbonyolultabb, általunk most alkalmazott gépekig. E főpróbának csupán egy melléktermékét ismertetjük az alábbiakban.

2. Lássunk tehát néhány „előre látott”, közérdekűnek ítélt kérdést és az azokra adott válaszokat. Az egyszerűbbtől fogunk a bonyolultabb felé haladni: először igyekszünk a „legártatlanabb”, csupán a betűkre vonatkozó kérdéseket feltenni, aztán a „rázósabbaknak” vélteket — a stiliztikát, a grammatikát, az etimológiát illetően.

a) Hány b e t ű b ű l állhat egy-egy magyar szó? Megjegyzendő, pontosabban szólva itt nem betűkről, hanem „*n*-ekről”, jelelemekről kell beszélnünk: a kétjegyű *sz* betű nálunk kettőnek számít, a háromjegyű *dzs* — három betűnek és így tovább. (A továbbiak során, elektronikus gép segítségével hívásával, ezen a pontatlanságon, mert kétségtelenül az, segíthetünk majd.) Azt is újból le kell szögeznünk, hogy a szavak s z ó t á r i a l a k j á n a k hosszát mértük. A szövegből vett szavak átlagos hossza ennél nagyobb lesz a toldalékok miatt. A kérdésre a válasz az 1. sz. táblázatból olvasható le.

A másfélezer átvizsgált szó zömét tehát az 5—12 betűsek alkotják. Valószínűleg hasonló lesz a helyzet a megvizsgálandó 60 000 szó esetén is. Már a „szélekre” vonatkozóan bátortalanabban nyilatkozhatunk. Valószínűleg lesz 20 betűsnél hosszabb szavunk is (Ország László említette nekünk a 23 betűből álló *megfellebbezhetetlenség*-et, tehát legalább egy 20-nál több betűs szavunk már biztosan van). Valószínű, hogy az egybetűs szavak aránya, bár az itt különösen gyakori homonimákat külön-külön számítjuk, kisebb lesz az egész anyagban, mint az itt tapasztalt.

¹ Mintánk adatai tehát egészükben nagyobb mértékben fognak eltérni a teljes feldolgozás adataitól, mint az ekkora tesztnél megengedhető volna statisztikailag. Ezen az egészen belül egyes adatesoportok másokhoz képest kisebb vagy nagyobb mértékben közelítik meg a tényleges értékeket. E relatív megbízhatóságra több helyütt fogunk utalni a szöveg során.

Szóhossz betűben	Mennyiség	Szóhossz betűben	Mennyiség
1	10	11	95
2	11	12	70
3	17	13	42
4	37	14	23
5	144	15	21
6	145	16	2
7	209	17	9
8	225	18	4
9	218	19	3
10	178	20	1

1. sz. táblázat

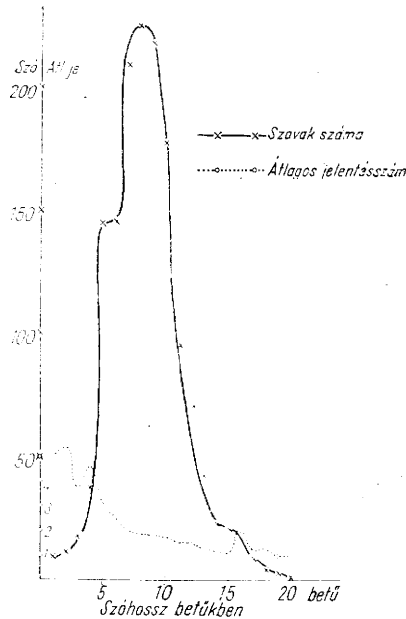
b) Igaz-e az a Zipf-törvény, mely megfelelést állít fel a szóhossz és a jelentések száma között? Mint ismeretes, ez a törvény igen durva megfogalmazásban úgy szól, hogy minél hosszabb egy szó, annál kevesebb jelentése van és megfordítva. E kérdésre a választ a 2. sz. táblázat adja. Itt az „átlagos jelentésszám”-ot a lehető legegyszerűbb és legdurvább módon úgy kaptuk meg, hogy az adott hosszúságú — 1, 2, 3 stb. betűből álló — szavak jelentéseit összeadtuk és az eredményt osztottuk az adott hosszúságú szavak mennyiségével. Tehát például: mint láttuk, volt összesen 10 egybetűs szavunk. Ezek között volt 4 háromjelentésű, 3 négyjelentésű, 2 ötjelentésű, 1 pedig 16 jelentésű (a mi mintánkban ez az α névmás—névelő volt). A jelentések száma tehát összesen 50, ez osztva tízzel — 5, ez az egybetűs szavak átlagos jelentésmennyisége mintánkban.

