

Az első magyar sakk-könyv etimológiai statisztikája*

1. Ebben az előadásban két, számomra kedves kutatási téma fonódik össze: az egyik a magyar sakknyelv és története; a másik az etimológia, azon belül az etimológiai statisztika.

Az eredetkategóriák előfordulási arányai szótárakban és szövegek feldolgozásával egyaránt vizsgálhatók. A címből könnyű kitalálni, hogy mostani elemzésem a második típushoz, azaz a használati statisztikák közé tartozik. Etimológiai szempontú használati statisztikát először tíz évvel ezelőtt állítottam össze, a 20. század utolsó évtizedében született elbeszélések részletei alapján (Horváth 2000). Mivel abban az írásomban áttekintettem a hazai etimológiai statisztikai kutatások történetét, beleértve fontos elméleti-módszertani részleteket is, a szakirodalmi előzmények közül ezúttal csak keveset emelek ki, azokat is csupán röviden.

* Előadásként elhangzott az MTA Nyelvtudományi Intézetében 2009. november 17-én.

Az etimológiai statisztika alapjául szolgáló korpusz – akár egyetlen szövegből, akár több szöveg részleteiből áll – nem szokott terjedelmes lenni. Ez érthető, hiszen az eredetstatisztika készítése nagy figyelmet igénylő, aprólékos, fáradságos munka. Különösen az volt a számítógépek megjelenése előtt, a hagyományos, cédulázó gyűjtés és a kézi adatfeldolgozás korában. Nem véletlen, hogy az eddigi egyetlen olyan használati eredetstatisztika, amelynek korpusza könyvméretű, már számítógép segítségével készült. Szerzője egyik tehetséges tanítványom, Csiszár Gábor (2002): szakdolgozatában Kosztolányi első verseskötetének etimológiai statisztikáját állította össze, 13 440 szelőfordulását és 16 410 tőelőfordulását. Munkája azonban publikálatlan.

A gyakorisági vizsgálatok nagyobb része a választott korpuszt önmagában, a szinkrón vagy diakrón összehasonlításra, vagy meg is valósítják azt. A lehetőség már Tolnai Vilmos nevezetes eredetstatisztikájában (1924) is megvolt. A diakrón egybevető elemzés klasszikusa Benkő Loránd tanulmánya (1962) négy korszak egy-egy szövegének szembesítésével. Ennek a hatására készítette el tanítványaival Hajdú Mihály hét bibliafordítás összehasonlító etimológiai vizsgálatát (1997), és magam is ettől indítva vállalkoztam a Halotti beszédnek és Kosztolányi azonos című versének az etimológiai statisztikai összevetésére egy eddig publikálatlan előadásomban (Horváth 2004). Az Ómagyar Mária-siralommal foglalkozó elemzésemben (Horváth 2002) viszont csak alkalmi megjegyzésekkel utaltam a mai elbeszélésekben mért arányokra (Horváth 2000).

Az említett előzményeket és körülményeket figyelembe véve döntöttem úgy, hogy elkészítem egy teljes könyv szókészletének etimológiai statisztikáját. Választásom az első magyar sakk-könyvre esett. A mű 1758-ban jelent meg Budán, Landerer Ferenc Leopold nyomtatásában, *Sách, avagy királyos játéknak szabott rend-tartási* címmel. Nyelvezetével már 2008-ban tartott intézeti előadásomban is foglalkoztam. Mivel az akkor elhangzottak bővebb, írott változatban már megjelentek a *Magyar Nyelvben* (Horváth 2010), most csak egy-két fontos részletet ismétel meg.

A *Sách* a címlapot követően 37 kis (A/6-os) szövegoldalból álló könyvecske. Minden bizonnyal eredeti magyar alkotás, azaz nem fordítás; afféle átmenet a szabálykönyv és a kezdőknek szóló tankönyv között. Szerzője ismeretlen. Nyelvhasználata alapján megkockáztattam azt a feltevést, hogy talán a Kisalföld szülötte lehet.

A kiadás példányszáma ismeretlen; jelenleg három példány létezéséről tudunk biztosan, egy negyediké pedig valószínű. A könyv feltehetőleg nem volt széles körben ismert. Hosszú időre még a létezése is feledésbe merült. Megjelenése után 133 évvel, vagyis 1891-ben fedezte fel a nyilvánosság számára Makovecz Gyula, a milleniumi évek kiemelkedő sakkozója és lapszerkesztője. – A magyar sakkterminológia tudatos megalapozására a *Sách* ismerete nélkül, 1859-ben és 1860-ban, a Vasárnapi Újság sakkrovatának előkészítésével és megindításával, valamint a második (akkor elsőnek hitt) magyar nyelvű sakk-könyvnek a megjelentetésével került sor, Cseresnyés István és Rozsnyay Mátyás érdeméért.

Azt, hogy a *Sách* szókészletét választottam elemzésem tárgyául, a sakk és a sakknyelv iránti érdeklődésemem kívül a Kazinczy-év motíválta. Kíváncsi voltam, hogy az ő születésénél csupán egy évvel korábban, még a nyelvújítás előtt, de már nem messze attól, milyen lehetett egy úttörő vállalkozásnak mondható ismeretterjesztő munka szókészletének eredetbeli összetétele.

A gyakorisági vizsgálatot több lépésben végeztem el, tehát ennek megfelelően mutatom be. Az előkészítés után először az eredetkategóriáknak a könyv teljes szókészletére érvényes arányait állapítottam meg. Az eredményt összevettem a korábbi vizsgálataimban, más nyelvemlékekkel kapcsolatban mértekkel. Ezt követően kiválasztottam a sakszókincshez tartozó elemeket, és körükben külön is készítettem etimológiai statisztikát, az eredményt pedig összevettem a könyv egészére érvényessel.

2. Az előkészítő szakaszban döntenem kellett arról, hogy az etimológiai kategóriák arányait mire nézve kívánom megállapítani: a szövegben közvetlenül megjelenő szavakat vagy töveiket tekintve-e. Az elhatározás nem volt nehéz: a szakirodalmi előzmények és saját korábbi tapasztalataim

alapján nyilvánvalónak tartottam, hogy mind a szóstatisztikának (2.1.), mind a tőstatisztikának (2.3.) az elkészítése célszerű. Újdonságként egy közbülső fokozatot is beiktattam: a lemmastatisztikát (2.2.). – Azt is fontosnak tartottam, hogy ezekben a részvizsgálatokban mind **állományi**, mind **előfordulási statisztika** készüljön. Az állomány felmérése azt tisztázza, melyek azok a szavak, amelyek a szövegben jelen vannak, illetőleg amelyekre a szöveg szavai visszavezethetők; ilyenkor a készlet elemeit úgy tekintjük, mintha mindegyikük csupán egyszer szerepelne. Ezzel szemben az előfordulási statisztika tekintettel van az ismétlődésekre: az elemek minden felbukkanását figyelembe veszi.

