

Óvodások beszéde etimológiai megközelítésben¹

1. Bevezetés

A magyar nyelvvel kapcsolatban száz éve folynak etimológiai statisztikai kutatások. Hozzá kell azonban tenni, hogy kevesen és ritkán vállalkoztak effajta vizsgálódásra. Mivel a témakör kutatástörténete iránt érdeklődők kritikai értékelést is tartalmazó szemlét találhatnak első ilyen tárgyú írásomban (Horváth 2000), a részletes áttekintés helyett itt csupán két klasszikussá vált tanulmányt emelek ki: Tolnai Vilmosét (1924) és Benkő Lorándét (1962). Magam is több etimológiai statisztikai vizsgálatot végeztem már el: részint kortárs szövegek alapján, részint nyelvtörténeti anyagon (Horváth 2000, 2002, 2004, 2010, 2012, 2016, 2017).

Az etimológiai statisztikák egy része szótári alapú; más részük használati típusú, azaz szövegekben méri fel a szavak eredetkategóriáinak az arányait. A használati statisztikák a legutóbbi időkig csakis írott szövegek alapján készültek. Néhány évvel ezelőtt úgy döntöttem, hogy változtatok ezen a szokáson, és a beszélt nyelvet is bevonom a vizsgálataim körébe. Erre azon a konferencián kerítettem sort, amely a különféle generációk nyelvhasználatával foglalkozott az ELTE Mai Magyar Nyelvi Tanszékének szervezésében (Horváth 2016). Az akkori előadás tekinthető a mostani közvetlen előzményének. Ott az MTA Nyelvtudományi Intézetében épülő BEA adatbázis (l. pl. Gósy 2008) segítségével a felnőttek három nemzedékének spontán narratíváit hasonlítottam össze szóképzésük eredete szempontjából. Ez a munka kíváncsivá tett arra, milyen lehet az eredetkategóriák képviselője a gyermeknyelvben, például az óvodások beszédében. Ennek a kiderítéséhez alkalmas segédeszközként kínálkozott a GABI gyermeknyelvi adatbázis (vö. Bóna 2017; Vakula–Váradi 2017), és ösztönző hatású volt az anyanyelv-elsajátítási konferencia meghirdetése. Vonzott az út járatlansága is, hiszen tudomásom szerint a gyermeknyelvi és az etimológiai kutatás összekapcsolása mindkét területen újdonságot jelent.

Etimológusként természetesen arra irányul az érdeklődésem, hogy az egyes eredetkategóriák – és különösen a főkategóriák – mekkora képviselőhöz jutnak az óvodások nyelvében. Mennyiben és milyen mértékben más még ebből a szempontból a nyelvhasználat ebben az életkorban, mint a felnőtteké?

És mi érdekelheti a gyermeknyelv kutatóját etimológiai vonatkozásban? Személyes tapasztalatok hiányában erre nehezebb felelnem. Az biztos, hogy nem remélhet olyan közvetlen, gyakorlati értelemben vett hasznot, mint például a lo-

¹ Előadásként elhangzott az ELTE BTK-nak *Az anyanyelv-elsajátítás folyamata hároméves kor után* című konferenciáján 2018. május 24-én. Itt is megköszönöm Bóna Juditnak az engedélyezést a GABI adatbázis felhasználásához, valamint az adatközlők kiválasztásában nyújtott segítségét. Tanulmányom megírása kapcsolódik az NKFI *Új magyar etimológiai szótár. Második ütem* című, K 124127. számú projektumához; ennek a munkálatai Gerstner Károly vezetésével folynak a Nyelvtudományi Intézetben.

gopédiától. De úgy gondolom, hogy az oktatás számára sem felesleges – amolyan háttértudásként – annak a feltérképezése, hogy a gyerekek milyen származású szavak készletével élnek, mennyire hatolnak be nyelvhasználatukba az idegen eredetű elemek, milyen súllyal vannak meg benne a belső keletkezésű szavak különböző típusai: a szóképzés, az összetétel, a hangutánzás-hangfestés és egyebek.

Az óvodások szókészletének megvizsgálása előtt azt feltételeztem, hogy beszédükben nagyobb az alapnyelvi (uráli, finnugor, ugor kori) örökség képviselete, mint a felnőttekben, hiszen a gyerekeknek még sem idejük és alkalmuk, sem pedig szükségük nem volt arra, hogy a mindennapok alapszókincsén felül akkora fogalomszói készlettel dúsítsák fel a sajátjukat, amekkora vetekedhetne a felnőttekével. Mivel a bonyolultabb, elvontabb témák az óvodások narratíváiban aligha jelentkeznek, a hipotéziseim közé tartozott az is, hogy az ilyen tárgyokban nagy valószínűséggel felbukkanó idegen szavak és komplex morfémaszerkezetű elemek az óvodásokra nem jellemzők, tehát nem szoríthatják háttérbe az alapnyelvi örökség képviselőit. Arra szintén számítottam, hogy az óvodások narratíváiban jóval kevesebb a folyamatos, megszakítás nélküli szöveg, mint a felnőttekéiben, kényszerűen több viszont a felvételkészítő kérdéseire való reagálás, például tagadószókkal, névmásfélékkel: mindez feltehetőleg ugyancsak kedvez az alapnyelvi örökség reprezentánsainak.

A felnőttek beszédével való összevetésen kívül célul tűztem ki az óvodás lányok és fiúk szóhasználatának az összevetését, noha ebben a tekintetben jelentős különbségekre nem számítottam.

