

SZEMÉLYI SZÁMÍTÓGÉP ÉS FONETIKAI-FONOLÓGIAI KUTATÁSOK

Papp Ferenc

MTA Nyelvtudományi Intézete, Budapest

PC-ket nevezett kutatásokra több módon lehet alkalmazni. Itt két irányra szeretnék rámutatni: azoknak a szavaknak a teljes listázása, melyek ún. archifonémákat meg azonos magánhangzó fonémakettősöket tartalmaznak (1.); a hosszú valamint az álhosszú többjegyű betűket tartalmazó címszavak listájának hasonló összeállítása (2.). Példáim mindkét esetben az ÉrtSz.-re, tehát egy lexikográfiai korpuszra fognak korlátozódni. Természetesen semmi akadálya sincs annak, hogy ugyanezt a munkát szövegeken, így az MTA Nyelvtudományi Intézete Nagyszótári Osztályán készülő hatalmas szöveg-thészauruszon elvégezzük. Esetenként utalni fogunk erre a lehetőségre s annak sajátosságaira. Megjegyzem még, hogy a két irányt egyesíti azok hasonló felszíni – tehát pl. számítógépes – megoldása: adott karaktertöbbséget keresünk a file-ban: ez természetesen rutinfeladat (miután KG a technikai-programozási kulcsot megtalálta hozzá).

1.1. gyöngy, pinty. Tekintsük meg ezt a két, itt fettel szedetett tőszót, az 1. sz. függelékben bővebben adott válogatást. Bennük természetesen nem az a probléma, amit a HSzab.¹¹ említ, hogy ti. az egyszerűsítés elve érvényesül leírásukban, bár bizonyos értelemben ez a primitív és félrevezető meghatározás is utal valamire. Nevezetesen: ilyen esetekben kétségtelenül [ɲ] hangot ejtünk: gyö[ɲɟ], pi[ɲc]. Ezt egyszerűsítve "n" betűvel írjuk – ámde jól tudjuk, hogy betűkkel nem hangokat, hanem fonémákat jelölünk. Milyen fonéma van ezeken a helyeken, /n/ vagy /ɲ/, ennek megfelelően mit kell itt írunk, "n" vagy "ny" betűt? A természetes ejtéshez (hangokhoz) közelebb álló fonológiai irányzatok számára olyan fonéma van, amilyen hangot ejtünk – esetünkben tehát / /, eszerint "ny" betűt is kellene írunk. Más irányzatok megkövetelik, hogy a kérdéses hangot szabad helyzetében vizsgáljuk. Amilyen hangot ott hallunk, olyan fonémával van dolgunk: *vasgolyó* – /j/ fonéma, mert szabad helyzetben [vɔj], s-sel; *vagongyár* /n/ fonéma, mert [vɔgon] [n]-nel stb., függetlenül az aktuális [ʒ], [ɲ] stb. ejtéstől.

Könnyű volt a helyzetünk a fenti esetekben s a hozzájuk hasonló összetett szavakban: az összetételt, képzős alakulatot részeire bonthatjuk s így "kiugrat-hatjuk" az előtag (a tő) végső mássalhangzójának szabad karakterét. De mit tehetünk a fejezet címébe foglalt s ahhoz hasonló tőszavakban? Hiszen ezek azért tőszavak, hogy felbonthatatlanok, nincs mit kiugratni: a [ɲ] ezekben bizony mindig

csak [ɲ], nem tudhatjuk, hogy a /n/ fonéma realizációi-e, avagy a ɲ-éi. (Ezért kvázi mindegy, hogy mit írunk, ezért lehet éppen itt "egyszerűsíteni" és "n"-et írni és nem az ezer és ezer más helyen, ahol persze kötelezően "ny"-et vetünk papírra-táblára. Ez egyeseknek borzasztó, de így van: a magyar helyesírás elég következetesen fonématis. Csak "elég következetesen" és nem teljesen – egy természetes úton kialakult rendszertől ne várjunk természetellenes, teljes következetességet.) (Ezt a kérdést l. még Hell kandidátusi értekezésében: kézirat [1983], 127, 131.)

