

A Toldi tőkészletének eredetéről*

1. A vizsgálat előzménye. Használati típusú, azaz szövegre vonatkozó (nem pedig szótári alapú) etimológiai statisztika készítésekor a vizsgálat irányulhat mind a szövegben közvetlenül megjelenő szavak eredetkategoróriák szerinti arányaira, mind pedig töveiknek a származás szerinti képviselőjére. Már első etimológiai statisztikám összeállításakor (HORVÁTH 2000) úgy gondoltam, hogy mindkétfajta felmérés fontos. Az akkor és később szerzett elemzési tapasztalataim arra készítettek, hogy bevezessek egy harmadik statisztikafajtát is: a lemmák eredetstatisztikáját. A lemma a szónál elvontabb egység, lényegében megfelel a lexikográfiából ismert címszónak. A lemmák eredetstatisztikájának előkészítésekor a szavakról lemeztünk bizonyos képzőfajtákat (az igenévképzőket, valamint a ható, a műveltető és a szenvedő igék képzőit), és így adunk minősítést a megmaradt szórésznek. (Részletesebben l. HORVÁTH 2010: 426, 2017: 453.)

A Toldi etimológiai statisztikájának összeállítását néhány évvel ezelőtt többféle okból határoztam el. Egyrészt Arany János születése kétszázadik évfordulójának, az Arany-évnek a megünneplése ösztönzött rá. Másrészt hasznosnak, sőt szükségesnek éreztem, hogy – a korábbi statisztikák többségével ellentétben – egy teljes és terjedelmes műben mérjem fel az eredetkategoróriák arányait, mégpedig olyanban, amelyet nagyon sokan ismernek és szeretnek. A Toldit abban a reményben választottam, hogy elemzésem elvei, módszerei és eredményei hasznosíthatók lesznek az oktatásban is az általános iskolától kezdve egészen a doktoranduszképzésig, persze a megfelelő szinthez igazítva. A Toldi választása mellett szólt az is, hogy előkészítésképpen nem kellett cédulázó jellegű munkát végezmem, hiszen alapanyagul rendelkezésemre állt PÁSZTOR EMIL Toldi-szótára (ToldiSz.).

Kezdetől fogva biztos voltam abban, hogy a Toldi etimológiai statisztikáját a szavakra, a lemmákra és a tövekre egyaránt kiterjedve érdemes és célszerű elkészítenem. Ugyanakkor tisztában voltam azzal, hogy ez igencsak időigényes feladat. Természetesen arról sem feledkezhettem meg, hogy eredményeimnek legalább egy részét illő az Arany-évben közzétennem; emellett ügyelnem kellett arra, hogy ez beleférjen egy előadásnak, illetve folyóiratcikknek a keretébe. Mindezt figyelembe véve 2017-ben úgy döntöttem, hogy a Magyar Nyelvtudományi Társaság felolvasó ülésén elhangzó előadásomban és annak írott változatában (HORVÁTH 2017) a szavak és a lemmák eredetstatisztikájának bemutatására keríték sort. Hangsúlyoztam azonban (HORVÁTH 2017: 453, 465), hogy a tö eredet-statisztika összeállításáról és publikálásáról sem mondok le, csupán későbbre halasztom. Sajnos a halasztás különféle okokból hosszabb lett, mint reméltem. Most azonban végre bemutathatom a Toldi-beli tövek eredetstatisztikáját mint az előző tanulmány kiegészítőjét és párját.

* A tanulmány elkészítését az NKFI Új magyar etimológiai szótár. Második ütem című, K 124127. számú projektuma támogatta. (Ennek a munkálatai, amelyekben senior kutatóként veszek részt, GERSTNER KÁROLY vezetésével folynak a Nyelvtudományi Intézetben.) Köszönöm Hoffmann Istvánnak és másík, anonim lektoromnak támogató véleményüket.

2. A tőeredet-statisztika előkészítése. Mint a bevezetőben említettem, etimológiai statisztikámat a ToldiSz.-ra alapozva állítottam össze. Hogy az eredet-statisztika számára miféle szűréseket, átsorolásokat hajtottam végre az alapanyagon, azt az előző írásomban részletesen ismertettem (HORVÁTH 2017: 451–452), így ennek a megismétlését nem tartom szükségesnek.

A Toldiban meglévő tövek listáját a lemmáké alapján hoztam létre. A tövek megkereséséhez és eredetbeli minősítésükhöz szükség esetén az EWUng. segítségéhez fordultam.

Az összetett lemmákat elemeikre bontottam, beleértve az *és*, *ember* típusú elhomályosult összetételeket is. A származékokról levágtam a képzőt vagy képzőket, függetlenül attól, hogy tőlexémához vagy fiktív tőhöz jutok-e ezzel. Ugyanígy tettem a megszilárdult jeles vagy ragos alakulatok végződésével is. Természetesen bőven akadtak olyan esetek, amelyekben az említett műveletekből egymás után többet is el kellett végezni. Így jutottam el például a *vadmadár-tojás* lemmától a *vad*, *madár* és *tojik* tövekhez, az *üggvel-bajjal* lemmától pedig az *ügy* és *baj* tövekhez.

Vannak olyan eredetkategóriák, amelyeknek a képviselői a közvetlen előzményükhöz képest nem tartalmazznak morfématöbbletet. Ilyen a szófajváltás és a szóhasadás.