2. Anyag, adatközlők és módszerek

Mint említettem, az óvodások beszédének vizsgálatához a GABI gyermeknyelvi adatbázist használtam fel. Azon belül – igazodva a felnőttek nyelvhasználatára vonatkozó felmérésemhez (Horváth 2016) – a spontán narratívának nevezett egységekkel dolgoztam. A gyűjtéskor ebben a feladattípusban a felvétel vezetője arra kérte (és áthidaló kérdésekkel, megjegyzésekkel abban segítette) az óvodást, hogy beszéljen egy, az érdeklődési köréhez tartozó, személyes témáról: az óvodai vagy családi életről, játékokról, ünnepekről, a tévében látott műsorról stb.

Etimológiai statisztikai elemzésemhez a GABI adatközlői közül (Bóna Judit segítségével) öt éves óvodásokat választottam ki, 5-5 lányt és fiút. Az ő narratíváikat egészükben feldolgoztam, de csak a saját szövegüket vettem figyelembe a számításokhoz; a felvételkészítő kérdéseit, megjegyzéseit természetesen nem.

A használati statisztikák összeállítói hol a szövegben közvetlenül megjelenő szavakat minősítik eredetük szerint, hol visszavezetik őket a tövükre (vagy másféle etimológiai előzményükre), hol pedig mind a kétféle statisztikát elkészítik (részletesebben l. Horváth 2000). Az első magyar sakk-könyvet vizsgálva bevezettem egy harmadik statisztikafajtát is: ez a **lemmastatisztika**. A lemma a szónál elvontabb egység, lényegében megfelel a szótárírásban megszokott címszónak. Míg a szóstatisztikában minden képzett szó külön egységnek számít az alapszavához képest, a lemmastatisztikában az igenévképzőket, valamint a ható,

a műveltető és a szenvedő igék képzőit lemetsszük, és úgy adunk eredetminősítést a maradék szórésznek. Mindhárom statisztikafajtának megvan a maga haszna, de a konferencia-előadás keretében csak az egyiknek az elkészítésére vállalkozhatam. A felnőtt nemzedékek nyelvhasználatának a vizsgálatához (Horváth 2016) igazodva most is a lemmastatisztika mellett döntöttem.

A lemmák eredetminősítésében elsősorban a magyar nyelv legjobb teljes etimológiai szótára, az *Etymologisches Wörterbuch des Ungarischen* (= EWUng.), továbbá morfémaszerkezetük volt a segítségemre. Néhol felhasználtam azokat a minősítéseket is, amelyeket Gerstner Károly (2003) a *Magyar értelmező kéziszótár* megújított kiadásának (= ÉKsz.²) a szócikkeihez fűzött.

Mivel a tulajdonnevek származtatása más szavakétól eltérő, speciális feladat, az etimológiai statisztikai vizsgálatok többségében szokásos módon én is kizártam őket a felmérésből.

A korábbi elemzéseimhez hasonlóan az óvodások narratívái alapján is készítettem mind állományi, mind előfordulási statisztikát. Az **állományi statisztika** csak arra figyel, melyek azok a lemmák, amelyek az óvodások szövegeiben megjelennek, ismételt felbukkanásukra nincs tekintettel. Az **előfordulási statisztika** viszont minden egyes lemmaadatot figyelembe vesz, természetesen a szövegek más pontjain ismétlődőket is. A kétféle statisztika alapja tehát a másfajta nyelvészeti munkákban „type : token” vagy „típus : példány” elnevezésekkel alkalmazott megkülönböztetés.

A lemmaszámlálással kapcsolatban ki kell térnem arra, hogyan kezeltem a szövegek beszélt nyelvi voltával összefüggő jelenségeket. – A lejegyzésben „Ö”-vel jelölt „nyökögő hezitálást” lemmaadatként nem vettem figyelembe. – Nem vehettem számításba az ilyenféle elharapott szótöredékeket sem: *f, kö-*. Ha viszont csak kissé gondatlan ejtés miatt maradt le a szóalak végéről egy-két hang, de a lemmát kétségtelenül azonosítani lehetett (*há* ’hát’ típus), akkor beszámítottam a lemmaadatot. – A kapkodásból vagy bizonytalankodásból adódó *amed- ameddig* típusú ismétléseket a lemma egy adatának tekintettem, a *hogy hogy ...* típusúakat viszont annyinak, ahányszor az egészében (azaz nem töredékében) kimondott alak elhangzott; például ebben a részletben kettőnek: „*azt hogy hogy futkájoztam*”. – A *jáco tásasoztun* [’társasoztunk’] típusú önjavítások esetében a második elemet fogadtam el érvényesnek, az vált a statisztika részévé.

Akárcsak a korábbi elemzéseim többségében, az etimológiai statisztikát ezúttal is két változatban készítettem el. Ezek a bizonytalan és a vitatott eredeztetések kezelésében térnek el egymástól. Az elsőben mind a bizonytalan, mind a vitatott eredetű elemek önálló kategóriaként szerepelnek, hasonlóan például az összetett lemmákhoz vagy az olasz jövevényekhez. A másik változathoz viszont felszámoltam a bizonytalan és a vitatott eredetűek kategóriáját: képviselőiket a számításba vehető, illetőleg a valószínűbb magyarázatuknak a kategóriájához soroltam át. Így került például az első statisztikában a bizonytalan eredetűekhez tartozó *s* a másodikban a finnugor örökség reprezentánsaihoz, míg az elsőben a vitatott eredetűek között szereplő *zöld* a másodikban az iráni jövevényekhez. (Az átsorolás elveiről és módszeréről bővebben l. főleg Horváth 2000: 180, 321–2.) – Felszámoltam a főkategórián belüli tisztázatlanság kategóriáit is, vagyis az alapnyelvi örökségen

belül a tisztázatlan rétegükét, a belső keletkezésűeken belül pedig a tisztázatlan kialakulásmódúakét. (Ennek egyes részleteihez l. a 3. pontban a kategóriák szemléjét.) – A kétféle statisztikát a 3. pontban található táblázatok jelenítik meg.