Természetesen lényeges kérdés a szószámlálás módja, vagyis az, mit tekintünk külön egységnek, és mit nem. Az egybe- és különírás tekintetében a Sách nem mondható következetesnek. Úgy tűnik, az ingadozások hátterében a helyesírás kialakulatlansága, bizonytalansága áll, nem pedig grammatikai természetű különbség. Úgy gondolom tehát, nem mondható bántóan anakronisztikus megoldásnak, hogy a szószámláláskor az ingadozó eseteket igyekeztem a mai helyesírási norma szellemében egységesíteni. – A sakk-könyv sok korabeli nyomtatványhoz hasonlóan, de a mai szokástól eltérően kedveli bizonyos viszony- és segédszók kötőjeles kapcsolását; például így: □*zám-zerint-való* (Sách A3a), *Királynénak-is* (A2b). Az ilyen eseteket a ma szokásos különírásnak megfelelően kezeltem. Az igekötős igéket pedig a könyvbéli írásmódtól és az igekötő pozíciójától függetlenül mindig egybeírtaknak tekintettem.

Az egyes szavak, lemmák, tövek eredetminősítését a hagyományos cédulázáshoz hasonlóan, de a csoportosítás és az összegzés meggyorsítása érdekében a számítógép rovatkezelő programja (Excel) segítségével végeztem el.

Az eredetminősítésekben lehetőleg az EWUng.-ot követtem; még azokban a ritka esetekben is, amelyekben a saját véleményem eltér a szótártétől. Ha a vizsgált elem az EWUng.-ban nem szerepel, akkor minősítéséről alaktani felépítése alapján döntöttem. Olyan eset szinte nem is adódott, amelyben sem a szótárt, sem a morfológiát nem hívhattam segítségül. (Bizonyos értelemben kivétel az *egy* névelő, valamint a *s* viszonyzó; megítélésükről, kezelésmódjukról l. Horváth 2000: 320, ill. 176.) – A „valószínű” eredetminősítést mindenütt egyenértékűnek tekintettem a biztossal.

Az egyes elemek eredetminősítését befejezve összegezhettem, táblázatokba foglalhattam az eredményeket. Mind a szavakat, mind a lemmákat, mind a töveket illetően kétféle statisztikát készítettem, két-két táblázatot állítottam össze. Ezek a bizonytalan és a vitatott eredetű elemeknek a kezelésmódjában különböznek egymástól. Az 1. típusban az összes bizonytalan eredetű elem közös, más kategóriáktól elkülönülő csoportba került, ugyanígy a vitatott eredetűek is. A 2. típusban viszont megszüntettem a „bizonytalan eredetű” és a „vitatott eredetű” kategóriát: elemeiket szétosztottam más kategóriák között annak megfelelően, mi a legvalószínűbb magyarázatuk. A bizonytalan eredetű elemek az EWUng. szerint esetleg számításba vehető származtatásuk szerint kapták meg besorolásukat. A vitatott etimológiák körében pedig a szótárban első helyen álló magyarázatot vettem figyelembe.

A kétfajta statisztikai besorolást szeretném egy-egy példával is megvilágítani. – Az *út* ige az EWUng. szerint bizonytalan eredetű, esetleg ősi örökség a finnugor korból: így a szó az 1. statisztikában a „bizonytalan”, a 2.-ban a „finnugor” minősítést kapta. – Az *öröm* töve a szótár szerint vitatott eredetű: első magyarázata szerint finnugor örökség, a második szerint török származású. Az 1. statisztikában ezért a „vitatott”, a 2. statisztika számára pedig a „finnugor” minősítést adtam.

Tisztázatlanság fennállhat etimológiai főkategórián belül is. A *had* szó például biztosan ősi örökség, de nem tudjuk, hogy a finnugor vagy az uráli korból való-e. Minősítése az 1. statisztika számára: „ősi, tisztázatlan rétegből”. Mivel az első lehetőség a valószínűbb, a minősítés a 2. statisztika számára: „finnugor”.

2.1. A szóstatisztika összeállításakor minden képzett alakulatot külön egységként vettem figyelembe, beleértve a ható, a műveltető és a szenvedő igéket, valamint az igeneveket is. Az igekötős igéket összetett szavakként vettem számításba. Az „igekötő + ige + képző” felépítést viszont

– például: *elnyerhet, felkapás, elzárva* – származéknak, szóképzés eredményének, tehát *elnyer-het, felkap-ás, elzár-va* tagolásának minősítettem, nem pedig összetételnek.

Az „ősi örökség magyar képzővel” kategóriát (pl.: *mond, lesz, fekete, foszt*) egységesnek tekintettem, nem differenciáltam annak megfelelően, hogy a tő melyik alapnyelvi rétegbe tartozik.

Közismert tény, hogy az iráni, a török és a szláv jövevények esetében sokszor igen nehéz vagy éppen lehetetlen megnevezni egyetlen konkrét nyelvet, amelyből a magyar a szót átvette. Éppen ezért és a statisztika elaprózásának elkerülésére ezeket a jövevényeket egységes tömbök képviselőiként kezeltem még azokban az esetekben is, amelyekben az átható nyelv közelebből meghatározható (volna). A latinnal és az újlatin nyelvekkel kapcsolatban efféle összevonást eleve nem terveztem; anyagomban az újlatin nyelveket egyébként is csak az olasz képviseli.

Az **1. szóeredet-statisztikai táblázat** elárulja, hogy a Sách szóállományát 796 elem alkotja, 3391 előfordulással. Egy szóra tehát átlagban 4,26 előfordulás jut.