A szófajváltás egy-két érintettje egymorfémás. Ilyen az *a* névelő és a *de* kötőszó. Ezeket a lemmákat valójában nem tőre vezetjük vissza, hanem velük azonos alakú etimológiai előzményükre: az uráli korból örökölt *a* mutató névmásra, illetőleg az onomatopoetikus *de* partikulára. Ilyenkor kényszerből, terminológiai egyszerűsítésként az etimológiai előzményt a „tövek” közé sorolom, és ezt tekintem a tőeredet-statisztika részének. Jóval gyakoribb eset azonban, hogy a szófajváltás eredménye végső soron valóban morféμάkra bontható. Ilyenkor a „lemma → tő” irányú átsorolást több lépcsőben, előbb nem morfológiai, majd morfológiai művelettel végeztem el. Például így: *múlva* névutó → *múlva* határozói igenév → *múlik* ige, ez pedig finnugor örökség.

A szóhasadás esetében szintén efféle altípusokkal van dolgunk. A *csíny* lemmának például a *csín* nem a töve, de erre vezethetjük vissza, így a tőstatisztikába a *csín* került a szláv jövevényeket képviselve. A *vörös* esetében viszont két lépcsőből állt a visszavezetés: előbb nem morfológiai művelettel a *véres*-hez, majd a képző elhagyásával a *vér*-hez juthattam el, ez pedig a tövek állományában a finnugor örökség képviselője.

Az elvonással létrejött lemmákra is érvényes, hogy etimológiai előzményük hozzájuk képest nem tő, sőt magától értetődően hosszabb náluk. Ennek ellenére mint etimológiai előzményt az elvonás alapját kellett a „tövek” közé sorolnom. A *vád* esetében például a *vádol* igét, ez pedig szláv jövevény.

A belső keletkezésű lemmák más eredetkategóriáinak reprezentánsait az imént ismertetett elvek és módszerek szerint soroltam be a tövek (illetve „tövek”) eredetkategóriáihoz.

A továbbiakban a tőeredet-statisztika szereplőit az egyszerűség kedvéért egységesen tövekként emlegetem, akkor is, ha nem morfológiai visszavezetéssel jutottam hozzájuk.

A tövek eredet szerinti minősítéséhez – mint vizsgálatom közvetlen előzményében (HORVÁTH 2017) és más etimológiai statisztikai elemzéseimben is – az EWUng.-ot hívtam segítségül. Az EWUng. után megjelent származtatási ötleteket – a szó- és lemmastatisztika elvéhez (vö. HORVÁTH 2017: 453) igazodva – nem vettem figyelembe.

Természetesen a kategorizálásban is alkalmazkodtam szó- és lemmavizsgálatomhoz: az alapnyelvi örökség rétegeit megkülönböztettem egymástól; nem differenciáltam a konkrétabb szóátadó nyelvek szerint az iráni, a török és a szláv jövevényeket; a valószínű eredeztetéseket pedig egyenértékűeknek tekintettem a biztosakkal.

A statisztikát ezúttal is két változatban készítettem el. Az I.-ben helyet kapott a bizonytalan és a vitatott eredetűek kategóriája. A II.-ban viszont megszüntettem őket: a bizonytalan származásúak az EWUng. szerint esetleg számításba vehető eredeztetésüknek megfelelő helyre kerültek, míg a vitatottak az EWUng.-ban első helyen szereplő minősítésüket kapták. Ehhez hasonlóan számoltam fel a II. változat számára a tisztázatlan rétegből örökölt alapnyelvi elemeknek a kategóriáját. (A két változat közötti átsorolás finomabb részleteihez l. főleg HORVÁTH 2000: 180, 321–322, valamint persze a következő pontban a statisztikai táblázatok bemutatását.)

A szó- és lemmastatisztikához hasonlóan a tőstatisztika mindkét változata tartalmaz állományi és előfordulási statisztikát. Az állományi statisztika a tőkészlet elemeivel egy-egy egységként számol, ismétlődésüket, annak a mértékét figyelmen kívül hagyja. Az előfordulási statisztika lényege viszont éppen az ismétlődés, az egyes tövek adatszámja.

3. A Toldi töveinek eredetstatisztikája. A mostani vizsgálatom korábban elkészült párja szerint a Toldi lemmaállománya 2672 tagú, összesen 9601 előfordulással (HORVÁTH 2017: 462).

Mekkora tőkészletre számíthatam ennek alapján? Számszerű becslésre nem mertem vállalkozni, hiszen a „lemmák → tövek” irányú átrendezés kétféle, egymással ellentétes következménnyel jár. Egyfelől a lemmaállományhoz viszonyítva számos új szereplő megjelenését eredményezi. Másfelől bőven akadnak olyan tövek, amelyekhez több vagy akár nagyon sok lemma gyűlik össze, és ez persze az állomány elemszámának csökkenését hozza magával. Régebbi vizsgálataimban (HORVÁTH 2000, 2010) azt tapasztalhattam, hogy a tövek állománya a szavakénál, illetve a lemmákénál jóval kisebb, előfordulási számuk összessége viszont meghaladja amazokét. Ebből kiindulva most is ilyen irányú eltérésekre számíthattam. Az eredmény igazolta ezt a feltevést, sőt nem várt mértékben: mind a tőállomány nagysága, mind a tőelőfordulások száma szinte ugyanolyan arányban különbözik a lemmákétól, ahogyan az első magyar sakk-könyvben (vö. HORVÁTH 2010: 427, 429–430). Ebből a megfigyelésből kiindulva olyan lehetőség adódik, amelyet itt és most csak felvetésként merek és akarok megfogalmazni: a jövőben talán érdemes volna „becslési mutatót” keresni az állomány nagyságára, illetve az előfordulási számra nézve a szavak, a lemmák és a tövek közötti összefüggésekben.