Hossz	Átl. jelentés	Hossz	Átl. jelentés	Hossz	Átl. jelentés
1	5,0	7	2,0	14	1,2
2	5,3	8	1,9	15	1,1
3	3,7	9	1,8	16	2,0
4	4,6	10	1,7	17	1,2
5	3,1	11	1,5	18	1,3
6	2,6	12	1,5	19	1,0
		13	1,3	20	1,0

2. sz. táblázat

A feltett kérdésre kapott válasz még plasztikusabbá válik egy grafikonon (l. 1. sz. ábra). E grafikon ábrázolja, hogy adott hosszúságú szóból mennyi van anyagunkban, és ugyanakkor azt is, hogy az ehhez a hosszúsághoz tartozó szavak átlagos jelentésszáma mennyi. Szemmel látható, hogy a Zipf-törvény gyönyörűen „kijön” a középső zónában, vagyis azoknak a szavaknak az esetén, amelyekből sok van; ugyanakkor jelentős ingadozás tapasztalható a két szélén, az egészen rövid és az egészen hosszú szavak esetén. Vajon azért van-e ez, mert ezekből a szavakból természetesen kevés van s itt az e s e t l e g e s eltéréseknek nagyobb a szerepük, vagy azt kell majd mondanunk, hogy a Zipf-törvény csak a „középső zónában” érvényes, a két szélén más törvényszerűségek uralkodnak? Ilyen kis anyag alapján még nem érdemes megkísérelni a választ erre a nemzetközileg is vitatott kérdésre. Amikor azonban együtt lesz a 60 000 egyedet tartalmazó anyag, erre alapozva olyan biztonsággal válaszolhatunk erre a kérdésre, amilyennel eddig — gépek híján — a Zipf-törvény egyetlen bírálója sem tehetette. (Természetesen maga Zipf sem.)

c) Milyen betűvel végződhet egy magyar szó szótári alakjában? Úgy tetszhet, pontosan a magyar nyelvben ez teljesen „mindegy”, más szóval: nem tudunk olyan végződés szempontjából „terhelt” betűkről, mint amilyen például a latinban az *-a* vagy a *-s* (az *-us*-ból) stb. Anyagunk mást mutatott. A részletes válaszcímzésével itt csupán arra mutatunk rá, hogy összes szavaink 43,54%-a a következő betűk valamelyikére végződött: *k* (7,47%), *a* (8,73%), *t* (11,07%), *s* (16,27% — ebből közel 1%-ot a *cs*, *zs* végűek tesznek ki). Márpedig, ha szavaink közel fele négy betű valamelyikére végződhet, ez azt mutatja, hogy a magyar nyelvben nincs olyan „rendetlenség”, mint az első pillanatban vélhető lenne; érdekes volna megnézni, hogy például éppen a



1. ábra

latinban van-e ténylegesen akár csak ekkora rend is e szempontból. S ehhez az adatunkhoz hozzá kell tennünk még azt, hogy viszonylag pontosnak tekinthető bizonyos okoknál fogva: a jelzett százalékos aránytól való eltérés várhatóan a legkisebbek között lesz az egész anyagban tapasztalt arányhoz képest. Mintavételünk pontatlanságából fog adódni például, hogy feltehetően valamivel több lesz az *-e*-re végződő szavak aránya, de az *-a*-val még akkor sem fog versenyezni ez a szám, vagy ez legalábbis kevésbé valószínű.