A Sách 1. szóeredet-statisztikai táblázata

Eredetkategoría	Szó	%	Szóelőfordulás	%
uráli	9	1,13	51	1,50
finnugor	17	2,14	166	4,90
ugor	6	0,75	16	0,47
ősi, tisztázatlan rétegből	5	0,63	49	1,45
ősi örökség magyar képzővel	15	1,88	72	2,12
ősi összesen	52	6,53	354	10,44
származék	396	49,75	991	29,22
fiktív tő származéka	7	0,88	10	0,29
elvonás	1	0,13	1	0,03
szilárdulás	99	12,44	367	10,82
szófajváltás	20	2,51	574	16,93
jelentéskülönülés	1	0,13	1	0,03
szóhasadás	5	0,63	57	1,68
összetétel	147	18,47	526	15,51
szóösszevonás	1	0,13	6	0,18
belső, de tisztázatlan	4	0,50	39	1,15
belső összesen	681	85,55	2572	75,85
iráni	1	0,13	1	0,03
török	5	0,63	14	0,41
német	3	0,38	28	0,83
latin	6	0,75	12	0,35
olasz	1	0,13	6	0,18
szláv	15	1,88	172	5,07
vándor	2	0,25	3	0,09
jövevény összesen	33	4,15	236	6,96
bizonytalan	14	1,76	101	2,98
vitatott	4	0,50	81	2,39
ismeretlen	12	1,51	47	1,39
tisztázatlan összesen	30	3,77	229	6,75
Összesen	796	100,00	3391	100,00

A főkategóriákat tekintve a kép arról tanúskodik, hogy a Sách szóállományából a belső keletkezésűek kiemelkedő, 85% feletti arányban részesednek. Képviseletük a szóelőfordulások között valamivel kisebb, de ott is meghaladja a 75%-ot. A 10%-nyi különbség arra vall, hogy a belső keletkezésű elemek átlagára kevésbé jellemző az ismétlődés. A többi főkategóriában ennek az ellentéte figyelhető meg: az előfordulási részesedés nagyobb az állományinál; a 10%-ot viszont csak az ősi örökség képviselőinek az előfordulása lépi túl.

Az eredetkategóriák közül a származékoké, vagyis a képzett szavaké a kiemelkedő vezető szerep: a Sách szóállományának majdnem minden második tagja ilyen. Az előfordulások között szintén a származékok dominálnak, de ott 30% alatti részesedéssel, tehát reprezentánsaik átlagára nem jellemző az ismétlődés.

Az állomány gyakorisági listáján a származékokat hozzájuk hasonlóan belső keletkezésű szavak kategóriái követik: az összetétellel, illetőleg a rag- vagy jelszilárdulással létrejött elemekéi. Más kategóriák részesedése itt nem éri el a 10%-ot. Az ősi rétegek közül finnugor elemből van a legtöbb, a jövevények közül pedig szlávból. A tisztázatlan eredetűek mindhárom kategóriájának 2% alatti a képviselete.

A Sách 2. szóeredet-statisztikai táblázata

Eredetkategória	Szó	%	Szóelőfordulás	%
uráli	9	1,13	51	1,50
finnugor	27	3,39	303	8,94
ugor	8	1,01	22	0,65
ősi örökség magyar képzővel	20	2,51	140	4,13
ősi összesen	64	8,04	516	15,22
származék	396	49,75	992	29,25
fiktív tő származéka	9	1,13	16	0,47
elvonás	1	0,13	1	0,03
szilárdulás	102	12,81	405	11,94
szófajváltás	21	2,64	574	16,93
jelentéskülönülés	1	0,13	1	0,03
szóhasadás	5	0,63	57	1,68
összetétel	148	18,59	527	15,54
szóösszevonás	1	0,13	6	0,18
belső, de tisztázatlan	1	0,13	1	0,03
belső összesen	685	86,06	2580	76,08
iráni	2	0,25	9	0,27
török	6	0,75	18	0,53
német	3	0,38	28	0,83
latin	6	0,75	12	0,35
olasz	1	0,13	6	0,18
szláv	15	1,88	172	5,07
vándor	2	0,25	3	0,09
jövevény összesen	35	4,40	248	7,31
ismeretlen	12	1,51	47	1,39
Összesen	796	100,00	3391	100,00

A szóelőfordulások gyakorisági listáján a származékoké mögött a szófajváltásoké áll a második helyen, csaknem 17%-os részesedéssel. Képviselői közül nem kevés igencsak gyakran ismétlődik; ezt jól mutatja az is, hogy a kategória állományi részesedése csupán 2,5%-os. – A belső keletkezésűek két említett kategóriáját még két ilyen követi 10% feletti eredménnyel: az összetétellel, valamint a szilárdulással létrejötték. Más főkategóriákból előfordulásukat tekintve is a finnugor és a szláv elemek tűnnek ki. – A tisztázatlan eredetűek három kategóriájából egyiknek a részesedése sem éri el a 3%-ot. Azt, hogy a vitatott eredetűek aránya itt majdnem az ötszöröse az állományinak, az az és az ez mutató névmások sok ismétlődése okozza. Az ismeretlen eredetűek képviselője viszont még az állománybeli 1,5%-tól is elmarad.

Ami az egyes szavak gyakoriságát illeti, a Sächban a két határozott névelő a listavezető: az *a* 252, az *az* 116 alkalommal jelenik meg. Oroszlánrészük van abban, hogy a szóelőfordulások között a szófajváltás kategóriája a második helyre ugrik előre. – A szavak listájának a harmadik helyén szintén belső keletkezésű, de összetétellel létrejött elem áll: az *és* kötőszó, 90 előfordulással. Az összes szó közül a negyedik hely a sakkterminológia leggyakoribb eleméé: a szófajváltással létrejött *mezőnek* 88 adata van. Közvetlenül utána következik a leggyakoribb figuratív, a szláv eredetű *király*, 70 előfordulással. – Az ősi örökséghez tartozó elemek leggyakoribb reprezentánsa, a *nem* tagadószó, 58 megjelenésével a hetedik a listán. A tisztázatlan eredetűek közül pedig a már említett két mutató névmás tűnik fel a legsűrűbben: az *az* 44 és az *ez* 35 adata a tizedik, illetőleg a tizenharmadik helyhez elég.

A 2. **szóeredet-statisztikai táblázat**ot a korábban ismertetett elvek szerint, a bizonytalan és a vitatott származatások felszámolásával hoztam létre. Így lépett elő benne főkategóriává az ismeretlen eredetű szavaké a tisztázatlan származásúaké helyett.

Az átrendezéssel a finnugor elemek nyerték a legtöbbet, különösen az előfordulásokat tekintve: ottani részesedésük 9%-os lett. Az ősi örökség együttes képviselője elsősorban ennek köszönhetően vált 15%-ossá. Különösen fontos erősítésként került át (a vitatottaktól) a finnugorok körébe az *az* és az *ez* mutató névmás, de az ő táborukat erősítette például az 1. statisztika szerint még az ősi szavak tisztázatlan rétegéhez tartozó *had* (22 adattal) és *két* (20) is.