A Toldi tőállománya 1147 tagú, összesen 11707 előfordulással. Egy-egy tőre tehát átlagban 10,21 előfordulás jut.

Az eredetkategoriai képviselőit az 1. táblázat mutatja be.

1. táblázat
A Toldi I. típusú töeredet-statisztikája

Eredet	Tő	%	Előfordulás	%
uráli	77	6,71	3446	29,44
finnugor	154	13,43	2475	21,14
ugor	79	6,89	648	5,54
tisztázatlan rétegből	38	3,31	1073	9,17
örökség összesen	348	30,34	7642	65,28
onomatopoetikus	119	10,37	445	3,80
permi	2	0,17	8	0,07
iráni	11	0,96	65	0,56
török	85	7,41	419	3,58
német	38	3,31	84	0,72
latin	18	1,57	36	0,31
olasz	17	1,48	31	0,26
román	1	0,09	2	0,02
szláv	124	10,81	502	4,29
tisztázatlan nyelvből	5	0,44	14	0,12
közvetített	7	0,61	11	0,09
vándor	12	1,05	27	0,23
jövevény összesen	320	27,90	1199	10,24
bizonytalan	132	11,51	1311	11,20
vitatott	47	4,10	235	2,01
ismeretlen	181	15,78	875	7,47
tisztázatlan összesen	360	31,39	2421	20,68
Összesen	1147		11707	

A következőkben az összkép jellemzésében és a részleteket taglalva természetesen többször hivatkozom a régebbi töeredet-statisztikáimra, valamint a Toldi lemmastatisztikájára. A lapszamos utalás folytonos ismétlésének megelőzésére itt adom meg az I. típusú (azaz a bizonytalanság és a vitatottság kategóriáit szerepeltető) táblázatok lelőhelyeit: a 20. század végi elbeszélésekből összeállított korpuszom tőstatisztikája: HORVÁTH 2000: 326; az első (a 18. századból való) magyar sakk-könyv tőstatisztikája: HORVÁTH 2010: 430; a Toldi lemmastatisztikája: HORVÁTH 2017: 462–463.

A Toldi t ő á l l o m á n y a a főkategóriák reprezentáltságának sorrendjét illetően meglepetéssel szolgál: a tisztázatlan eredetűek képviselője a legnagyobb benne, az állomány tagjainak csaknem a harmada közéjük tartozik. Maga a tekintélyes részesedés viszont nem meglepő: ezt megközelítőt találtam a 20. század végi

elbeszélésekből összeállított szövegmintámban és az első magyar sakk-könyvben is. Az eredmény mindenesetre látványos jele annak, mennyivel több a nehézség, a tisztázatlanság a tövek etimologizálásában, mint a lemmákéban, a lemmaállományban ugyanis a tisztázatlan eredetűek részesedése mindössze 8%-nyi.

A tisztázatlan eredetű tövek állományának a fele ismeretlen származású. Ez a Toldi teljes tőállományából 16%-os részesedést jelent. Szintén 10% feletti a bizonytalan eredetűek képviselése. A vitatott eredetűeké ennél jóval kisebb: az 5%-ot sem éri el.

A tőállományból az alapnyelvi örökség is 30%-nál nagyobb arányban részesedik. Képviselése azonban meglepő mértékben kisebb a 20. század végi elbeszélésekben és a 18. századi sakk-könyvben látottnál, azokban ugyanis megközelítette a 40, illetve az 50%-ot.

Az alapnyelvi rétegek közül a kiemelkedő vezető szerep – a másik két említett korpusz tőállományához hasonlóan – a finnugoré.

A jövevények főkategóriájának képviselése sem sokkal marad el a 30%-tól. Részesedésük nagyobb a 20. század végi elbeszélésekben mértnél, az első sakk-könyvbelinél pedig különösen.

Régebbi felméréseim eredményei alapján és a lemmastatisztika ismeretében biztosan lehetett számítani arra, hogy a jövevénytövek közül szláv eredetűből van a legtöbb, őket pedig a török származásúak követik. A Toldi-beli tőállományban a szlávok a 10%-os küszöböt is átlépik.

Az etimológiai statisztikákban szokásoshoz illően a főkategóriákat tekintve a Toldi lemmaállományában is a belső keletkezésűek domináltak. A „lemmák → tövek” irányú átsorolás természetéből adódóan (és persze szintén a szokásos foratókönyv szerint) azonban ez a főkategória szinte teljesen kiürült, pontosabban az onomatopoetikus elemek őrződtek meg belőle, ezek vették át a helyét. Képviseljük a Toldi tőállományában 10%-os, hasonló a szláv jövevényekéhez. Ez a részesedés az első magyar sakk-könyvben mértnek a többszöröse, a 20. század végi elbeszélésekben mértet szintén felülmúlja.