d) Milyen betűvel végződhetnek az egyes szófajok? Itt a legérdekesebb (bár, igaz, statisztikailag a legkevésbé megbízható) képet a melléknév mutatja: a vizsgált melléknév zöme *s*-re (67 db), *i*-re (29 db), *ó*-ra (11) és *n*-re (10) végződik — szemmel láthatóan egyes gyakori képzők miatt. Főneveink legkedveltebb végződése szintén a *s* (179), illetőleg a melléknévknél kevésbé „kiaknázott” *a* (108), meg — ismét egy képző miatt — a *g* (53). Igéink túlnyomó többsége szótári alakjában *l*-re (75), *k*-ra (66), illetőleg *t*-re végződött (62); e megoszlás függése a képzőktől (illetőleg az *-ik* ragtól) szintén kézenfekvő.

e) De egyáltalán: vajon milyen arányban tartalmazza a szókészletünk az egyes szófajokat? Anyagunkban 275 ige, 861 főnév, 158 mellék-

név, 5 névmás, 23 határozószó, 2 névutó, 3 kötőszó, 5 indulatszó, 5 módosító szó volt. A két-, illetőleg többszófajúak legnépesebb csoportját a főnév-melléknevek, illetőleg a melléknév-főnevek alkották (mint *eszkimó, takarodó, vasutas* stb. — a végső kidolgozásban e két kategória között az ÉrtSz.-nak megfelelően különbséget fogunk tenni), összesen 109 egyeddel, de volt egy ige-melléknevünk (*tartozik*), egy főnév-határozószónk (*vasárnap*), egy indulatszó-főnevünk (*abcúg*) stb. is. Ezekről az arányoktól az egész anyagban közepes értékű eltérés várható.

f) S ha így észrevétlenül áttértünk a grammatikára: Mit mondhatunk példának okáért nyelvtani szempontból a legnagyobb számban képviselt főnevekről?

861 főnevünk túlnyomó többsége (667 egyed) az 1. tőtípushoz tartozott, vagyis nem változik a ragozás során (mint: *hajó, ablak, vártömlő* stb.). Nem kis öröme szolgálat ez a tény mind a magyarul tanuló idegen ajkúnak, mind a magyarra fordító gépnek. További 116 esetben a tő a végződés legelőbbje előtt egyszerűen csak megnyúlik (*alma, körte*). Elenyésző ezek után azoknak a főneveknek a száma, ahol egyéb változások történnek. Persze, tekintetbe véve azt a körülményt, hogy minden egyes általunk nyert adatot — nagyon durván szólva — negyvennel kell megszorozni ahhoz, hogy az egész szótárban várható mennyiséget megközelítőleg megkapjuk, valószínű, hogy a „ritkább” tőváltozástípusok egy részét sem lesz érdemes egyszerűen a kivételek listájára tenni, hanem egy-egy nyelvtani szabályt kell alkotni rájuk. Végső soron nyilván a célszerűség fogja eldönteni, hogy mit tekintünk még „grammatikailag kezelhető”-nek, és mit grammatikai „kivétel”-nek (vagyis a szótárba utalandónak).