2.2. Korábbi gyakorisági vizsgálataimhoz képest (és tudomásom szerint másokéihoz képest is) újdonságnak számít a **lemmastatisztika**. A lemmatizációt, azaz „címszósítást” egyrészt annak érdekében hajtottam végre, hogy statisztikámat közelítsem a szótári alapúakhoz, összehasonlíthatóbbá tegyem velük. Másrészt kíváncsi voltam arra, hogy bizonyos szokványos képzőjű elemeknek az átértékelése, önállóságuk megszüntetése milyen mértékben csökkenti az állomány nagyságát és a belső keletkezésűek főkategóriájának részesedését, továbbá hogyan módosítja a főkategóriák közötti és az azokon belüli arányokat.

A lemmatizációt úgy végeztem el, hogy megszüntettem a ható, a műveltető és a szenvedő igéknek (az utóbbiakhoz számítva az *elütődik* típusúakat is), valamint az igeneveknek az önállóságát: besoroltam őket alapigéjükhez, természetesen előfordulási számukkal együtt. Ugyanakkor – tekintettel a szófajváltás tényére, illetőleg annak az igenevekkel kapcsolatban megfigyelhetőnél nagyobb mértékére – nem terjesztettem ki a műveletet sem az *-ás/-és* képzős (vagy más deverbális) főnevekre, sem az *-ú/-ű* képzős mellénevekre. Önálló egységekként tartottam meg az igékötös igéket is.

Az elmondottak szemléltetésére elegendőnek tartok egyetlen példát. Az *üt* ige a szóstatisztikában 5 adattal szerepel. Ugyanez az ige ugyanott még a következő képzett, illetve összetett szavakban van jelen (betűrendben, zárójelben az előfordulási számmal): *beütés* (1), *elüt* (2), *elütés* (3), *elüthet* (5), *elütmi* (5), *elütődik* (1), *elütött* (igenév) (1), *elüttetődik* (1), *járóütés* (1), *kiütés* (1), *ütés* (16), *üthet* (2), *ütő* (igenév) (1), *ütőhatalom* (8). Ez összesen 15 szó, 53 adattal. – A lemmatizáció során az *elüt* és az *ütő* beolvadt az *ütbe*; az *elüthet*, *elütmi*, *elütődik*, *elütött*, *elüttetődik* szavak pedig az *elüt* képviselőivé váltak. Az *üt* lemma előfordulási száma ezzel 8, az *elüté* pedig 15 lett. A többi szó változatlan önállósággal és adatszámmal került át a lemmastatisztikába. Így a példabeli lemmaszám a szóállományhoz képest 15-ről 8-ra redukálódott, az előfordulási szám viszont természetesen változatlanul 53 maradt.

Az **1. lemmaeredet-statisztikai táblázat**ból kitűnik, hogy a teljes állomány a lemmatizációval a szavakéhoz viszonyítva 796 elemről 711-re, azaz mintegy 11%-kal csökkent. Az átlagos előfordulási szám 4,77.

A művelet a főkategóriák arányai között nem okozott látványos átrendeződést. A várakozásnak megfelelően némi veszteség érte a belső keletkezéseket, és ezzel a másik három főkategória nyert. Az állományt tekintve azonban egyiknek az aránynövekedése sem éri el a 2%-ot, az előfordulásokra nézve is csak az ősi örökség nyeresége 3%-nyi.

A Sách 1. lemmaeredet-statisztikai táblázata

Eredetkategória	Lemma	%	Lemma-előfordulás	%
uráli	11	1,54	64	1,89
finnugor	19	2,67	190	5,60
ugor	7	0,98	27	0,80
ősi, tisztázatlan rétegből	5	0,70	49	1,45
ősi örökség magyar képzővel	18	2,53	126	3,72
ősi összesen	60	8,43	456	13,45
származék	244	34,41	751	22,15
fiktív tő származéka	12	1,69	17	0,50
elvonás	1	0,14	1	0,03
szilárdulás	99	13,90	367	10,82
szófajváltás	20	2,81	574	16,93
jelentésselkülönülés	1	0,14	1	0,03
szóhasadás	6	0,84	61	1,80
összetétel	189	26,54	624	18,40
szóösszevonás	1	0,14	6	0,18
onomatopoetikus	1	0,14	2	0,06
belső, de tisztázatlan	5	0,70	40	1,18
belső összesen	579	81,46	2444	72,07
iráni	1	0,14	1	0,03
török	6	0,84	15	0,44
német	4	0,56	29	0,86
latin	6	0,84	12	0,35
olasz	1	0,14	6	0,18
szláv	15	2,11	172	5,07
vándor	2	0,28	3	0,09
jövevény összesen	35	4,92	238	7,02
bizonytalan	18	2,53	117	3,45
vitatott	6	0,84	83	2,45
ismeretlen	13	1,83	53	1,56
tisztázatlan összesen	37	5,20	253	7,46
Összesen	711	100,00	3391	100,00

Fontosabb ennél a belső keletkezési elemek körében lezajlott átrendeződés. A származékok vezető szerepe megmaradt, az előzetes várakozást igazolva mégis ez a kategória a lemmatizáció vesztese: az állományban 15%-kal, az előfordulások között 7%-kal csökkent a részesedése. A visszaesés kü-

lönböző mértékét az magyarázza, hogy a lemmatizációval sok származék veszítette el önállóságát, de ezeknek általában csekély az előfordulási számuk. – Az összetett lemmák viszont a származékok rovására történt gyarapodásról tanúskodnak, persze különösen az állományt tekintve.

A lemmatizáció a szóstatistikához képest új kategóriának, az onomatopoeitikusnak a felbukkanásával is járt. Képviselője a *kaphat* származék helyébe lépő *kap* ige lett.

Az egyes lemmák gyakorisági listájának élmezőnye nemigen különbözik a szavakétól. Ez nem is meglepő, hiszen zömmel olyan elemek alkotják, amelyek nem kaphatnak erősítést a lemmatizációtól; a művelet kedvezményezettjei (a szokványos képzések alapigéi) viszont a szóstatistikából akkora hátrányt örökölték, amekkorát adatszámuk gyarapodásával sem képesek behozni. A legfel-tűnőbb a *lesz* ige előretörése: szóként csak 11 adata van, lemmaként viszont a beolvaszott származékoknak (*lehet* 13, *lévő* 12, *lenni* 2, *lévén* 1) köszönhetően 39, és ezzel a lemmalista tizenkettedik helyezette.

A 2. lemmaeredet-statisztikai táblázat ugyanolyan elvek szerint különbözik a párjától, mint a két szóstatistika.