Az egyes kategóriáknak, főkategóriáknak a képviselői természetesen nagyon különböző mértékben ismétlődnek, ezért a t ő e l ő f o r d u l á s o k között egészen mások az arányok, mint a tőállományban. Mint az 1. táblázat előtt említettem, a Toldiban egy-egy tőre átlagban 10,21 előfordulás jut. Ezzel az ismétlődési hányadossal vethető össze a táblázat segítségével az egyes főkategóriáké: örökség: 21,96; onomatopoetikus elemek: 3,74; jövevények: 3,75; tisztázatlan eredetűek: 6,73. Az alapnyelvi rétegekből örökölt töveknek az ismétlődése tehát az átlagnak a dupláját is meghaladja, míg a másik három főkategória ezzel összefüggésben elmarad az átlagtól.

A főkategóriákra nézve látványos az alapnyelvi örökség térhódítása. Képviselése itt a kétharmadosat is megközelíti. Igaz, a sakk-könyvbéli arányt nem éri el.

Az örökség tényerése miatt a többi főkategória részesedése persze csökken. A tisztázatlan eredetűeké még így is tekintélyes: 20% feletti. A jövevényeké feleakkora. Az onomatopoetikus elemeké 4% alá kerül. Régebbi vizsgálataim anyagában ezekhez nagyjából hasonló arányokat találtam. Csupán az onomatopoetikus elemeknek a sakk-könyvbéli igen kevés előfordulása „lóg ki” a képből, de a szövegtípus tulajdonságai, szaknyelvi jellege miatt ez nem meglepő.

Az alapnyelvi örökség rétegeiből a tőelőfordulások között az uráli dominál. Míg az állományban a képviselőinek száma csak a fele a finnugorokénak, sőt valamicskével még az ugorokétól is elmarad, az összes előfordulást tekintve az uráliak kerülnek a kategóriák listájának élére: részesedésük önmagában is megközelíti a 30%-ot. Ez nem is csoda, hiszen akadnak olyan képviselőik, amelyekből sok lemma származik, köztük igen gyakoriak is. Példaként talán elég a leggyakoribb tőnek, az *a* névmástőnek az esetét említenem. Önálló mutató névmási lemmaként csupán kétszer jelenik meg a Toldiban. Tőként viszont nagyon sok névmásnak, névmási határozószónak az alapja vagy összetevője, és ezek között gyakoriak is vannak: az *úgy* (47 lemmaadattal) és az *ott* (25-tel), valamint az *aki* (22-vel). De a legfontosabbat még nem is említettem: a szófajváltással keletkezett *a* névelő, a lemmák közül a leggyakoribb is az *a* névmástőre megy vissza a maga 666 (!) adatával.

A Toldi uráli származású tövei közül, mint a példa elárulta, az *a* névmástő a leggyakoribb 877 adattal. A listán palatális párja, az *e* követi, szintén magasan kiemelkedve a tövek sorából (719 adattal; magának az *e* mutató névmásnak a 24 előfordulásán kívül ezekkel: *is* 131, *és* 89, *egy* névelő 59, *sem* 54, *így* 35, *egy* számnév 32, *én* 28 stb.). Százas nagyságrendű adat tartozik még a *mi* kérdő névmási tőhöz (381; magának a névmásnak a 20 adatán kívül ezek: *mint* kötőszó 46, *mert* 31, *mintha* 28, *minden* 26, *mely* 21, *mikor* 21 stb.), a *ho-* névmástőhöz (315; benne: *hogy* kötőszó 94, *ha* 54, *mintha* 28, *hanem* 24, *hogy* 'hogyan, ahogy' 24 stb.), továbbá a *föl-*höz (110; egyebek mellett minden megjelenő *fel* igekötős igével). Ötven feletti még a *megy* (63) és a *lát* igéhez (51), továbbá a *fél* névszóhoz (51) összegyűlő előfordulások száma is.

A finnugor rétegből örökölt tövek részesedése az összes előfordulást tekintve kisebb az uráliakénál, de szintén 20% feletti. A legtöbbször előforduló képviselőjük a *nem* tagadószó, amelyhez számos összetett lemmától is gyűlnek adatok, noha kevés olyantól, amely gyakorinak mondható (összesen 343 adat; magának a tagadószónak a 181 előfordulásán kívül idetartoznak: *sem* 54, *hanem* 24 stb.). A finnugor eredetű tövek előfordulási listájának a második helyén az az elem áll, amelyet azonosító címkeként az EWUng. *mög* címszavával nevezek meg (228 tőelőfordulás; idetartoznak a *meg* igekötős igék, továbbá: *még* 29, *meg* kötőszó 12 stb.). Kétszáz feletti a *van-*hoz sorolható adatszám is (206; *van* 151 + *vagy* kötőszó 25, *valami* 6 stb.). Százas nagyságrendben fordul még elő a *má-* tő (149; benne: *most* 40, *már* 24, *majd* 22, *más* 13, *ma* 11, *másik* 11 stb.). Ötvennél több tőadat képviseli a *bél* (65) és a *mag* (58) főnevet, valamint a *lesz* igét (61) és a *ne* tiltószót (53).

Az ugor korból örökölt tövek állománya kissé nagyobb az uráliakénál. Az ismétlődés azonban sokkal kevésbé jellemző rájuk, mint uráli és finnugor társaikra, tehát az előfordulásokra nézve messze elmaradnak tőlük. Leggyakoribb képviselőik közül is csak kettőnek van ötvennél több adata: a *kiv-*nek nevezhetőnek (89; a *ki* igekötős igék, valamint néhány határozószó) és a *jó-*nak (68; *jó* 34 + *jól* 5, *jobban* 3 stb., ill. *igaz* 1, *jog* 1 stb.).