Ami az etimológiai kategóriák önálló vagy összevont kezelését illeti, szintén alkalmazkodtam az előző elemzéseimben követett elvekhez. A valószínű magyarázatokat egyenértékűeknek tekintetem a biztosakkal. Az alapnyelvi rétegeket megkülönböztettem egymástól. A konkrét szóátadó nyelvek szerint nem differenciáltam az iráni, a török és a szláv jövevényeket, az olaszt mint leánnyelvet viszont elkülönítettem a latintól. Az idegen mintára, de magyar elemekből alkotott tükörosszetéseket a belső keletkezésűek főkategóriájában helyeztem el, de megkülönböztetve a többi összetett lemmától (vö. Horváth 2016: 311–4). – Egy-két további kategóriához a 3. pontban, a táblázatok bemutatásakor fűzök megjegyzést.

A statisztikai táblázatokban a 10 óvodás adatközlő használta lemmakészlet – akárcsak annak idején a felnőtt nemzedékeké (Horváth 2016: 311–5) – összeítve szerepel. Az egyéni eredmények bemutatása nehezen követhetővé tenné az elemzést, és terjedelmi okból sem lehet megvalósítani.

3. Eredmények

A 10 éves óvodás narratíváiban összesen **655 lemma** jelenik meg, **2855 előfordulással**. Egy lemmára tehát átlagosan 4,36 adat jut.

Az eredetkategóriákat képviselő lemmák állományának és előfordulásainak számadatait, valamint a százalékarányokat az 1. és a 2. táblázat mutatja be. A felnőtt nemzedékek körében mért eredményekkel való összehasonlításra (vö. Horváth 2016: 311–5) csak az 1. táblázat alkalmas, mivel annak idején a felnőttek szövegeiből csak 1. típusú (azaz a bizonytalan és a vitatott eredet kategóriáit megtartó) statisztikát állítottam össze.

Az 1. táblázatból kitűnik, hogy az óvodások narratíváinak **lemmaállományában** az etimológiai főkategóriák közül a belső keletkezésűeké a vezető szerep: 70% feletti a részesedésük. A jövevénylemmáké és az alapnyelvi örökség reprezentánsaié – ettől sokkal elmaradva – valamivel 10% feletti, a tisztázatlan eredetűeké pedig a készlet 6%-a.

A belső keletkezésűek a **lemma-előfordulások** körében is messze megelőzik a többi főkategóriát, de egy-egy képviselőjükre a korpuszátlagnál kisebb adatszám jut (4,16), így az előfordulásokat tekintve a részesedésük csak megközelíti, de nem éri el a 70%-ot. A jövevények ismétlődési hányadosa ennél sokkal kisebb (1,66), ezért ez a főkategória az előfordulásokat tekintve az utolsó helyre csúszik vissza, 5%-nál is kisebb képviseléssel. Sokkal jellemzőbb viszont a korpusz átlagánál a szövegbeli ismétlődés az alapnyelvi örökséghez tartozó lemmákra (7,54), így az előfordulásokra nézve a részesedésük megközelíti a 20%-ot. Jóval az átlag feletti a tisztázatlan eredetűek ismétlődési hányadosa is (6,20), tehát az előfordulások között nagyobb a képviselőségük az állománybelinél: 9%-nyi.

1. lemmaeredet-statisztikai táblázat

Eredet	Lemma	%	Előfordulás	%
uráli	14	2,14	118	4,13
finnugor	27	4,12	299	10,47
ugor	5	0,76	29	1,02
tisztázatlan rétegből	3	0,46	12	0,42
kétforrású örökség	1	0,15	3	0,11
örökség magyar képzővel	20	3,05	67	2,35
<i>Örökség összesen</i>	<i>70</i>	<i>10,69</i>	<i>528</i>	<i>18,49</i>
származék	119	18,17	235	8,23
fiktív többlől	20	3,05	37	1,30
elvonás	2	0,31	3	0,11
szóróvidülés	2	0,31	4	0,14
szóösszehúzás	2	0,31	2	0,07
jel- vagy ragszilárdulás	69	10,53	245	8,58
önállósulás	3	0,46	3	0,11
szófajváltás	18	2,75	601	21,05
szóhasadás	7	1,07	75	2,63
szóvegyülés	1	0,15	3	0,11
összetétel	197	30,08	583	20,42
szóösszevonás	2	0,31	6	0,21
tükörösszetétel	3	0,46	4	0,14
onomatopoetikus	13	1,98	55	1,93
belső, de tisztázatlan	9	1,37	89	3,12
<i>Belső keletkezésű összesen</i>	<i>467</i>	<i>71,30</i>	<i>1945</i>	<i>68,13</i>
iráni	3	0,46	3	0,11
török	9	1,37	12	0,42
német	8	1,22	12	0,42
angol	2	0,31	2	0,07
latin	8	1,22	11	0,39
olasz	3	0,46	5	0,18
szláv	27	4,12	56	1,96
több nyelvből	1	0,15	1	0,04
közvetített	2	0,31	2	0,07
vándor	7	1,07	13	0,46
nemzetközi	7	1,07	11	0,39
<i>Jövevény összesen</i>	<i>77</i>	<i>11,76</i>	<i>128</i>	<i>4,48</i>
bizonytalan	19	2,90	65	2,28
vitatott	9	1,37	104	3,64
ismeretlen	13	1,98	85	2,98
<i>Tisztázatlan összesen</i>	<i>41</i>	<i>6,26</i>	<i>254</i>	<i>8,90</i>
<i>Összesen</i>	<i>655</i>		<i>2855</i>	