Am a fő problémát — a tő általunk vallott felfogása mellett, melyre itt nem kívánunk kitérni, de mely az előzőkből kitapintható volt — nem is annyira a tőtípus meghatározása fogja okozni, hanem az, hogy a tőhöz milyen alakban járulnak az egyes ragok. A kérdés nehezebbik fele tehát így hangzik: A legtöbb esetben változatlan tőhöz mikor kell *-t, -at, -et, -ot*, illetőleg *-öt* ragot illesztem; a többes szám végződése mikor *-k, -ak, -ek, -ok*, illetőleg *-ök*; az egyes szám harmadik személyű birtokos személyjel mikor *-a, -e, -ja*, illetőleg *-je*. Az ÉrtSz. szerkesztőinek tapasztalata szerint e három alak ismeretében a többi 711 (ha elfogadjuk az Antal javasolta 714-et, mint egy magyar főnév összes paradigmatis alakjainak számát) automatikusan képezhető. Ha tehát e kérdésre megadtuk a választ, a szótártól teljesen elszakadva, egy szigorúan vett grammatikai kérdésre válaszoltunk. Mégpedig arra, hogy milyen szabályok adhatók a magyar főnév ragozására vonatkozóan. Maga a gép által nyújtott lista természetesen csak lista lesz; a grammatikus rendszerező kezének és eszének kell abban felfednie a törvényszerűségeket; e törvényszerűségek egy részét (vagy jó részét? egyelőre ezt még nem tudjuk) már ismerjük is, különösen egyes szótárak magyar nyelvtani kompendiumaiból, meg egyes idegen ajkúak számára készült magyar nyelvtanokból. Igen messzire vezetne, ha anyagunk alapján most itt kívánnánk erre a kérdésre válaszolni, ezért csak egyetlen, apró és jelentéktelen példával elégszünk meg. — A magyar főnév automatikus szintézisének algoritmusain dolgozva felvetődött az a kérdés, hogy vajon minden egyes palatális ragozástípusnak megfelel-e egy veláris és megfordítva. Úgy találtuk, hogy van egy palatális séma, amelynek nincs veláris megfelelője, de nincs olyan veláris séma, amelynek palatális megfelelője ne volna. A „pár nélküli” palatális sémára mindössze egyetlen példát sikerült találnunk: *modell* — *-t, -ek, -je* (vagyis *modellt, modellek, modellje*). Furdalt bennünket a lelkiismeret, vajon valóban csak egyetlen és még hozzá ilyen ritka (az ÉrtSz.-ban nem is szereplő) példánk van-e erre, és főleg, hogy vajon valóban nincs-e olyan főnév, amely a *-t, -ak, -ja* komplexumot kapja (ez utóbbi lenne ugyanis a *modell* veláris „párja”: a tárgyeset végződése előtt nincs kötőhangzó, az *-e-k* helyén meg *-a-* áll). Kísérleti mennyiségi anyagunk mindkét vonatkozásban megnyugtatót bennünket. Még két, a *modell*-hez hasonlóan viselkedő főnevet találtunk benne: a *pelerin* és az *eszer*

szavakat, ezek egyike sem szokottabb az általunk talátnál érzésünk szerint. Veláris pár meg nem akadt az anyagban — ha lesz is hát, feltehetően kevés lesz a „teljes” anyagban is belőle¹.

Még egyetlen érdekességet ezzel kapcsolatosan. A ragokból természetesen egyértelműen kiderül, hogy valamely főnév veláris, vagy palatális hangrendű-e. Úgy találtuk, hogy a *veláris* főnevek száma közel kétszerese a palatálisakénak. A teljes anyagban bizonyos okoknál fogva elég jelentős eltérés mutatkozhat majd ehhez képest a palatálisok javára.

g) Vajon hogyan bántak az ÉrtSz. szerkesztői a stílusminősítésekkel? Mint ismeretes, 28 különféle stílusmegjelölést alkalmaztak hatalmas munkájukban. Bizonyos alapelveken túl egységes elbánásmódra nem törekedhettek, ők maguk sem tudhatták gép híján, végül is mely szavak kerültek ebbe vagy abba a stílusrétegbe. Így, remélem, számukra is újdonságként közölhetem, hogy anyagunkban — és ettől a teljes anyagban közepes eltérés várható csupán — a szavak 73,3%-a esetén a „fejben”, vagyis összegező stílusminősítésként semmiféle megjegyzéssel sem éltek. (O. Nagy Gábor volt szíves felhívni a figyelmemet arra, hogy ez nem egy esetben egyáltalán nem azért van, mintha a szó valóban teljesen semleges volna. Éppen ellenkezőleg: annyira különböző stílusértéke van egyes jelentéseinek, hogy nem lehetett őket „egy kalap alá hozni”. A továbbiak során erre a körülményre figyelni fogunk, bár véglegesen megoldani nem fogjuk tudni.)