Az alapnyelvből örökölt tövek között akadnak olyanok, amelyekkel kapcsolatban nincs tisztázva, hogy melyik rétegből valók. A Toldiban állományuk

az ugorokénak a felénél is kisebb, sűrű ismétlődésüknek köszönhetően azonban az előfordulásokra nézve bőven megelőzik amazokat. Rendkívül gyakran megjelenő képviselőik is vannak: *az* (368; *az* mutató névmás 105 + *az* névelő 157, *azután* 24, *akkor* 20, *azért* 12 stb.) és az *el* (206; az *el* igekötős igék és származékaik, valamint néhány határozószó és névutó). Ötvennél több adat képviseli még a *ki* névmást (87; *ki* 50 + *aki* 22, *senki* 11, *akárki* 2, *valaki* 2), az *ez* mutató névmást (64; *ez* névmás 43 + *ennyi* 4, *ekképp* 3, *ezalatt* 3 stb.) és az *út* főnevet (56; *út* 13 + *azután* 24, *utána* 5 stb.). Az *az* és *ez* esetében a rétegbeli vitatottság oka az, hogy nincs tisztázva, mikor csatlakozott a *-z* névmásképző a tőhöz (a részleteket l. az EWUng.-ban; a tőstatisztikai besoroláshoz vö. még HORVÁTH 2002: 275). Szeretném itt előrebocsátani, hogy a tisztázatlan rétegből örökölt tövek leggyakoribb képviselőinek nagy része a II. típusú statisztikában a finnugor réteghez került át, és ez persze nagyban hozzájárult a finnugor tövek ottani térnyeréséhez.

A főkategóriák között a belső keletkezésük helyébe lépő *onomatopoeitikus* elemek ismétlődési hányadosa sokkal kisebb az átlagosnál. Érthető tehát, hogy előfordulási részesedésük jóval elmarad az állománybelitől. Kiemelkedően leggyakoribb képviselőjük a *de* kötőszónak az előzményeül szolgáló *de* partikula (84; *de* kötőszó 75 + *dehogy* 6, *dejszen* 1, *nemde* 1, *nemdenem* 1). Utána a *jaj* következik (24; *jaj* 21 + *eljajdul* 1, *jaj-keserves* 1, *jajszó* 1). A hangutánzó ígétövek közül a *sír* jelenik meg legtöbbször (16; *sír* 9 + *sírat* 3, *sírás* 2, *sírály* 1, *sír-ri* 1).

A *jövevények* ismétlődési hányadosa ugyanolyan kicsi, mint az onomatopoeitikus töveké. A *jövevénytövek* közül, akárcsak az állományban, az előfordulásokat tekintve is szláv eredetűből van a legtöbb, őket pedig a törökök követik. Ezen a két kategórián kívül másnak itt nincs is 1%-nál nagyobb képviselete.

A szláv *jövevények* közül egynek az előfordulási száma sem éri el az ötvenet, a húszat viszont haté is: *király* (48; *király* 41 + *királyi* 4, *királyfi* 1, *királyudvar* 1, *kiskirály* 1), *vitéz* (30; *vitéz* 26 + *vitézség* 2, *vitézi* 1, *vitézlő* 1), *cseh* (26), *dolog* (23; *dolog* 22 + *dolgozik* 1), *szolga* (23; *szolga* 21 + *szolgál* 1, *szolganép* 1), *beszéd* (22; *beszéd* 10 + *beszél* 8, *elbeszél* 4).

A török származásúak közül négynek van húsznál több adata: *kép* (32; *kép* 10 + *képzél* 4, *miképp* 4, *ekképp* 3 stb.), *erő* (31; *erő* 14 + *erős* 13, *erősít* 1, *erőskegyetlenül* 1, *erősség* 1, *erőtlen* 1), *bú* (22; *bú* 5 + *bús* 5, *búsul* 3 stb.), *idő* (21; *idő* 7 + *midőn* 9, *amidőn* 2, *esztendő* 2, *koránsem* 1 [vö. EWUng. *korán*² a.]).

A többi *jövevénykategóriában* nincs olyan tő, amelynek az előfordulási száma húsz feletti.

A kategóriák gyakorisági listáján a török elemeket a németek követik. A leggyakoribb képviselőik: *marha* (9; *marha* 7 + *aprómarhanyáj* 1, *marhavásár* 1), *kanna* (6), *pár* (6), *pint* (6; *pint* 3 + *pintes* 2, *őtpintes* 1), *tarsoly* (6), *tenyér* (6; *tenyér* 5 + *tenyérnyi* 1).

Iráni elemből a Toldi tőállományában nincs sok, de a *jövevények* közül ennek a kategóriának a legnagyobb az ismétlődési hányadosa. A legtöbbször megjelenő reprezentánsa az *asszony* (14; *asszony* 10 + *anyámasszony* 2, *özvegyasszony* 2). Érdekes a *tíz* esete: önálló lemmaként nincs meg a Toldiban, viszont tőként annyi adatot gyűjt magához, hogy az a második leggyakoribb iráni elemmé avatja (12;

tizedik 3, *tizenkét* 3, *kilenc-tíz* 1, *tizedel* 1, *tizenegyedik* 1, *tizenkettedik* 1, *tíz-tizenkét* 1, de az utóbbi a tő kétszeres jelenléte miatt 2 tőadatként kerül a statisztikába).

A latin és az olasz jövevények a tőállományt tekintve az irániak előtt állnak, sokkal ritkább ismétlődésük miatt azonban az előfordulásokra nézve jóval elmaradnak tőlük.