A 2. lemmaeredet-statisztikai táblázata

Eredetkategoría	Lemma	%	Lemma-előfordulás	%
uráli	11	1,55	64	1,89
finnugor	31	4,36	337	9,94
ugor	9	1,27	33	0,97
ősi örökség magyar képzővel	25	3,52	198	5,84
ősi összesen	76	10,68	632	18,64
származék	246	34,60	755	22,26
fiktív tő származéka	14	1,97	23	0,68
elvonás	1	0,14	1	0,03
szilárdulás	102	14,35	405	11,94
szófajváltás	21	2,95	574	16,93
jelentésselkülönítés	1	0,14	1	0,03
szóhasadás	6	0,84	61	1,80
összetétel	190	26,72	625	18,43
szóösszevonás	1	0,14	6	0,18
onomatopoeitikus	1	0,14	2	0,06
belső, de tisztázatlan	1	0,14	1	0,03
belső összesen	584	82,14	2454	72,37
iráni	2	0,28	9	0,27
török	8	1,13	21	0,80
német	4	0,56	29	0,86
latin	6	0,84	12	0,35
olasz	1	0,14	6	0,18
szláv	15	2,11	172	5,07
vándor	2	0,28	3	0,09
jövevény összesen	38	5,34	252	7,43
ismeretlen	13	1,83	53	1,56
Összesen	711	100,00	3391	100,00

Az átrendezéssel itt is a finnugor elemek nyerték a legtöbbet, különösen az előfordulásokat tekintve. Az ősi örökség részesedése az állományra nézve is átlépte a 10%-os küszöböt, az előfordulások között pedig megközelítette a 20%-ot.

2.3. A tőstatisztika számára az összetett szavakat elemeikre bontottam, természetesen az igekötős igéket is beleértve. Az elhomályosult összetételekkel, amilyen például az *ember*, ugyanígy tettem. – A származékszavak tövéről itt nemcsak azokat a képzőket vágtam le, amelyeket a lemmatizációkor, hanem az összeset, függetlenül attól, hogy tölexémához jutok-e, vagy fiktív tőhöz. Szintén lementettem a töről a megszilárdult alakulatok rag- vagy jelmorfémáját. – Természetesen és gyakran megesezt az is, hogy az összetétel felbontására és a morfémák lementszésére együtt kellett sort kerítenem. A lemmatizációnál említett *járóútés* esetében például így jutottam el a *jár* és az *út* tövekhez.

Bizonyos etimológiai kategóriáknak a képviselői a közvetlen előzményükhöz képest nem tartalmazznak morfémátöbbletet. Ilyen a szófajváltás, a jelentéselkülönülés és a szóhasadás. – Ezeknek az eredménye tőszó, illetőleg toldalékolhatatlan szófaj is lehet. Ilyen például a *húz* → *hoz* szóhasadás, valamint a *de* (partikula) → *de* (kötőszó) szófajváltás. Az efféle esetekben nem „tő : szó”, hanem „etimológiai előzmény : etimológiai következmény” típusú szembenállással van dolgunk. A statisztika szempontjából azonban ugyanúgy a visszavezetés mozzanata a lényeges, mint a tényleges tövek esetében, ezért a tőstatisztikában az ilyen elemeknek is helyet adok, a tövekével megegyező státussal és az egyszerűség kedvéért rájuk ugyancsak a „tő” megnevezést használva. – Ha az imént felsorolt szóalkotásmódok produktuma és egyben előzménye több morfémából áll, akkor a visszavezetés végül valóban tövet eredményez. Ilyen például a sakk-könyvbeli egyetlen jelentéselkülönülésnek az esete: a szöveg *hajt* ' <növény> hajtást növeszt' igéje a 'mozdít, üz stb.' jelentésű *hajtól* különült el, annak pedig fiktív *haj-* töve van.

Az elvonás eredményére is érvényes, hogy hozzá képest a közvetlen etimológiai előzmény nem tő; sőt a produktumra az előzményhez viszonyítva éppen a morfémahiány a jellemző. A Sách szókészletének egyetlen olyan eleme, amely elvonással keletkezett, a *harc*: ez közvetlenül a *harcol* igére, az pedig a fiktív *hor-* töre vezethető vissza.

Az etimológiai visszavezetésben fontosnak tartottam annak az elvnek a betartását, hogy a magyar nyelv határain kívülre csak egy lépést szabad tenni. A *közép* szó például ugor kori örökség. Hiába ismerjük fel, hogy képző van rajta, és hiába tudjuk, hogy a *köz* finnugor örökség, a képzés itt „alapnyelvi belügynek” számít, nem a magyarban történt, ezért a *közép* esetében a visszavezetésben csak az ugor korig mehetünk, a tőstatisztika számára a *közép* minősítése csupán „ugor” lehet.

Az **1. tőeredet-statisztikai táblázatból** kiderül, hogy a Sáchnak az etimológiai visszavezetéssel nyert tőállományát 289 elem alkotja, 4112 előfordulással. Egy-egy töre tehát átlagban 14,23 előfordulás jut.

A tőstatisztika képe merőben más, mint a szó- és lemmastatisztikáé. Az etimológiai visszavezetés szinte teljesen felszámolta a belső keletkezésű elemek kategóriáit; ezzel a főkategória helyébe az onomatopoeitikus elemeké lépett, csekély részesedésekkel. A többi főkategória viszont természetesen igen sokat nyert. Az ősi tövek száma megközelíti az állomány felét, illetőleg az előfordulások háromnegyedét. Nem meglepő, hogy igencsak megnőtt a tisztázatlan eredetűek részesedése is: az állományban 30% feletti, az előfordulások között 18%. A jövevények állományi részesedése megközelíti a 20%-ot, az előfordulási pedig a 10%-ot. Mindez arra is vall, hogy az utóbbi két főkategóriában a tövek átlagára nem jellemző a sok ismétlődés: a tisztázatlan eredetűek előfordulási hányadosa 8,34, a jövevényeké 6,93, mindkét érték jóval elmarad az átlagtól. Ezzel szemben az ősi tövek az átlagnál sokkal hajlamosabbak az ismétlődésre: előfordulási hányadosuk 21,33.

Az ősi örökség kategóriái közül a „magyar képzővel” típus a tőstatisztikában természetesen nincs meg. A vizsgálat eddigi menetében az alapnyelvi rétegek közül mindvégig a finnugoré volt a ve-

zető szerep. Az állományt tekintve ez a tőstatisztikában is így van. Az előfordulási statisztikában viszont az uráli tövek ugranak az első helyre az átlagot szinte elképesztő mértékben felülmúló ismétlődési hányadosuknak (37,71!) köszönhetően.