¹ Más szavakkal: Munkánk során a következő két kérdésre kellett választ találnunk: a) Melyek azok a főnevek, amelyek tárgy esetben *-t*, többes számban *-ek*, egy. 3. birtokos alakban *-je* végződést kapnak? b) Vannak-e olyan főnevek, melyek *-t*, *-ak*, *-ja* végzódéseket kapnak? Lyukkártya-rendszerű feldolgozás nélkül nehezebb megfelelni különösen a második kérdésre, mint egy nívós keresztretjtvény kérdéseire, pedig annak megválaszolásától igen sok függött algoritmusaink felépítése során. Mintánk vizsgálata rögtön kimutatta, hogy az egész ÉrtSz.-ban néhány tucat *modell*-típusú főnév rejtőzhet (s azokat mind, vagy nagyrészt a szonánsra — de nem *m*-re — végződő kölesónszavaink között kell keresnünk). — E típus veláris párja pedig vagy üres halmaz, vagy nem tartalmaz tíznél sokkal több egyedet. Ez az eredmény azért volt rögtön kiolvasható a feldolgozásból, mert az egyik táblázatban a főnevek ragjaik alakja szerint vannak felsorolva, a következő kulcs szerint: A „Ragok” rovat első két oszlopa a tárgyeset alakját jellemzi, mégpedig: *-t = 01*, *-at = 02*, *-et = 03*, *-ot = 04*, *-öt = 05*; a második két oszlop tartalma: *-k = 01*, *-ak = 02*, *-ek = 03*, *-ok = 04*, *-ök = 05*; a harmadik két oszlop: *-a = 01*, *-e = 02*, *-ja = 03*, *-je = 04*. Világos, hogy a (*modell*) *-t*, *-ek*, *-je*-féle szavakat a „Ragok” rovat 010304 száma alatt kell keresnünk: ott meg is találtuk a fent idézett két szót (*eszer*, *pelerin*) és csak ezt a kettőt. Világos, hogy ezt megelőző ragszámú szó (10 303) nincs a magánhangzó-illeszkedés miatt (az 10 303 számnak ui. a fentiek értelmében a következő ragok felelnének meg: *t* — ez az 1 — *-ek* — 03 a második helyen — és *-ja* — 03 az utolsó két helyen). Az *eszer*—*pelerin*—*modell* csoportot megelőző szavak tehát 10 302 ragszámúak: ez az *ember* (*-t*, *-ek*, *-e*)-féle főnevek népes csoportja. A *modell*-félék szomszédai a másik oldalról (növekvő ragszámmal) a 10 305-ösök volnának, ilyenek azonban nincsenek, mert a birtokos személyjel összesen csak 4 alakot vehet fel, az utolsó helyen 5-ös tehát nem állhat. Következik elméletileg a 10 400: a végső két 00 mutatja, hogy itt olyan — ritka — szavaknak kellene állniuk, melyeknek nincsen birtokos személyjeles alakjuk. És így tovább egészen a legmagasabb, 50 504 számig (tehát *-öt*, *-ök*, *-je* — pl. *görög*). A legalacsonyabb ragszámmal viszont a plurale tantumok szerepelhetnének, hiszen ott az egyes accusativus helyén 0 (ill. 00) áll.

Példaként arra, hogy mik „kerültek össze” egy-egy közös stílusminősítés alá, felsorolom anyagunk *biz* (*bizalmas*) jellel ellátott szavait. Írme: *bakfis, balek, ellóg, fél-eszű, guberál, gulyáságyú, habarodik, hibbant, lagzi, lábatlankodik, ledoktorál, ledorongol, loespocs, lódítás, lófrál, öregúr, pemzli, pesztonka, pénzmag, szaki, taksa*. „Durva” jellel ellátott szavainkat nem is merem felsorolni (köztük van anyagunkban, összetétel utótagjaként, két ízben is, egy egytagú szavunk, melynek finnugor eredete nemrég nyert teljes bizonyítást.) „Hivatalosak” viszont, példának okáért, a következők: *abbeli, baleset-elhárítás, ellenindítvány, ellenintézkedés, ellenjegyez, ellennyugta, eszközés, kvóta, lófogatú, négyzetcentiméter, négyzetméter, pénzellátás, szakbizottság, társadalombiztosítás*. — Az ÉrtSz. jó ismerői bizonyára rájöttek e példák némelyikéből arra, hogy ahol két vagy több stílusminősítés is volt egyszerre a fejbén, ott csak az elsőt vettük figyelembe; a továbbiakban ezen még fogunk tudni némileg finomítani. Megjegyzendő, egy megbeszélés során elhangzott olyan vélemény, hogy a stílusminősítést egyáltalán ne vegyük figyelembe, annyira szubjektív. Mi ellenkező véleményen vagyunk, s e véleményünket a fentiekkel igyekeztünk alátámasztani.

h) E r e d e t szempontjából a következő megoszlást tapasztaltuk mintánkban.