A latin jövevények közül a *mód* tűnik fel legtöbbször (7; *mód* 6 + *módnélküli* 1). Utána ezek következnek: *cifra* (4), *cimbalom* (4; *cimbalom* 2 + *cimbalmos* 2).

Az olaszok között ezeknek van kettőnél több adatuk: *part* (5), *pajzs* (3), *rest* (3; *rest* 1 + *restell* 2), *vendég* (3; *vendég* 2 + *vendégoldal* 1).

A Toldi tőállományában két permi jövevény van jelen. A gyakoribb a lemmaként is meglévő *kenyér* (6). Az *ezüst* újdonság a lemmakészülethez képest (2; *ezüstkoszorú* 1, *ezüstös* 1).

Akárcsak a lemmaállományban, a tövekében is az egyetlen román elem a *cimbor*a (2 adattal).

Egyes jövevényekkel kapcsolatban nincs tisztázva, melyik az a nyelv, amelyből a magyarba kerülhettek. Ennek a csoportnak az egynél többször felbukkanó reprezentánsai: *bér* (6; *bér* 1 + *béres* 5), *köntös* (3), *rúd* (3).

Szintén akadnak olyan jövevények, amelyek nemcsak egy nyelvből jutottak el a magyarba, hanem talán egy másiknak a közvetítésével is. Az ilyeneket (akárcsak szó- és lemmastatisztikámban) az átadó nyelvektől függetlenül a „közvetített” elnevezésű jövevénykategóriába vontam össze. Egynél többször megjelenő képviselőik: az olaszból talán latin közvetítéssel is átvett *tréfa* (3; *tréfa* 2 + *tréfaság* 1), valamint a latinból esetleg szláv közvetítéssel is átkerült *alamizsna* (2) és *pogány* (2).

A vándorszók is a (tágabb értelemben vett) jövevények főkategóriájába sorolhatók. Képviselőik a tőállományban az irániakhoz hasonló számban vannak jelen, de sokkal ritkábban ismétlődnek azoknál. Közülük ezek a leggyakoribbak: *paripa* (5), *levente* (4), *szoba* (4; *szoba* 3 + *hálószoba* 1), *tarisznya* (4).

A tisztázatlan eredetűek főkategóriájában a bizonytalan származású tövek ismétlődési hányadosa lényegében megfelel a Toldi egész tőanyagára érvényes átlagnak, a vitatott és az ismeretlen eredetűeké ellenben messze elmarad attól. Emiatt a bizonytalan eredetűek részesedése nagyjából akkora az előfordulások között, mint az állományban, míg a vitatott és az ismeretlen származásúaké az állományinak a felét sem éri el.

A bizonytalan és a vitatott eredetű tövek leggyakoribb képviselőiről úgy szölok, hogy a II. típusú (azaz a bizonytalanság és a vitatottság kategóriáit felszámoló) statisztikában érvényes minősítésüket is jelzem.

A bizonytalan eredetűek négy leggyakoribb képviselőjének a finnugor származatása jöhet szóba. Ez a II. típusú statisztikában természetesen jelentős nyereség ennek az alapnyelvi rétegnek.

A lista elején őríasi főlényvel és hatalmas adatszámmal a *s* áll (513; *s* 132 + *is* 131, és 89, *sem* 54, *nincs* 18, *se* 12, *sincs* 10 stb.; megjegyzés: a *sincs* a *s* elemet kétszeresen tartalmazza, tehát az adatszám valójában 10 helyett 20). Majdnem százszor fordul elő a *nagy* (98; *nagy* 85 + *nagyon* 9, *nagyapa* 1, *nagyfejű* 1, *nagy-*

le 1, *nagyúri* 1). Utána a *le* következik (49; a *le* igekötős igék tartoznak hozzá, valamint az imént említett *nagy-le*). A negyedik pedig a *vég* (38; *vég* 10 + *végre* 13 és sok más egy-két adattal).

A bizonytalan eredetűek olyan képviselőiből, amelyekre nézve nem a finnugor származás vehető számításba, a talán törökből átvett *ér* igének van a legtöbb adata (29; *ér* ige 5 + *érez* 5, *ért* 5, továbbá kisebb adatszámú származékok és igekötős összetételek).

A bizonytalan származású tövek között persze olyanok is akadnak, amelyek talán onomatopoetikus eredetűek. Közülük háromnak van kettőnél több adata; mindhárom tő fiktív: *csill-* (3; *csillapul* 3), *kár-* (3; *káromkodás* 1, *káromkodik* 1, *káromol* 1), *sim-* (3; *simogat* 2, *végigsimogat* 1).

Meglepőnek tűnhet, de olyan bizonytalan eredetű tövek is vannak, amelyek a II. típusú statisztikában az ismeretlen származásúakhoz kerülnek. Ez akkor történhet meg, ha az EWUng. a „bizonytalan eredetű, talán származékszó fiktív tőből” minősítést használja, a fiktív tövet pedig ismeretlen eredetűnek tartja. Az efféle tövekből a Toldiban az *ism-* a leggyakoribb (7; *ismer* 3, *megismer* 2, *ismeretlen* 1, *kiösmerszik* 1).

A vitatott eredetű tövekből az *n*-nel címkézhető a leggyakoribb. Ez az *én* személyes névmás utótagja (34; *én* 28 + *engem* 3, *énrám* 1, *énrólam* 1, *enyém* 1). Többféle származtatása közül leginkább az uráli névmási eredet jön számításba.