A főkategóriáknak az 1. táblázatból látható képviselőjét összehasonlítottam a felfelé narratíváiban talált képpel (vö. Horváth 2016: 314–5). Ehhez χ^2 próbát

használtam. Az eltéréseket $p < 0,05$ esetén, azaz a véletlennek 5%-nál kevesebb esélyt hagyva tekintettem szignifikánsnak.

Az alapnyelvi örökségből való lemmák – mind az állományt, mind az előfordulásokat tekintve – sokkal nagyobb arányban vannak jelen az óvodások narratíváiban, mint a felnőttekéiben ($p < 0,001$). A felnőttek beszédében az állományi részesedésük még a 6%-ot sem éri el, és az előfordulási is csupán 13%-nyi.

A belső keletkezésűekkel éppen fordítva áll a dolog: a felnőttek javára mutatkozik igen jelentős különbség ($p < 0,001$). Náluk a lemmaállományban 81%-os, míg az előfordulások között 75%-os ennek a főkategóriának a képviselője.

A jövevénylemmák reprezentáltságában nem mérhető szignifikáns eltérés az óvodás és a felnőtt adatközlők beszéde között. Ezeknek az állományi részesedése a felnőttek narratíváiban 10%-os, míg az előfordulási arányuk 5,5%.

A tisztázatlan eredetű lemmák az óvodások beszédére sokkal jellemzőbbek (az állományban $0,01 > p > 0,001$; az előfordulásokban $p < 0,001$). A felnőtteknél az ilyen lemmák az állományból 3,5%-kal, míg az előfordulásokból 6,7%-kal részesednek.

A vizsgálat tehát igazolta azt a hipotézist, hogy az óvodások beszédében nagyobb a képviselője az alapnyelvi örökség elemeinek, mint a felnőttekéiben. Helyesnek bizonyult az a feltételezés is, hogy a belső keletkezésű lemmák aránya még nem éri el a felnőttkorit, jóllehet már az óvodások narratíváiban is fölényesen vezetik a főkategóriák listáját. Nem igazolódott viszont az a sejtés, hogy az idegen eredetű lemmák képviselője elmaradhat a felnőttek körében megfigyelttől; minden bizonnyal azért nem, mert az anyagban jórészt igazán meghonosodott jövevénytiszavakkal van dolgunk, érezhetően idegenekkel kevésbé.

Etimológiai szempontból az óvodás fiúk és lányok beszéde között jelentős különbségekre előzetesen nemigen számítottam. Sok eltérést nem is találtam, de feltűnő, hogy a jövevénylemmák képviselője – mind az állományra, mind az előfordulásokra nézve – szignifikánsan nagyobb a lányok beszédében, mint a fiúkban ($0,05 > p > 0,02$): az állománybeli részesedés a fiúknál 8,42%, a lányoknál 12,81%; az előfordulási a fiúknál 3,66%, míg a lányoknál 5,33%. Még egy szignifikáns eltérésre bukkantam, de erre csak az előfordulások szempontjából: a tisztázatlan eredetű lemmák képviselője a fiúk szövegeiben nagyobb (náluk 10,01%, a lányoknál 7,74%; $0,05 > p > 0,02$). A nemek nyelvhasználatában megmutatkozó különbségekre egyelőre nem tudok magyarázatot adni; lehetséges, hogy pusztán az adatközlők csekély számával függenek össze.

Az 1. táblázat összképét, valamint a nemzedékek és a nemek beszédének különbségeit az eredetbeli főkategóriák arányaival jellemeztem. Szeretném még bemutatni a körükön belül megtalálható etimológiai kategóriák képviselőjét az óvodás adatközlők beszédében, példákkal is szemléltetve. A lemmák mellett előfordulási számukat akkor adom meg, ha nagyobb egynél.

Az óvodások narratíváiban az **alapnyelvi** rétegek közül – a korábbi elemzéseim tapasztalata szerint egyáltalán nem váratlanul – a finnugornak a képviselője legnagyobb. A leggyakoribb reprezentánsai ezek: *van* (129 adattal), *nem* tagadószó (76), *ő* (27), *kell* (18). – Az uráli rétegből származó lemmákból feleannyi van, mint a finnugor koriakból. Közülük a *tud* fordul elő legtöbbször (54),

ezt a *megy* (19), a *mi* kérdő névmás (14), a *mi* személyes névmás (6) és a *név* (5) követi. – Az ugor rétegből pedig ezek a leggyakoribbak: *jó* (19), *domb* (4), *fog* (segéd)ige (4).