A nem összetett szavak esetén e szavak 23%-a nem szerepelt Bárczinál, 17,6%-a finnugor eredetű, 18,5%-a önálló (belső magyar fejlődés eredménye), 3,5%-a török, 6,6%-a szláv, 3,5%-a német, 1,9%-a latin-görög, 0,3%-a neolatin, 1,2%-a egyéb, 23,7%-a bizonytalan.

Ha az összetételekben szereplő töveket vizsgáljuk (maguk az összetételek már mind „önállóak”, vagyis belső magyar fejlődés eredményei, egészükben tehát nem érdemes jellemezni őket), ettől némileg eltérő képet kapunk. Nevezetesen: itt jóval kevesebb az olyan elem, amely Bárczinál nem szerepel, főként a finnugor eredetűek, kisebb részt az „önállóak” kategóriája javára.

Meg kell jegyeznünk, hogy elhangzott olyan tekintélyes vélemény is: ne vegyük figyelembe munkánk során a SzófSz. adatait, hanem az etimológiát egyáltalán ne érintsük, vagy a készülő új etimológiai szótár anyagát dolgozzuk fel lyukkártyákra. Nektünk az a véleményünk, hogy az etimológiát kár volna figyelembe nem venni. Érdekesek ezek az eredmények önmagukban is, különösen, ha például összevetjük őket azokkal a statisztikai adatokkal, amelyeket Vermes Stefánia magának a SzófSz.-nak az egyszerű statisztikai feldolgozásával kapott (vö. NyK. LI, 435). Hiszen míg Vermes adatai lényegében tőszókinésünkre vetnek világot, a mi adataink azt fogják mutatni, hogy e tőszókinés hogyan „kamatozott”, hogyan ágazott szerte a magyar szókinés egészében (ezzel kapcsolatos már például fenti megjegyzésünk az összetételekre vonatkozóan). De meg jó lesz az etimológia különféle kombinációkban is, amikor például egy bizonyos végű *e r e d e t* ű szavakra vagyunk kíváncsiak. — Ami másrészt a SzófSz. adatainak bizonyos mértékű elavultságát illeti, ez kétségtelen. Viszont ma ez az egyetlen klasszikus standard munka, mely „az asztalon hever”. Mihelyt megjelenik egy jobb nála, gépi vagy emberi munkával könnyűszerrel korszerűsíthetjük majd az anyagot; ezt nem tudnánk ilyen könnyen megtenni, ha nem ragaszkodnánk mereven egyetlen lezárt forráshoz, mint amilyen a SzófSz. Másrészt azt is szeretnénk remélni, hogy munkánk éppen a jelen etimológusainak is fog segíteni abban, hogy az a készülő, korszerű etimológiai szótár könnyebben, gyorsabban, elegánsabban elkészülhessen. Ezért vesszük figyelembe az etimológiát is, és ezért vesszük azt éppen a SzófSz. alapján figyelembe.

3. Munkánknak eredetileg a Szóvégmutató Szótár elnevezést adtuk (Kiss Lajos névadó javaslatát téve a magunkévá). Valóban, e munka ez is lesz: a magyar szavakat jobbról számított ábécé rendben is fogja közölni. Ez igen sok szempontból célszerű, sőt célszerűbb, mint a balról számított ábécérend. A célszerűség taglalásába nem kívánunk

belemenni, egyszerűen közlünk itt két listát „a tergo”, vagyis jobbról számított betűrendben — úgy, ahogy azt a gép kiadta. Íme; a *g* betű egy része:

lagymatag, vastag, vastagság, balság, lágyság, feleség, ellenség, édeség, balog

Részlet a *k* betűből:

rarak, lágység, ellenségeskedik, édesedik, vastagodik, lábatlankodik, elmagyarosodik, rarakódik, edződik, ellik, ellenkezik, lábadozik, eszik, feleszik, felesküzik, ellök.