A Sách 1. tőeredet-statisztikai táblázata

Eredetkategória	Tő	%	Tőelőfordulás	%
uráli	31	10,73	1169	28,43
finnugor	59	20,42	819	19,92
ugor	34	11,76	345	8,39
ősi, tisztázatlan rétegből	15	5,19	632	15,37
ősi összesen	139	48,10	2965	72,11
onomatopoetikus	6	2,08	25	0,61
iráni	2	0,69	7	0,17
török	13	4,50	54	1,31
német	6	2,08	36	0,88
latin	7	2,42	14	0,34
olasz	4	1,38	38	0,92
szláv	22	7,61	236	5,74
vándor	2	0,69	3	0,07
jövevény összesen	56	19,38	388	9,53
bizonytalan	39	13,49	484	11,77
vitatott	9	3,11	32	0,78
ismeretlen	40	13,84	218	5,30
tisztázatlan összesen	88	30,45	734	17,85
Összesen	289	100,00	4112	100,00

A jövevények főkategóriáján belül a szlávok vezető szerepe megmaradt. Az etimológiai visszavezetés leginkább talán a török elemeknek kedvezett, de az előfordulásokat tekintve figyelmet érdemel például az olasz elemek megnövekedett részesedése is. (Az utóbbi elsősorban a *talpas* származékból kapott *talpnak* köszönhető.)

A tisztázatlan eredetűek között a vitatottak részesedése a legkisebb. A bizonytalan és az ismeretlen eredetű tövek száma majdnem megegyezik, de a bizonytalanokra sokkal jellemzőbb az ismétlődés: előfordulási hányadosuk (12,41) megközelíti az átlagot, míg az ismeretlen eredetűeké (5,45) messze elmarad attól.

A tövek gyakorisági listáját az uráli korból örökölt *a* névmás vezeti, 357 előfordulással. Második a szintén ősi, de tisztázatlan alapnyelvi rétegből való *az*, 279 adattal (minősítéséhez vö. Horváth 2002: 275). A harmadik és a negyedik két uráli névmás: az *e* (242) és a kérdő-vonatkozó *mi* (199). Az ötödik pedig a bizonytalan eredetű (talán finnugor kori) *s* (191). A jövevények listavezetője a szláv *király* (82; adatszám a szóstatisztikához képest a *királyi*, *királyné*, *királyos* csatlakozásával nőtt).

A 2. tőeredet-statisztikai táblázat ugyanolyan elvek szerint különbözik a párjától, mint a korábban bemutatott pároknak a tagjai. Az ismeretlen eredetűek kategóriája természetesen itt is főkategóriává lép elő.

Az átrendezéssel az ősi tövek állományi részesedése 60% fölé növekedett, az előfordulások között pedig megközelítették a 85%-ot. A jövevények sokkal kisebb mértékben gyarapodtak. Az ismeretlen eredetűek tőállományi részesedése 15%-nyi lett.

A Sách 2. töredet-statisztikai táblázata

Eredetkategória	Tő	%	Tőelőfordulás	%
uráli	37	12,80	1192	28,99
finnugor	90	31,14	1844	44,84
ugor	48	16,61	409	9,95
ősi összesen	175	60,55	3445	83,78
onomatopoetikus	7	2,42	26	0,63
őspermi	1	0,35	6	0,15
iráni	3	1,04	15	0,36
török	18	6,23	68	1,65
német	7	2,42	38	0,92
latin	7	2,42	14	0,34
olasz	4	1,38	38	0,92
szláv	22	7,61	236	5,74
vándor	2	0,69	3	0,07
jövevény összesen	64	22,15	418	10,17
ismeretlen	43	14,88	223	5,42
Összesen	289	100,00	4112	100,00

A bizonytalan és vitatott etimológiák felszámolása az alapnyelvi rétegek közül a finnugornak kedvezett a legjobban: a tőállományban 30% feletti lett a részesedése, az előfordulásokat tekintve pedig 45%-osra nőtt képviselőtelével újra messze megelőzte az urálit.

3. A Sách etimológiai statisztikájának elkészülése természetesen felvetette azt a kérdést, hogy az itt mért arányok miben és mennyire különböznek a más nyelvmlékek vizsgálatában megfigyeltektől. Részletes elemzés helyett egy-egy diakrón szó-, illetve tőstatisztikai táblázatot készítettem a fő-kategóriák arányainak összevetéséhez. (A lemmastatisztikát mellőznöm kellett, hiszen ilyen most készült először.)

A táblázatokhoz (a szemléleti-módszertani egységesség érdekében) saját korábbi vizsgálataimat (Horváth 2000, 2002, 2004) használtam fel: azokból és a mostaniból is az 1. típusú szó- és tőstatisztikát. A nyelvmlékeket időrendben helyeztem el, így a Sách középre került. A rovatok felső sora az állományt, az alsó az előfordulásokat képviseli.

Csak a Sáchot vetem össze a korábban vizsgált szövegekkel, azoknak az egymással való összehasonlításáról itt lemondok. Annak érdekében, hogy megállapíthassam, melyek azok a különbségek, amelyek statisztikailag valóban jelentősek, valószínűség-számítási (χ^2) próbának vettem alá az adatokat. Az eltéréseket a véletlennek 5%-nál kisebb esélyt hagyva tekintettem szignifikánsnak.

3.1. A diakrón szóstatisztika azt bizonyítja, hogy a belső keletkezésű elemek részesedése a Sáchban lényegesen nagyobb, mint a többi szövegben. Ez a szóállományra és a szóelőfordulásokra egyaránt érvényes.

Az ősi örökség vonatkozásában éppen fordított a helyzet: képviselőinek aránya a Sách szóállományában és szóelőfordulásai között szignifikánsan kisebb, mint a többi nyelvmlékben.

A jövevénytiszavakkal kapcsolatban csak az állományt illetően figyelhetők meg jelentős különbségek, mégpedig a sakk-könyv hátrányára a két Halotti beszéddel és a 20. század végi elbeszélésekkel szemben.