Van ennek a, mint látjuk, éterien elméleti kérdésnek egy nagyon gyakorlati vonatkozása. Nevezetesen: a helyesírás, kivált a magyar gyermekeké. (Másutt a magyar gyermekekkel egy sorban fogjuk emlegetni a magyarul tanuló külföldit. Itt ezt nem tehetjük, legalábbis ha könyvből mesterségesen tanuló külföldit tartunk szem előtt: az ilyen idegenajkú előbb tanulja meg a betűt, s végig az marad az uralkodó számára.) Milyen hibákat vét a magyar kisiskolás a *gyöngy*, *pinty*-féle szavakban? Amikor a társadalmilag helyes fonématis írásmódot beléverjük az iskolában – mennyire verjük ki belőle a jó fonetikus hallást? Vagy inkább az a helyzet, hogy mire ezek a ritkább szavak gyermekünk elé kerülnek, az már sziklaszilárdan ismeri a társadalmilag elvárt norma elveit, s ezért – a magyarban – megszűnt becsületesen hallani? A gimnáziumra alkalmasint az a helyzet, hogy e szavak nem okoznak helyesírási problémát?

Éppen a magyar helyesírás fonématis volta miatt még néhány hasonló elem megbújhat a *ndj*, *ntj* stringek képe alatt is. Ezek nem lévén túl számosak, itt fel is sorolom őket. *ndj*: *mindjárt*, *mindjobban*, *rendjel*; *ntj*: *elrikkantja magát*, *szentjánosáldás*, *-bogár*, *-kenyér*, *szintjel*, *-zés*, *-ő*. De ezek az ÉrtSz.-címszavak talán csak egy kivétellel (*mindjárt*) egészen nyilvánvalóan, szigorúan szinkron szempontból is összetettek vagy kvázi szövegből vett nagyobb darabok. Az előbbiekre nyilván más szabályok érvényesek (vö. legutóbb pl. Deme 1992, 30), az utóbbiak szinte maguktól felhívják a figyelmet arra, hogy szövegtörzsekben nem csupán más statisztikus törvényszerűségek fognak uralkodni (ez természetes), de más grammatikai-morfológiai helyzetek is fognak szerepelni (*mondja*, *pontja*; [minden] *szentje* stb.), emiatt szövegekben várhatóan gyakoribbak is. Akár így, akár úgy alakul azonban, a szövegek prezentálta ilyen stringek feltehetően nem lesznek archifonématisak – így nevezhetjük a *gyöngy*, *pinty*-féle feloldhatatlan tőszói alakulatokat –, alkalmasint mind átlátszó morfológiai okoknál fogva, a fonématis elveknek megfelelően írónak így vagy úgy.

1.2. Korpuszainkban, lexikográfiaiakban éppenúgy, mint élő szövegen, előfordul, hogy két azonos magánhangzó-fonéma kerül közvetlenül egymás mellé (*alálás*, *beebédél* stb. – 1. 2. sz. függelék). Hogy itt két fonémáról van szó, azt világosan mutatja mind a 15 milliónk nyelvérzéke még akkor is, ha gyors, hanyag ejtéssel egészen egybe is rántjuk a két (azonos) magánhangzó-elemet¹. Az

osztatlan hangpatak fonémákra bontása éppen a szociális faktor dolga, az nyelvenként-hagyományonként esetleg más és más lehet, ezért hivatkoztunk itt a 15 millió magyarra, hazánk határain belül és kívül. Ámde ha megnézzük Olasz Gáborék műszerrajzolta görbéit, ijedten látjuk, hogy e helyen nincs szakadás, amint várnók – valami tehát társadalmilag (fonológiaiilag) van is, meg – fizikailag – nincs. (Vö. Olasz 1989, 63.) Vagyis a "Fonnyadt babarcú öregasszony" és a "Gyönyörű, babaarcú fiatalasszony" homofónok, de nem homográfok – csak ennyi lenne köztük a különbség? Vajon találunk-e még ehhez hasonló minimális párokat szótári vagy szöveganyagban?