Vannak olyan esetek is, amelyekben nem állapítható meg, hogy a lemma melyik rétegből öröklődött. Mivel ez egyfajta (igaz, főkategórián belüli) vitatottság, a 2. statisztikai táblázat összeállításakor felszámoltam ezt a kategóriát, és a lemmáit a leginkább számításba vehető réteg képviselőiként vettem figyelembe. – Összesen 3 lemma esetében tisztázatlan az alapnyelvi réteg. Mellettük most zárójelben azt a réteget tüntetem fel, amelynek a képviselőit a 2. táblázatban gyarapították: *jön* (ugor) (6), *két* (finnugor) (3), *kettő* (finnugor) (3).

A „kétforrású örökség” kategóriára egyetlen lemma, a *hall* ige (3 adat) miatt volt szükségem: ez két alapnyelvi szó összetartó fejlődésével keletkezett (a részleteket l. az EWUng.-ban).

Az olyan lemmáknak, amelyek „alapnyelvi fiktív tő + magyar képző” felépítésűek (a *fiktív tő* szak kifejezéshez vö. főleg Benkő 1984), de magának a morfémaegyüttesnek nincsenek rokon nyelvi megfelelői, az „örökség magyar képzővel” minősítést adtam. A kategória leggyakoribb képviselői: *lesz* (11), *mond* (9), *mese* (7).

Ami a **belső keletkezésű** lemmákat illeti, a felnőttek szövegeiben (vö. Horváth 2016: 314) a származékok (vagyis a képzés produktumai) és az összetettek szinte „kéz a kézben” vezették az állományi listát. Az óvodások beszédében ellenben szembeűnő a különbség: az összetételek részesedése a teljes lemmaállományból 30%, míg a származékoké csak 18%. Az viszont itt is igaz, hogy a lemmaállományt tekintve a kategóriák listájának – beleértve a nem belső keletkezésűeket is – ez az első két helyezette. Ugyanakkor az ismétlődés sem az összetett, sem a képzett lemmákra nem jellemző, így az előfordulásokból az összetettek csak 20%-kal, a származékok pedig mindössze 8%-kal részesednek.

Az összetétellel létrejött lemmák között mindamellettt vannak kiemelkedő gyakoriságúak: *és* (157), *is* (45), bár ezeknek az összetett volta elhomályosult. A két listavezető után szintén olyanok következnek, amelyek nem fogalomszói jellegűek: *ami* (21), *én* (19), *amikor* (11). A leggyakoribb fogalomszói összetétel a *szülinap* (10); aligha véletlenül, hiszen az óvodások számára nagyon fontos fogalmat jelöl. – Az összetett lemmák közé tartoznak az igekötős igék is. Közülük az *átmegy* bukkan fel a legtöbbször (5). – Bőven vannak olyan összetett szavak, amelyek csekély adatszámuk miatt önmagukban nincsenek nagy hatással a statisztika képére, érdekességképpen, az óvodások világát felidéző voltokra tekintettel mégis érdemesnek tartom néhányukat a felsorolásra: *kistesó* (4), *hercegnő* (2), továbbá egy-egy előfordulással: *babapiskóta*, *boszorkányruha*, *dínócsontváz*, *gumilövedék*, *hógolyógyűjtemény*, *jégszörny*, *kábojjáték*, *lufimeccs*, *pillangókورونا*, *pókemberjelmez*, *robotkutya*.

A származékokra is igaz, hogy nem fogalomszói elemek állnak a listájuk élén: *ilyen* (34), *olyan* (11). A *kutya* (10) és a *lehet* (9) csak ezek után következnek. – Itt is felsorolok jellegzetes „óvodásszavakat”: *bobozik* (2), *hercegnős* (2), valamint egy-egy adattal: *dínófigurákos*, *kézműveskedik*, *kisvakondos*, *nőrfőzik*, *okés*, *robotfejecske*, *robotka*, *transzformer*.

A lemmaállományi listán az összes kategóriát tekintve a jel- és ragszilárdulásé áll a harmadik helyen, 10%-ot meghaladó eredménnyel. Az előfordulásokra nézve kissé meg is előzi a származékokat. Képviselői közül a diskurzusjelölő *hát* (28) jelenik meg a legtöbbször; utána az *itt* (18) és az *akkor* (13) következik. Az említett lemmák ragszilárdulás eredményei; a jelszilárdulásai közül a *neki* (10) a leggyakoribb.

A korábbi etimológiai statisztikáimhoz hasonlóan itt is sajátos kétarcúság jellemzi a szófajváltás kategóriáját: kevés lemma alkotja, de sok adattal; rendkívüli az ismétlődési hányados: 33,39. Az állományi részesedés 3% alatti, míg az előfordulási 21%. Az utóbbi eredmény még a megszokottnál is előkelőbb helyre juttatja a kategóriát a gyakorisági listán: egyenesen az elsőre. Természetesen számos kiemelkedően gyakori lemma tartozik ide: az *a* névelő (219), a *hogy* kötőszó (87), az *igen* mondatszó (76), a *meg* kötőszó (67), az *egy* névelő (58), a *de* kötőszó (37) és az *az* névelő (36). Nagy különbséggel követi őket a *vagy* kötőszó (8). Fogalomszói elemként említést érdemel a *farkas* (2), valamint az igenév szófajváltásával létrejött, egyszerű felbukkanó *eső*.