Diakrón szóstatisztika

Szöveg(ek)	Ősi	%	Belső	%	Jövevény	%	Tisztázatlan	%
Halotti beszéd	33	25,58	66	51,16	12	9,30	18	13,95
	70	26,12	132	49,25	24	8,96	42	15,67
Ómagyar	29	30,85	57	60,64	3	3,19	5	5,32
Mária-siralom	42	32,06	77	58,78	7	5,34	5	3,82
Sách	52	6,53	681	85,55	33	4,15	30	3,77
	354	10,44	2572	75,85	236	6,96	229	6,75
Kosztolányi:	36	19,35	116	62,37	18	9,68	16	8,60
Halotti beszéd	62	22,92	175	60,76	18	6,25	29	10,07
20. sz. végi	74	10,01	562	76,05	60	8,12	43	5,82
elbeszélések	217	15,27	1027	72,27	79	5,56	98	6,90

A tisztázatlan eredetű szavaknak és előfordulásainak a részesedése a SÁCHban lényegesen kisebb, mint a két Halotti beszédben; a többi eltérés itt nem szignifikáns.

3.2. A diakrón tőstatisztika másféle, bonyolultabb differenciákra derít fényt.

Diakrón tőstatisztika

Szöveg(ek)	Ősi	%	Belső	%	Jövevény	%	Tisztázatlan	%
Halotti beszéd	62	56,88	2	1,83	15	13,76	30	27,52
	190	61,49	4	1,29	27	8,74	88	28,48
Ómagyar	56	70,89	4	5,06	3	3,80	16	20,25
Mária-siralom	110	74,32	9	6,08	10	6,76	19	12,84
Sách	139	48,10	6	2,08	56	19,38	88	30,45
	2965	72,11	25	0,61	388	9,44	734	17,85
Kosztolányi:	81	52,26	6	3,87	30	19,35	38	24,52
Halotti beszéd	247	68,99	8	2,23	31	8,66	72	20,11
20. sz. végi	179	39,17	34	7,44	110	24,07	134	29,32
elbeszélések	1120	62,47	66	3,68	178	9,93	429	23,93

Az alapnyelvből örökölt tőállomány a SÁCHban lényegesen kisebb arányú, mint az ÓMS.-ban, de sokkal jelentősebb, mint a 20. század végi elbeszélésekben. Az ősi tövek előfordulását illetően a sakk-könyv nemcsak az elbeszéléseket, hanem a HB.-et is szignifikánsan felülmúlja.

A belső keletkezésű tövek a SÁCHban mind onomatopoeitikusak, a többi szövegben is szinte mindig azok. Állományuk részesedését tekintve az elbeszélések előzik meg nagymértékben a SÁCHot. Az onomatopoeitikus tövek előfordulását illetően a sakk-könyv nemcsak az elbeszélésektől, hanem az ÓMS.-tól és a Kosztolányi-verstől is jelentősen elmarad. Műfajára gondolva ez nem is meglepő.

A jövevénytövek állományi és előfordulási részesedését tekintve a SÁCH nagyon emlékeztet társaira, csupán az ÓMS. ilyen állományát múlja felül tekintélyes különbséggel.

A tisztázatlan eredetű tövek állományának képviselőjére nézve sehol sincs jelentős differencia. Az ilyen tövek előfordulásában azonban a SÁCH mind a HB.-től, mind a 20. század végi elbeszélésektől jóval elmarad.

4. A SÁCH etimológiai vizsgálata természetesen azt a kérdést is felvetette, mekkora benne a **sakkterminológia** képviselője, és milyen annak az **eredetbeli összetétele**. A válaszadáshoz (szótárszerűsége miatt) a lemmastatistikát hívtam segítségül.

A szakszavak kiválasztása a teljes szókészletből nem könnyű, és aligha lehet mentes a szubjektivitástól. A részletezésről lemondva itt csak annyit jegyzek meg, hogy a terminológia szerveződését és az egyes elemeknek a hovatartozását koncentrikus körökben, a lólapra emlékeztetően képzelem el. A sakk esetében például a figurák „hivatalos” megnevezéseit a centrumba vagy ahhoz igen közel kell tennünk, míg alkalmi szinonimáikat a periférián vagy éppen a terminológia határán kívül helyezhetjük el.

Sakkterminológiai tárgyú bölcsészdoktori értekezésében Földi József (1968: 336) úgy véli, hogy a Sách körülbelül 40 fontos sakkaszt tartalmaz. Korábbi sakknyelvtörténeti előadásomban (Horváth 2010: 70) ezzel magam is egyetértettem. Hibásnak most sem tartom: annak a függvényében, hogy mit tekintünk sakkasztónak és azon belül fontosnak, ez az eredmény is lehetségesnek, reálisnak mondható. Ezúttal azonban – kevésbé szigorú szűrővel – lemmalistámból való válogatással arra jutottam, hogy a könyvnek **68 sakkterminusa** van, összesen **663 előfordulással**. Az előfordulási hányados tehát 9,75.

Azokat a lemmákat, amelyeket a sakkterminológiához soroltam, nem fogalomkörök szerint, hanem az etimológiai statisztikához illően eredetük szerinti csoportosításban listázom. Az eredetkategorikán belül a csökkenő gyakoriság sorrendjében haladok, zárójelben feltüntetve az előfordulási számot. A jelentést ott adom meg, ahol azt a poliszémia vagy a homonímia, illetőleg a terminus szokatlansága (nemegyszer csak a Sächra jellemző volta) indokolja.

A lista bemutatása előtt szeretnék még két megjegyzést tenni. – Vannak olyan lemmák, amelyek csupán egy bizonyos jelentésben tekinthetők sakkterminusnak, más(ok)ban nem. A *játék* például általános jelentésében nem az, 'játszma'-ként viszont igen. Az ilyenek esetében csak a terminust reprezentáló adatok számát vettem figyelembe. – A másik kiegészítést óvatosságból, a félreértések megelőzésére teszem. Az etimológiai minősítés mindenütt magának a lemmának az eredetét jelenti, vagyis nem arra utal, milyen hatásra vált a lemma vagy adott jelentése a sakkterminológia részévé. (A különbségre hadd említsék egy példát a Sächon kívülről: a *paraszt* szó szláv eredetű, de sakknyelvi 'gyalog' jelentése német mintát tükröz. Ha az első sakk-könyvben megvolna, statisztikámban szlávként szerepelne.)

Ősi örökség: uráli: *megy* 'lép' (13) | finnugor: *jár* 'lép' (7), *fal* 'a tábla széle' (2), *áll* 'lép vhová' (1) | ugor: *lép* ige (9) | ősi, tisztázatlan rétegből: *had* (22) | ősi örökség magyar képzővel: *fekete* (11).