Itt közlöm az ÉrtSz. e kettősöket tartalmazó címszavainak teljes táblázatos áttekintését; a hivatkozott függelékben hozom a népesebb teljes listáikat.

fon. kettős	Σ	100=183	lista
aa	47	25,68	babaarc, bábaasszony, bérmaanya, bérmaapa, faanyag, faarc, hasaalja... <i>teljes listát l. a 2.sz. függelékben</i>
áá	5	32,73	aláás, hozzááll, hozzáállás, rááldoz, rááll
ee	74	40,44	bcebédél, beecsetel..., becegyezés, bünbeesés, csereeszköz, csirkeeszű, lcmel, ... zeneegylet <i>teljes listát l. i. h.</i>
éé			0
ii	35	19,13	antiimperialista, férffing stb. <i>teljes listát l. i. h.</i>
íí			0
oo	10	5,46	kooperáció, kooperál, kooptál, koordináció, koordinál, koordináta, koordináta-rendszer, mikroorganizmus, zoológia, zoológus
óó	4	2,19	állóóra, blokkolóóra, disznóól, fogadóóra
öö			0
őő	6	3,28	előőr, erdőőr, mezőőr, úttörőőr, védőőrizet, védőőrség
uu	2	1,09	individuum, vákuum
úú			0
üü			0
űű			0
	183	100,00	

2. asszony, csecsszopó: hosszú és álhosszú magyar mshzó fonémákat tartalmazó címszók kigyűjtése. Az ÉrtSz. keretei között ezt megtettem a szegedi előadásomban, itt onnan kiemelem a fő táblázatot:

	/XXX/	/X-XY/	Σ /string		/X-XY/ példa/teljes felsorolás
ccs	88,89 40 2,44	11,11 5 2,75	45	2,47	arccsont, gerinccsatorna, -cserép, harccsoport, tánc-
ggy	41,84 41 2,50	58,16 57 31,32	98	5,38	anyaggyűjtés, bélyeggyűjtemény, ... foggyökér, ideggyenge, jéggyár, joggyakorlat
lly	100,00 31 1,89	0,00 0 0,00	31	1,70	0
nny	93,57 160 9,77	6,43 11 6,04	171	9,40	agyonnyom, ellennyugta, ... istennyila, ... különnyomat, negyvennyolcas, színnyomás ...
ssz	92,18 1250 76,31	7,82 106 58,24	1356	74,51	adásszünet, ... gyorszámoló ... [79], csecsszopó [13], ésszerű [5], törzsszám [5], össz-szövet-ségi [4]
tty	100,00 116 7,08	0,00 0 0,00	118	6,37	0
zsz	0,00 0 0,00	100,00 3 0,65	3	0,16	halászsákmány, horgászszing, vadászsákmány
$\Sigma/..$	1638 90,00	182 10,00	1820	100,00	

És külön a (d)dz(s) stringgel kapcsolatos adatok:

	/dz/	/d-z/	Σ /string		teljes felsorolás
dz	95,28 121 79,08	4,72 6 100,00	127	79,87	beszédzavar, csendzavarás, földzárlat, rendzavarás, smaragdzöld, vadzab
ddz	0	0	0	0,00	0
dzs	100,00 32 20,92	0,00 0 0,00	32	20,13	[tehát *VADZSEB nincs!]
ddzs	0	0	0	0,00	0
$\Sigma/..$	153 96,23	6 3,77	159	100,00	

Ezek a példák elméletileg is, de talán kivált gyakorlatilag érdekesek. Hogyan is olvasunk, amikor magyarul olvasunk? Ha mi, jól olvasó magyar felnőttek, már el is felejtettük ezt (= az olvasás folyamata automatikussá vált bennünk) – ám hogyan olvasnak gyermekeink/unokáink, hogyan olvasnak magyar szöveget a magyarul tanuló külföldiek, hiszen ebben a két csoportban még többé-kevésbé nem automatizáltak ezek a folyamatok. Én itt természetesen nem kívánok erre a kérdésre válaszolni. Csak, előadásom címének megfelelően, nyersanyagot kívántam szolgáltatni efféle kísérletek elvégzéséhez, felhívni a figyelmet arra, még mennyi efféle nyersanyag áll a rendelkezésünkre géppel könnyen összegyűjthető módon mind a szótárakban, mind a szövegekben. – És még egy, részint ezzel összefüggő, gyakorlati kérdés – az elválasztásé.