A fiktív tövek származékaiból két olyan lemmának van a legtöbb adata (7-7), amelyek jellemzőek az óvodáséletre: *csoport* és *ovi*. Meglepő volt számomra az *igi* 'igen' felbukkanása (2), mivel máshol még nem talákoztam vele. Az óvodásokra jellemzőnek érezhetjük a *boszi* és a *cuki* lemmákat, noha csupán egyszer jelennek meg.

Elvonással jött létre a *szörny* (2) és a *szundi*, rövidüléssel a *dínó* (3) és a *kicsi*, szóösszehúzással (azaz a szó közepének elmaradásával) a *cipő* és a *tán*, szóösszevonással (azaz csonkulásos összetétellel) az *asziszem* (5) és az *aszondom*. Önállósulás eredménye a *kalap*, a *radír* és a *testvér*; szóvegyülése a *kufli* [*kukac* × *kifli*] (3). A szóhasadással létrejött lemmák között vannak ezeknél jóval gyakoribbak is: *még* (37), *mert* (19), *pedig* (14). A beszédkorpusz mindhárom tükörösszetétele német mintára keletkezett: *feladat* (2), *járőr*, *műanyag*. Az onomatopoeitikus elemekből a leggyakoribb a *kap* ige (15). Ezt mondatszók követik: *aha* (9), *kár* <madárhang> (8), *ühüm* (8).

Akadnak olyan lemmák is, amelyeket ugyan biztosan belső keletkezésűeknek tarthatunk, de ezen belül a kialakulásuk módja tisztázatlan. Közülük nem is egy elég gyakori, így az előfordulások között magának a kategóriának sem jelentéktelen, 3% feletti a képviselete. A példák mellett zárójelben azt a kategóriát is feltüntettem, amelybe a 2. táblázat megszerkesztésekor az adott lemmát átsoroltam: *ügy* (ragszilárdulás) (26), *így* (ragszilárdulás) (23), *már* (származék) (13), *egy* számnév (ragszilárdulás) (11).

A **jövevénylemmák** közül – más etimológiai statisztikáimhoz hasonlóan – a szláv eredetűeké a vezető szerep. Az állománybeli részesedésük 4%-os. Mivel azonban a sűrű ismétlődés nem jellemző rájuk (ahogy a többi jövevényre sem), az előfordulásokat tekintve csak 2%-nyi a képviseletük. A leggyakoribb reprezentánsuk a gyerekek világához jól illően a *robot* (9), ezt a *barát* (6) és a *csinál* (6) követi.

A szlávokon kívül a török, a német és a latin jövevényeknek is 1%-nál nagyobb az állománybeli arányuk. A törökök képviselői közül a *kis* (3) és a *kor* (2) fordul elő

egynél többször. A németek közül a *lila* (3), a *ja* (2) és a *pont* (2). Jövevényigeként említést érdemel a *rajzol*, az óvodás szóhasználat megidézőjeként pedig a *tuti*. A latinokból ezeknek van egynél több adatuk: *iskola* (2), *korona* (2), *múmia* (2).

A korpuszbeli angol eredetű jövevények (a vártnál talán kevesebb): *helovin* 'halloween', *oké*. Akadnak még olasz származásúak is: *pálya* (2), *torta* (2), *katonna*, továbbá irániak: *kard*, *kincs*, *vár* főnév. Több nyelvből (a franciából és a németből) került át a magyarba a *frufru*.

Vannak olyan jövevények is, amelyek esetében fontosabb átadójuk mellett egy másik nyelv is közvetítő szerepet játsz(hat)ott. Ilyen a *január* (a latinból, később a németből is), valamint a *szappan* (az olaszból, esetleg szerbhorvát közvetítéssel is).

Tágabb értelemben a jövevénytiszavak közé tartoznak a vándorszók és a nemzetközi szavak is. A vándorok közül a korpuszban egynél többször megjelenő lemmák: *papagáj* (4), *kártya* (3), *könyv* (2). A nemzetközieseknek mindegyik képviselőjét felsorolom annak a szemléltetésére, miféle elemek jutottak be közülük az óvodások szókincsébe: *autó* (3), *helikopter* (2), *legó* (2), *balerina*, *logopédia*, *szuper*, *telefon*.

A **tisztázatlan eredetűek** főkategóriájához a bizonytalan, a vitatott és az ismeretlen származású lemmák tartoznak. A bizonytalanok és a vitatottak példáinak felsorolásakor zárójelben azt a minősítésüket is megadom, amellyel a 2. táblázatban szereplő statisztikához hozzájárultak.

A bizonytalan eredetűek ismétlődési hányadosa kisebb az átlagosnál (3,42). Közülük csak a *játszik* (örökség magyar képzővel) igazán gyakori (26 adat). Ezt a *játék* (szintén örökség magyar képzővel) (6) és az *s* (finnugor) (5) követi.

A vitatott eredetűek kategóriája annyiban a szófajváltásra emlékeztet, hogy kevés képviselője van, de nagyon sok előfordulással. Az ismétlődési hányados kiugróan nagy (11,56), ennél csupán a szófajváltásnál mért érték magasabb (33,39). A vitatott eredetűek leggyakoribb reprezentánsa az *az* névmás (a 2. táblázathoz finnugor) (75 adat); a listán palatális párja, az *ez* (az áttrendezéssel szintén finnugor) követi (19). Érdekességképpen, a gyermeki spontaneitás jellemzőjeként említem meg az egyszer felbukkanó *áu* mondatszót (a 2. statisztikában onomatopoeitikus).