A gyakorisági listán szorososan követi az *úr* (31; *úr* 14 + *ország* 7, *úri* 3, *bátyádurad* 1, *nagyúri* 1, *országút* 1, *teremburádat* 1, *úradta* 1, *uramfia* 1, *úr-maga* 1); ennek a török eredeztetése a valószínűbb. Az utána következő *ör-* tőnek pedig a finnugor (25; *öröm* 12, *örül* 9, *megörvend* 1, *örömet* 1, *örömkönnnyü* 1, *örömszaj* 1).

Az olyan vitatott eredetű tövek közül, amelyeknek az onomatopoetikus magyarázatuk a meggyőzőbb, a *báty-* fordul elő legtöbbször (16; *bátya* 14, *bátyádurad* 1, *testvérbátya* 1).

Egyszer még az a különleges helyzet is előáll, hogy a „vitatott” minősítést a II. típusú tőstatisztika számára „ismeretlen”-re kell cserélni; a részletekhez l. az EWUng. *üld* szócikkét (a visszavezetett lemma az *üldöző*, 1 adattal).

A tisztázatlan származásúak harmadik csoportjából, az ismeretlen eredetű tövek közül gyakoriságával magasan kiemelkedik a *csak* (74; *csak* 61 + *ugyancsak* 4, *csakhamar* 3, *csakhogy* 2, *csakúgy* 2, *csaknem* 1, *nemcsak* 1). Rajta kívül sok adata van még a *ked-* tőnek is; ez a *pedig*-et és körét vonja magához (47; *pedig* 43, *pediglen* 2, *éspedig* 1, *mégpedig* 1).

Az I. típusú tőeredet-statisztika lezárásaként itt mutatom be a Toldi leggyakoribb töveinek összegzett listáját. A kategóriák szemléjében közölt részleteket, magyarázatokat nem ismétlem meg, de zárójelben megadom az eredetminősítéseket, szükség esetén (kettősponttal bevezetve) megtoldva őket a II. típusú statisztikában érvényessel: 1. *a* 877 (uráli) | 2. *e* 719 (uráli) | 3. *s* 513 (bizonytalan; finnugor) | 4. *mi* (kérdő) névmás 381 (uráli) | 5. *az* 368 (tisztázatlan rétegű örökség; finnugor) | 6. *nem* tagadószó 343 (finnugor) | 7. *ho-* 315 (uráli) | 8. *mög* 228 (finnugor) | 9–10. *el* 206 (tisztázatlan rétegű örökség; finnugor); *van* 206

(finnugor) | 11. *má-* 149 (finnugor) | 12. *föl* 110 (uráli) | 13. *nagy* 98 (bizonytalan: finnugor) | 14. *kiv* 89 (ugor) | 15. *ki* névmás 87 (tisztázatlan rétegű örökség: finnugor) | 16. *de* partikula 84 (onomatopoetikus) | 17. *csak* 74 (ismeretlen) | 18. *jó* 68 (ugor) | 19. *bél* 65 (finnugor) | 20. *ez* 64 (tisztázatlan rétegű örökség: finnugor) | 21. *megy* 63 (uráli) | 22. *lesz* 61 (finnugor) | 23. *mag* 58 (finnugor) | 24. *út* 56 (tisztázatlan rétegű örökség: finnugor) | 25. *ne* tiltószó 53 (finnugor) | 26–27. *fél* névszó 51 (uráli); *lát* 51 (uráli). Önmagáért beszél, hogy ennek a listának 23 (!) eleme az alapnyelvi örökséghez tartozik. Ráadásul a két jelenlévő bizonytalan származású tő is talán örökség. Az onomatopoetikus és az ismeretlen eredetű töveknek egy-egy reprezentánsuk szerepel a listában, a jövevényeknek egy sem.

A II. típusú tőeredet-statisztika képét a 2. táblázat rögzíti.

2. táblázat

A Toldi II. típusú tőeredet-statisztikája

Eredet	Tő	%	Előfordulás	%
uráli	108	9,42	3634	31,04
finnugor	225	19,62	4300	36,73
ugor	114	9,94	908	7,76
örökség összesen	447	38,97	8842	75,53
onomatopoetikus	140	12,21	491	4,19
permi	3	0,26	9	0,08
iráni	19	1,66	115	0,98
kaukázusi	1	0,09	2	0,02
török	109	9,50	591	5,05
német	46	4,01	110	0,94
latin	22	1,92	43	0,37
francia	1	0,09	1	0,01
olasz	19	1,66	41	0,35
román	1	0,09	2	0,02
szláv	128	11,16	522	4,46
tisztázatlan nyelvből	1	0,09	1	0,01
közvetített	1	0,09	1	0,01
vándor	13	1,13	28	0,24
jövevény összesen	364	31,73	1466	12,52
ismeretlen	196	17,09	908	7,76
Összesen	1147		11707	

Ez a kép az I. típusú statisztikának a 2. pontban ismertetett elvek szerint való átrendezésével alakult ki. A bizonytalanok és a vitatottak kategóriájának megszűnésével a tisztázatlanok főkategóriája az ismeretlen eredetűekének adta

át a helyét. Az átsorolás az alapnyelvi örökség tisztázatlan rétegét is felszámolta. Nem tűnt el viszont teljesen a tisztázatlan nyelvből származó jövevények kategóriája, mivel egy képviselője ismeretlen nyelvből való: ez a *bükk*.