A magyar szóelválasztás – egyébként kár büszkélkednünk: ilyesfajta még a német, a francia, az orosz stb. (csak nem az angol!) – olyan transzparens, hogy szinte magától kínálta magát az automatikus elvégzésre, az algoritmizálásra. Évtizedekkel ezelőtt éppen ezért ezen tanítottunk magyar szakos hallgatókat programozni (már elfelejtettem, de – horribile dictu – lehet, hogy gépi nyelven). Feltehetően ugyanezért nem sokkal utánunk a berlini akadémia nyelvtudományi intézetében ugyanezt a magyar nyelvi feladatot megoldották. Ez még hagyján – a magyar nyelv/helyesírás mondott sajátosságán kívül itt közrejátszhatott esetleg az is, hogy nekünk, magyaroknak már akkor is megalapozottan jó hírünk volt a táboron belül, a kétfajta németek között pláne... De azután alkalmasint ugyanezért az angolok csak úgy adtak el nekünk egy nagy nyomdai berendezést, ha vele együtt megvesszük a magyar elválasztás őáltaluk összeállított szabályait. És a furcsa eredmény – illetőleg "eredmény" –: egyéb külföldi partnereink (legutóbb mintha egy olasz cégről hallottam volna) csak így hajlandók szállítani "író" – nyomdai, feliratozó stb. – berendezéseiket. Ugyanakkor valahol elveszhetett útközben az alapötlet, a magyar szóelválasztás könnyű automatizálhatósága. Állítólag ennek a körülménynek (a magyar elválasztás szabályai nem-ismeretének) köszönhetjük a minden oldalról, így a magyar sajtóból is hirtelen ránkzúduló rettenetes elválasztások tömegét. (Ti. sajtónk is import nyomdagépeken dolgozik.) És mi lenne a helyzet, ha a kérdés m a g v á t évtizedekkel ezelőtt nem oldottuk volna már meg? (Éppen csak a magvát. Az összetételek helyét a memóriában csupán jeleztük, tehát maradt *kö-rút és *ren-dőr, mindig csak e-lég és fe-lül.)

1. sz. függelék: a *gyöngy-*, *pinty-*féle szavak (bekezdve a tőszavak):

(Az egyes oszlopok értelme: a) hány tő; b) három oszlop együtt: 001 ige, 002 fn, 003 mn, két szám – kettős szófajúság; c) képző: 0 nincs képző, 1 képző a szó végén, 2 képző a szó belsejében, 3 itt is, ott is, 9 különös képző, rag mint képző stb.)²

ngy (135 rekord, 15 tőszó)

angyal	10020	gyöngyhímzés	20021
angyalarcú	20031	gyöngyírás	20021
angyalbögyörő	20020	gyöngykagyló	20020
angyalbőr	20020	gyöngykór	20020
angyalcsináló	20321	gyöngyköles	20020
angyalfia	20029	gyöngyös	10321
angyalhaj	20020	gyöngyözik	10011
angyali	10031	gyöngysor	20020
angyalka	10021	gyöngyszem	20020
angyalos	10031	gyöngyszín	20230
arkangyal	20020	gyöngyszínű	20031
ány	10020	gyöngytyúk	20020
ányi	10021	gyöngyvászón	20020
árpagyöngy	20020	gyöngyvirág	20020
babakelengye	20020	hangya	10020
begöngyöl	60011	hangyaboly	20020
békeangyal	20020	hangyamászás	20021
bongyor	10030	hangyasav	20020
csecsemőkelengye	20020	hangyaszorgalom	20021
daragyöngy	20020	hangyatojás	20021
drágagyöngy	20020	hangyász	10021
elrongyol	60011	igazgyöngy	20020
elrongyolódik	60011	ingyen	16329
fagyöngy	20020	ingyenes	10329
felgöngyöl	60011	ingyenélő	20329
felgöngyölít	60011	ingyenhely	20029
félingyenes	20329	ingyenjegy	20029
géprongy	20020	ingyért	10069
göngyöl	10011	kalangya	10020
göngyöleg	10021	kalangyál	10011
göngyölít	10011	kapcarongy	20020
göngysúly	20020	kelengye	10020
göröngy	10020	kengyel	10020
göröngyös	10031	kengyelfutó	20021
gyöngy	10230	kengyelszár	20020
gyöngyárpa	20020	kengyelszj	20020
gyöngybagoly	20020	kengyelvas	20020
gyöngybetű	20020	kigöngyöl	60011
gyöngyélet	20021	kigöngyölít	60011
gyöngyfonál	20021	kigöngyölődik	60N11
gyöngyfüggöny	20021	langy	10030
gyöngyfüzér	20021	langymeleg	20030
gyöngyhalász	20021	langyos	10031
gyöngyharmat	20020	langyosít	10011
gyöngyház	20020	langyosodik	10M11