2. lemmaeredet-statisztikai táblázat

Eredet	Lemma	%	Előfordulás	%
uráli	15	2,29	122	4,27
finnugor	36	5,50	410	14,36
ugor	7	1,07	36	1,26
kétforrású örökség	1	0,15	3	0,11
örökség magyar képzővel	27	4,12	106	3,71
<i>Örökség összesen</i>	<i>86</i>	<i>13,13</i>	<i>677</i>	<i>23,71</i>
származék	122	18,63	253	8,86
fiktív töből	21	3,21	41	1,44
elvonás	2	0,31	3	0,11
szőrövidülés	2	0,31	4	0,14
szőösszehúzás	2	0,31	2	0,07

Eredet	Lemma	%	Előfordulás	%
jel- vagy ragszilárdulás	75	11,45	315	11,03
önállósulás	3	0,46	3	0,11
szófajváltás	19	2,90	604	21,16
szóhasadás	7	1,07	75	2,63
szóvegyülés	1	0,15	3	0,11
összetétel	198	30,23	585	20,49
szóösszevonás	2	0,31	6	0,21
tükörösszetétel	3	0,46	4	0,14
onomatopoetikus	15	2,29	57	2,00
<i>Belső keletkezésű összesen</i>	<i>472</i>	<i>72,06</i>	<i>1955</i>	<i>68,48</i>
iráni	5	0,76	7	0,25
török	12	1,83	16	0,56
német	8	1,22	12	0,42
angol	3	0,46	3	0,11
latin	8	1,22	11	0,39
olasz	4	0,61	6	0,21
szláv	27	4,12	56	1,96
több nyelvből	1	0,15	1	0,04
közvetített	1	0,15	1	0,04
vándor	7	1,07	13	0,46
nemzetközi	8	1,22	12	0,42
<i>Jövevény összesen</i>	<i>84</i>	<i>12,82</i>	<i>138</i>	<i>4,83</i>
<i>Ismeretlen</i>	<i>13</i>	<i>1,98</i>	<i>85</i>	<i>2,98</i>
<i>Összesen</i>	<i>655</i>		<i>2855</i>	

Az ismeretlen eredetűek ismétlődési hányadosa (6,54) az átlagnál nagyobb, de nem kiemelkedő. A leggyakoribb képviselőik: *szokik* (37), *csak* (33).

A kategóriák szemlét a **lemmák összesített gyakorisági listájának** az élményével zárom, feltüntetve az eredetminősítéseket is: 1. *a* névelő (szófajváltás) 219 | 2. *és* (összetett) 157 | 3. *van* (finnugor) 129 | 4. *hogy* kötőszó (szófajváltás) 87 | 5–6. *igen* mondatszó (szófajváltás) és *nem* tagadószó (finnugor) 76–76 | 7. *az* névmás (vitattott, a 2. táblázathoz finnugor) 75 | 8. *meg* kötőszó (szófajváltás) 67 | 9. *egy* névelő (szófajváltás) 58 | 10. *tud* (uráli) 54. – Összehasonlításként a felnőttek lemmáinak sorrendjét is megadom (vö. Horváth 2016: 315): 1. *a* névelő, 2. *és*, 3. *van*, 4. *hát* diskurzuszjelölő, 5. *az* névelő, 6. *hogy* kötőszó, 7–8. *ez* névmás és *nem* tagadószó, 9. *az* névmás, 10. *is*, 11. *egy* névelő, 12. *mert*.

A 2. lemmastatisztikai táblázatot – ahogy erről a 2. pontban már szóltam – a bizonytalanság és a vitatottság kategóriáinak a felszámolásával hoztam létre az 1.-ből. Az átrendezés nyertese az alapnyelvi örökség főkategóriája: az állományban 13%-ra, az előfordulások között pedig 24%-ra nő a részesedése. A belső keletkezésűek és a jövevények nyeresége jóval kisebb. Az ismeretlen eredetűek pedig főkategóriává lépnek elő, változatlan képviseléssel.

Mindebből látható, hogy az óvodások narratíváinak a bizonytalan, illetőleg vitatott eredetűek minősített lemmáival kapcsolatban leginkább az alapnyelvi elemként

való származtatás jöhet szóba. Közvetve ezzel függhet össze az is, hogy a tisztázatlan eredetűek képviselete az óvodások narratíváiban szignifikánsan nagyobb, mint a felnőttekéiben, hiszen ez az alapnyelvi örökség vonatkozásában is így van.

4. Összegzés és kitekintés

Vizsgálatom, amelynek újdonsága az etimológiai és a gyermeknyelvi kutatás összekapcsolása, igazolta azt a feltételezést, hogy az óvodások beszédében még jóval nagyobb az alapnyelvből örökölt lemmák képviselete, mint később a felnőttekéiben. Ez minden bizonnyal abból adódik, hogy a lemmák alapnyelvi eredete és az alapszókincshez tartozásuk összefügg egymással; a gyerekeknek pedig még nem volt sem módjuk, sem szükségük arra, hogy az alapszókincs mellé akkora kiegészítő szókészlet birtokába jussanak, mint a felnőttek. Ugyanakkor a statisztika azt is megmutatta, hogy az etimológiai főkategóriák közül már az ötéves óvodások beszédében is a belső keletkezésű lemmáké a fölényes vezető szerep.

Az egyes kategóriák áttekintésekor az óvodások beszédében a felnőtt nemzedékekéhez képest (és más statisztikáimhoz viszonyítva is) feltűnő volt a származékok csekélyebb jelenléte. Az eredmény arra mutat, hogy a szóképzés szerepe később, az óvodáskor után nő meg.