Nem lett egészen üres a közvetítéssel átvett jövevények kategóriája sem: megmaradt benne az oszmán-törökből valószínűleg szerbhorvát közvetítéssel átkerült *korbács*. A többi olyan jövevény azonban, amely az I. statisztikában még ehhez a kategóriához tartozott, átkerült a „fő” átdadó nyelv reprezentánsaihoz, mivel közvetített voltuk az EWUng. szerint csak esetleges lehetőségként vetődik fel.

A bizonytalan és a vitatott eredeztetések felszámolása más irányú következménnyel, új jövevénykategóriák megjelenésével is járt. Az egyik a kaukázusi: képviselője az I. statisztikában a bizonytalan származásúak közé tartozó *réz* (2 adattal). A másik pedig a francia: képviselője, a *kilincs* az I.-ben a vitatottaknál szerepel.

Az I. és a II. statisztika közötti átrendezéssel minden megmaradó főkategória részesedése nő. Az igazi nyertesnek azonban az alapnyelvi örökség főkategóriája tekinthető: az állományban 8, az előfordulások között pedig 10 százalékponttal gyarapszik a képviselete. Az utóbbi azt jelenti, hogy a Toldi tőelőfordulásainak a háromnegyede az alapnyelvi örökséghez tartozik.

Ahogy az I. tőstatisztika elemzésének néhány részlete alapján feltételezhető volt, elsősorban a finnugor réteg térnyerése szembetűnő: mind az állományban, mind az előfordulásokat tekintve mindegyik kategóriáét megelőzi a részesedése.

4. Összegzés és kitekintés. A tőeredet-statisztika két változatának elkészítésével és bemutatásával teljessé vált a Toldi etimológiai statisztikája. Az első magyar sakk-könyv anyagának a vizsgálata után (vö. HORVÁTH 2010) ez a második olyan gyakorlati felmérésem, amelyet nem szövegrészletek vagy rövid nyelvmelék, hanem terjedelmes és egész mű alapján végeztem a szavakra, a lemmákra és a tövekre is kiterjedve, beleértve állományukat és előfordulásait.

Bízom abban, hogy ez a két felmérés, különösen a Toldié, összehasonlítási alapul szolgálhat más hasonlókhöz: mind az elveket és a módszereket, mind pedig a konkrét, számszerű eredményeket tekintve.

Úgy vélem, a Toldi etimológiai statisztikája az oktatásban is hasznosítható, persze az adott szint lehetőségeihez, igényeihez illő válogatással.

Természetesen a Toldi tőeredet-statisztikájának elkészülte sem jelenti azt, hogy az etimológiai kategóriákra nézve a művel kapcsolatban nem maradtak nyitott, kutatásra érdemes kérdések. Példaként most csak kettőt említek közülük. Meg kellene vizsgálni, miért kisebb a Toldiban a belső keletkezésű lemmák részesedése, mint amilyenre más felméréseim alapján számíthattam (vö. HORVÁTH 2017: 464). Érdekes lenne továbbá annak a feltárása is, mekkora a műben az alapnyelvi örökség részesedése akkor, ha a körét a fogalomszói elemekre korlátozzuk.

Kulcsszók: Toldi, etimológiai statisztika, a tövek eredetstatisztikája, állományi statisztika, előfordulási statisztika.

Hivatkozott irodalom

- EWUng. = *Etymologisches Wörterbuch des Ungarischen* 1–2. Hrsg. BENKŐ, LORÁND. Akadémiai Kiadó, Budapest, 1993–1995. + *Register*. Akadémiai Kiadó, Budapest, 1997.
- HORVÁTH LÁSZLÓ 2000. Etimológiai kategóriák arányai mai elbeszélésekben. *Magyar Nyelv* 96: 170–181, 316–332.
- HORVÁTH LÁSZLÓ 2002. Az Ómagyar Mária-siralom etimológiai statisztikája. *Magyar Nyelv* 98: 265–282.
- HORVÁTH LÁSZLÓ 2010. Az első magyar sakk-könyv etimológiai statisztikája. *Magyar Nyelvőr* 134: 421–436.
- HORVÁTH LÁSZLÓ 2017. A Toldi szókészletének eredetéről. *Magyar Nyelv* 113: 450–466. <https://doi.org/10.18349/MagyarNyelv.2017.4.450>
- ToldiSz. = PÁSZTOR EMIL, *Toldi-szótár: Arany János Toldijának szókészlete*. Tankönyvkiadó, Budapest, 1986.

On the origin of the stems in Arany's *Toldi*

The author has been working on etymological statistics for more than twenty years now. In 2017, he compiled a statistics of origin of the most widely known Hungarian epic poem, János Arany's *Toldi*, with respect to the inventory and occurrence (types and tokens) of the words and lemmas occurring in it. The present paper is a counterpart of the 2017 paper, complementing it in some respects. This time, the author explores the proportions of categories of origin with respect to the stems of the words included in the poem. On the one hand, the results are compared to those of the earlier study, and on the other hand, the present paper is intended to serve as a basis of comparison for similar studies in the future. The author hopes that the results of his analyses can be utilised at various levels of education from primary schools to doctoral studies.

Keywords: Toldi, etymological statistics, statistics of origin of stems, inventory statistics, occurrence statistics.

HORVÁTH LÁSZLÓ
Nyelvtudományi Intézet