langyosság	10021	rongybaba	20020
lengyel	10230	rongyabda	20020
lengyeles	10031	rongyol	10011
lengyelke	10021	rongyolódik	10011
lengyolcs	20020	rongyos	10321
lerongyolódik	60011	rongyoskodik	10011
lerongyolt	60031	rongyosleves	20023
öngyilkos	20329	rongyosodik	10N11
öngyilkosjelölt	30329	rongypapír	20020
öngyilkosság	20029	rongszedő	20321
öngyulladás	20029	rongszőnyeg	20021
öngyújtó	20029	spongya	10020
összegöngyöl	60011	szappangyökér	20020
őrangyal	20020	toprongyos	10031
papírrongy	20020	törlőrongy	20022
pongyola	10230	üveggyöngy	20020
pongyolaság	10021	vagongyár	20020
porrongy	20020	varangy	10021
rézangyal	20020	vándorkelengye	20020
ringy-rongy	20230	védangyal	20020
ringyó	10020	védőangyal	20022
rongy	10230	vízgyöngy	20020
rongyász	10021		

nty (43 rekord, 8 tőszó)

aranysarkantyús	20031	kulipintyó	10020
békaporonty	20020	lappantý	10020
billentyű	10021	leffentyű	10021
billentyűs	10031	megkontyol	60011
billentyűzet	10021	megsarkantyúz	60011
csappantý	10021	pattantý	10021
díszpinty	20020	pattantýs	10231
elszontyolódik	60011	pinty	10020
felkontyol	60011	pintyőke	10021
felkontyoz	60011	ponty	10020
fogantý	10021	poronty	10020
herkentyű	10020	röppentyű	10021
hópinty	20020	sarkantý	10021
kallantý	10021	sarkantýka	10021
kallantýs	10031	sarkantýs	10031
karmantý	10021	sarkantývirág	20022
kattantý	10021	sarkantýz	10011
konty	10020	szellentyű	10021
kontyalávaló	30029	szívbillentyű	20021
kontyos	10321	tükörponty	20020
kontyvirág	20020	zongorabillentyű	20021
kösöntý	10020		

2. sz. függelék: Az ÉrtSz. 'aa'-t, 'ee'-t és 'ii'-t tartalmazó címszavainak teljes listája. (A két rövidebb csak a címszavakat tartalmazza.)

aa (47 rekord)

babaarc	munkaasztal
bábaasszony	odaad
bérmaanya	odaadás
bérmaapa	odaadó
faanyag	odaajándékoz
faarc	óraadás
hasaalja	óraadó
hazaad	pályaaavató
hálaadás	példaadás
hálaadó	propagandaanyag
hátraarc	puskaagy
katonaanyag	rakétaautó
katonaasszony	rostaalj
kártyaadósság	ruhaakasztó
kártyaasztal	ruhaanyag
kornaasszony	szobaasszony
konyhaajtó	tapétaajtó
koponyaalap	utcaablak
libaaprólék	utcaajtó
macskaasztal	visszaad
mostohaanya	visszaakaszt
mostohaapa	visszaalakul
munkaadó	vitaanyag
munkaalkalom	

ee (74 rekord)