Úgy gondolom, e két felsorolásból az „a tergo” sorrend néhány előnye szembezőkő; további előnyökre is fény derül a munka során.

4. Nem lehetünk olyan hálátlanok, hogy e munkát egyik főszereplőjéről, az általunk alkalmazott „okos” gépekről ne szóljunk, ha csak egy-két szót is.¹

Lyukkártya-rendszerű elektromechanikus (tehát nem elektronikus) gépeken dolgozunk. Ezeknek a gépeknek az elve igen egyszerű. Egy-egy tétel (esetünkben egy-egy szó a rá vonatkozó mindenféle szótári információval) egy-egy kartonkártyára kerül lyukasztás útján. A kártya 80 függőleges oszlopot tartalmaz, minden egyes oszlopon (nullával együtt) tíz számjegy következik egymás alatt. Ha megállapodtunk abban, hogy a szófajt — mondjuk — a balról számított 42. oszlopban fogjuk jelölni, továbbá abban, hogy például az ige ebben az oszlopban = 1, a főnév = 2, a melléknév = 3 és így tovább, akkor a következő módon dolgozunk. Egy ilyen módon „kódolt” kézirat alapján (ahol tehát például a szófajok esetén az egyes szavak után a fent jelzett számok állnak) egy gépkezelő „legépli” a kéziratot; — csakhogy nem papírra, hanem az említett kártyákra, és ezeken a kártyákon nem betűk lesznek, hanem a megfelelő oszlopban a megfelelő helyen lyukak. A gépkezelő munkája itt alig különbözik a gépíróétől: valamivel kényelmesebb neki, mert kevesebb előtte a billentyű, meg a gép maga gondoskodik a kártyák betolásáról (nem a kezelőnek kell „papírt cserélnie”). A kilyukasztott kártyákkal — számos ellenőrző operáció után — a gép dolgozik.

Számos operációt mellözve, utalok arra, hogy a gépkomplexum egyes tagjai nem rendeznek, hanem „tábláznak”. Még hozzá elég sebesen. Egy legújabb típusú gép, melyből néhány darabot nemrég kapott hazánk, óránként 20—24 000 sort képes kiírni. Más szóval: az egész ÉrtSz.-t egy előzőleg végrehajtott tetszőleges rendezésben nem egészen 3 óra alatt leírja, természetesen mindenféle információval (utalás szófajra, eredetre stb.) együtt.

Így, bár a feldolgozás szerkesztőjeként e sorok írója fog szerepelni, megköszönve számos munkatársa szíves segítségét — szerzőként vajon nem a már nem fiatal, de még mindig életerős, munkabíró, a korral haladó, az elektronikával nem konkurráló, hanem azt szépen kiegészítő lyukkártya-rendszerű gépeket kell-e majd megneveznünk? Én a magam részéről, mint szerkesztő, okvetlenül e kissé szokatlan szerzőség mellett szavaznék, s most ez úton is meg szeretném köszönni az általam szerkesztett szerző rugalmasságát, pontosságát, gyorsaságát, maximális megértőkészségét a szerkesztő iránt.

Papp Ferenc

¹ Munkánk a gépeken csak azért vált lehetségessé, mert az Országos Ügyvitelgépészeti Felügyelet igazgatója, Botka Zoltán igen nagy készséggel felkarolta próbálkozásunkat, mindjárt az elejétől kezdve tisztánva látjelentőségét. A kísérleti anyag gépi feldolgozásának programjait Molnár Miklós, a KSH dolgozója készítette el, nélküle sem tudtunk volna egyetlen lépést sem tenni e téren. Úgy gondoljuk, mindkettejüknek méltán lesz hálás a hazai, de a külföldi filológia is, hiszen e munkát sok tekintetben úttörő jelentőségű (legalábbis Európában — de némileg annak határain túl is).