Ugyancsak szembevetendő a szófajváltás képviselőinek a szokásosnál is nagyobb aránya az előfordulási adatok között. A háttérben az állhat, hogy az óvodások hajlamosak a névelők, kötőszók ismételtetésére, egyfajta hezitálásként is.

A felmérés szerint az óvodás fiúk és lányok narratívái között etimológiai szempontból nem sok különbség akad. Úgy tűnik, a lányokra jellemzőbb a jövevényszavak használata, mint a fiúkra, de ez az eredmény ellenőrzést kívánna az adatközlők számának növelésével. Mivel most csak lemmaeredet-statisztikát készítettem, érdemes lenne összeállítani a vizsgált beszédanyag szavainak, különösen pedig a szavak töveinek etimológiai statisztikáját.

Azokból a felnőtt nemzedékekből, amelyeknek a nyelvhasználatát a BEA adatbázis segítségével megvizsgáltam (Horváth 2016), a 20 éveseké volt a legfiatalabb, ezúttal pedig a GABI gyermeknyelvi beszédatadabázis felhasználásával ötéves óvodások szövegeit dolgoztam fel etimológiai szempontból. Mivel az elemzés érdekes és jelentős különbségeket tárt fel az óvodások és a felnőttek beszéde között, érdemes lenne felmérni egy-két átmenetet alkotó korosztálynak, például a 10 és 15 éveseknek a nyelvhasználatát is. Ehhez a GABI adatbázis alkalmas segítségként kínálkozik.

SZAKIRODALOM

- Benkő Loránd 1962. Adatok a magyar szókincs szerkezetének változásához. *Nyelvtudományi Közlemények* 64: 116–36.
- Benkő Loránd 1984. *A magyar fiktív (passzív) tövű igék*. Akadémiai Kiadó, Budapest.
- Bóna Judit 2017. GABI – Gyermeknyelvi beszédatadabázis a kutatásban. In: Bóna Judit (szerk.): *Új utak a gyermeknyelvi kutatásokban*. ELTE Eötvös Kiadó, Budapest, 35–50.

- ÉKsz.² = Pusztai Ferenc (főszerk.) 2003. *Magyar értelmező kéziszótár. Második, átdolgozott kiadás.* Akadémiai Kiadó, [Budapest].
- EWUng. = Benkő, Loránd (Hrsg.) 1993–1997. *Etymologisches Wörterbuch des Ungarischen I–III.* Akadémiai Kiadó, Budapest.
- Gerstner Károly 2003. [Eredetminősítések az ÉKsz.² szócikkeihez.]
- Gósy Mária 2008. Magyar spontánbeszéd-adatbázis – BEA. In: Gósy Mária (szerk.): *Beszédkutatás 2008. Sokszínű beszédtudomány.* MTA Nyelvtudományi Intézet Kempelen Farkas Beszédkutató Laboratórium, Budapest, 194–207.
- Horváth László 2000. Etimológiai kategóriák arányai mai elbeszélésekben. *Magyar Nyelv* 96: 170–81, 316–32.
- Horváth László 2002. Az Ómagyar Mária-siralom etimológiai statisztikája. *Magyar Nyelv* 98: 265–82.
- Horváth László 2004. *Két Halotti beszéd az etimológia tükrében.* Előadás. MTA Nyelvtudományi Intézet, 2004. február 26.
- Horváth László 2010. Az első magyar sakk-könyv etimológiai statisztikája. *Magyar Nyelvőr* 134: 421–36.
- Horváth László 2012. *Három Halotti beszéd az etimológiai statisztika tükrében.* Előadás. MTA Nyelvtudományi Intézet, 2012. november 20.
- Horváth László 2016. Nemzedékek spontán beszéde etimológiai megközelítésben. In: Balázs Géza – Veszelszki Ágnes (szerk.): *Generációk nyelve.* ELTE BTK Mai Magyar Nyelvi Tanszék – Inter Nonprofit Kft. – Magyar Szemiotikai Társaság, Budapest, 309–17.
- Horváth László 2017. A Toldi szókészletének eredetéről. *Magyar Nyelv* 113: 450–66.
- Tolnai Vilmos 1924. Halhatatlan magyar nyelv. *Magyar Nyelv* 20: 50–9.
- Vakula Tímea – Váradi Viola 2017. Gyermeeknyelvi hangfelvételek rögzítésének és lejegyzésének tapasztalatai. In: Bóna Judit (szerk.): *Új utak a gyermeknyelvi kutatásokban.* ELTE Eötvös Kiadó, Budapest, 51–64.

Horváth László
tudományos munkatárs
Nyelvtudományi Intézet

SUMMARY

Horváth, László

Nursery-school children's speech in an etymological perspective

Etymological statistical research on Hungarian has been going on for a century now. Some of that literature is dictionary-based; other research projects are usage-based, that is, they explore proportions of etymological categories of words occurring in various texts. Most usage-based statistical surveys so far have been conducted on written texts. It was only recently that the examination of spoken texts in this respect was started, by the present author. A previous survey of his contrasted three generations of adult speakers with respect to the origins of the word stock they used. The present talk studies the shares of etymological categories in the spontaneous speech of five-year-old nursery-school children with the help of the GABI database. The combination of child language research with etymological research represents a novelty in both areas.

Keywords: speech database, child language, etymological statistics, lemma statistics