beebédél	60011	bűnbeesés	20029
beecsetel	60011	csereeszköz	20020
beemel	60010	csirkeeszű	20031
beemléz	60011	egybeesik	60019
beenged	60011	félreesik	60019
beerdősít	60011	félreeső	60039
beereszt	60011	kétségbeejt	20019
beeresztő	60321	kétségbejtő	20039
beerősít	60011	kétségbeesett	20329
beesett	60031	kétségbeesés	20029
beesik	60010	kétségbeesik	20019
beesteledik	60M11	közbeesik	60019
beeszik	60010	közbeeső	60039
beevez	60010	leegyszerűsít	60011
beleegyezés	60029	leegyszerűsödik	60011
beleegyezik	60019	leejt	60010
beleelegyedik	60019	leellik	60010
beleelegyül	60019	leelőlegez	60019
beleereszt	60019	leemel	60010
belesik	60019	leendő	10031
beleeszik	60019	leendőbeli	10321

leenged	60011	összeerősít	60011
leereszkedés	60021	összeesés	60021
leereszkedik	60011	összeesik	60010
leereszkedő	60031	összeesket	60011
leereszt	60011	összeesküszik	60011
leerősít	60011	összeesküvés	60021
leesik	60010	összeesküvő	60321
leeszik	60010	összeeszik	60010
leevez	60010	összeeszkábál	60011
összeegyezik	60011	padlóbeeresztő	70323
összeegyeztet	60011	teendő	10021
összeegyeztethetetlen	60031	teleereszt	60011
összeelegyedek	60011	teleesz	60019
összeelegyít	60011	tőkeerős	20033
összeenyvez	60011	zeneegylet	20029
összeereszt	60011	zeneelmélet	20021

ii (35 rekord)

antiimperialista
 férfiing
 hadiipar
 határkiigazítás
 háziipar
 kéziipar
 kiigazít
 kiigazodik
 kiigényel
 kiiktat
 kiindít
 kiindul
 kiindulás
 kiindulási
 kiinduló
 kiirt
 kiismer
 kiismerhetetlen

kiismerszik
 kiiszik
 kiizzad
 kiizzaszt
 lelkiismeret
 lelkiismeretes
 lelkiismeretesség
 lelkiismeretfurdalás
 lelkiismereti
 lelkiismeretlen
 lelkiismeretlenség
 lelkiismeretvizsgálat
 nekiindul
 nekiiramodik
 színiiskola
 színiiskolás
 tudniillik

¹ A konferencián szóban rosszul, scserbai teljes stílusban ejtettem az efféle szavakat: itt éppen a gyors, beszélgetési stílus az érdekes, ahol is a két elem a kiejtésben összeolvad. Erre az – ejtésbeli? hangzó stílusbeli? – hibámra Siptár Péter volt szíves felhívni a figyelmemet. Köszönöm. – S ha már az ilyen kritikai megjegyzésekért mondott köszöneteknél tartok: Tarnóczy Tamás a fent ecsetelt szembeállításomat (társadalmi vs. fizikai) túl élesnek találta. "... a fizikai rögzítmények (látható akusztikai lenyomatok) sohasem azonosak a valósággal. A valóság az agyunkban van!" – írta ennek kapcsán magánlevélben. Neki meg külön is köszönöm a hangosan ki sem mondott észrevételt: igyekeztem ennek megfelelően (át)fogalmazni.

² Az olvasó így fogalmat kaphat a DT 2 egészének a felépítéséről. Abban még, az itt közölteken kívül, információk vannak az egyes címszók belső (főleg alaktani) viselkedésére vonatkozóan. A szófajban előforduló nagybetűk – hexadecimális kódok, szintén csak a DT 2-ben.

Irodalom

- Deme László: Hangzásforma és írásforma. Nagy L. János (szerk.): Implom József helyesírási verseny / Rendszer és stilsztika a helyesírásban. Gyula 1992, 25–32.
- Hell György: A magyar főszókincs fonémastatisztikája, szótagtípus statisztikája és mássalhangzókapcsolódásai. Budapest 1983. [Kézirat]
- Olaszy Gábor: Elektronikus beszédelőállítás. A magyar beszéd akusztikája és formánszintézise. Budapest 1989.