Belső keletkezésű: származék: *járás* 'mozgás, menetmód; lépés' (33), *lovas* 'huszár' (31), *talpas* 'gyalog mint sakkfigura' (30), *veszt* főnév 'sakkadás; más bábnak a megtámadása' (26), *nyílas* 'futó' (23), *ütés* (16); *állás* 'alapállás', *költözés* 'sáncolás' (7-7); *királyos* (4); *elütés*, *kiugratás* 'az alapállás elhagyása; játszmakezdés' (3-3); *kikapás* 'az alapállás elhagyása' (2); *beütés*, *felkapás* 'az ellenfél alapsorának elérése', *húzás* 'lépés', *kiugrás* 'az alapállás elhagyása', *költözködik* 'sáncol', *lépés*, *menés* 'lépés', *mozdítás* 'lépés', *rablás* 'a király megfosztása a hadától', *ugrás* (1-1) | szófajváltás: *mező* 'a táblának egy kis négyzete; maga a tábla' (88), *hadakozó* 'sakkfigura' (3) | összetétel: *hadnagy* 'vezér' (21), *elűt* 'leüt' (15), *ütőhatalom* 'azoknak a mezőknek az összessége, amelyekre az adott figura üthet' (8), *királyné* 'vezér' (7), *elvesz* 'leüt' (4), *elvisz* 'leüt' (3); *elmegy* 'ellép', *kikap* 'kilép', elhagyja az alapállást; eléri az ellenfél alapsorát' (2-2); *főtiszt* 'kb. mint a mai nehéztiszt <a bástyára vonatkozva fordul elő>', *harcolójárás* 'szokásos lépés <szembeállítva a sáncolással>', *járástétel* 'lépés', *járóütés* 'ütéssel járó lépés [nem a menet közbeni ütést jelenti!]', *keresztülugrik*, *kiáll* 'kilép az alapállásból', *kimegy* 'kilép', *kivisz* '<gyalogot> bevisz az ellenfél alapsorára', *közkatona* 'gyalog', *lovasugrás* 'lőugrás', *mezőő* 'mező mint kis négyzet a táblán', *oldalmező*, *ugrójárás* 'lőugrás' (1-1).

Jövevény: török: *sereg* (1) | német: *torony* 'bástya mint sakkfigura' (24) | latin: *tábla* (4) | olasz: *szerecsen* 'sötét <mint figuraszín>' (6) | szláv: *király* (70), *tiszt* (42), *bajnok* 'figura' (16), *rend* 'sor' (11); *pázsit* 'sakktábla <mint harcmező>', *vitéz* 'tiszt' (1-1) | vándorszó: *sakk* főnév 'sakkadás' (2), *sakk-matt* főnév 'matt(adás)' (1).

Tisztázatlan eredetű: bizonytalan: *fehér* (26), *játék* 'játszma' (17), *út* (8) | ismeretlen: *ugrik* (7).

A lista kategóriáit és adatait táblázatba foglalva is bemutatom. A **terminológiai táblázat** képét a könyv teljes anyagának 1. lemmaeredet-statisztikai táblázatával összevetve érdemes értékelni.

Az összehasonlításból kiderül, hogy a sakcterminusok a teljes **lemmaállomány 9,56%**-át alkotják, az összes **lemmaelőfordulásból pedig 19,55%**-kal részesednek. Előfordulási hányadosuk (9,75) a kétszeresénél is nagyobb a teljes anyagban mérhetőnek (4,77). Nagyon is érthető, hogy egy szakkönyvben a terminusok az átlagosnál sokkal gyakrabban ismétlődnek, de hasznosnak érzem, hogy ezt az egyébként is feltételezhető jelenséget tényleges számaránnyal sikerült bizonyítanom.

A könyvbeli sakkterminológia kétharmadát belső keletkezésű elemek alkotják. Részesedésük azonban szignifikánsan kisebb a teljes lemmakészletre érvényesnél. A terminológiában az összetett lemmák száma majdnem eléri a származékokét; mindkét kategória 30%-nál valamivel nagyobb arányban részesedik az állományból.

A sakkterminológiában rendkívül megnő a jövevények képviselete: 18%-os részesedésükkel a főkategóriák sorában a második helyre lépnek elő. Ez az arány majdnem a négyszerese a könyv teljes lemmaállományában megfigyeltek. Itt is a szláv jövevények dominálnak.

A Sách sakkterminológiájának lemmaeredet-statisztikai táblázata

Eredetkategória	Lemma	%	Lemma-előfordulás	%
uráli	1	1,47	13	1,96
finnugor	3	4,41	10	1,51
ugor	1	1,47	9	1,36
ősi, tisztázatlan rétegből	1	1,47	22	3,32
ősi örökség magyar képzővel	1	1,47	11	1,66
ősi összesen	7	10,29	65	9,80
származék	22	32,35	195	29,41
szófajváltás	2	2,94	91	13,73
összetétel	21	30,88	75	11,31
belső összesen	45	66,18	361	54,45
török	1	1,47	1	0,15
német	1	1,47	24	3,62
latin	1	1,47	4	0,60
olasz	1	1,47	6	0,90
szláv	6	8,82	141	21,27
vándor	2	2,94	3	0,45
jövevény összesen	12	17,65	179	27,00
bizonytalan	3	4,41	51	7,69
ismeretlen	1	1,47	7	1,06
tisztázatlan összesen	4	5,88	58	8,75
Összesen	68	100,00	663	100,00

A terminusok körében az ősi örökségnek és a tisztázatlan eredetű elemeknek a reprezentáltsága nem különbözik lényegesen a lemmaállomány egészére érvényestől.

Ami a terminusok előfordulását illeti, a belső keletkezésűek aránya még mindig 50% feletti, de szignifikánsan kevesebb van belőlük, mint a könyv összes lemmamegjelenése között. Az össze-

tett terminusok jóval ritkábban ismétlődnek a származékoknál, ezért itt a részesedésük messze elmarad amazokétól.

Az előfordulásokat tekintve a jövevények adatszámja majdnem eléri a belső keletkezésűekének a felét. Ez természetesen szignifikánsan megnőtt, 27%-os képviseletet jelent. A szláv elemek előfordulási aránya önmagában is túllépi a 20%-ot.

A terminusok körét vizsgálva az ősi örökség reprezentánsai sokkal ritkábban bukkannak fel az összes lemmeelőfordulásra nézve tapasztalt aránynál, részesedésük itt 10% alatti. A tisztázatlan eredetűek képviselete viszont itt sem különbözik jelentősen a tágabb körű felmérésben látottól.

Mint már szó esett róla, a sakkterminusok gyakorisági listáját a szófajváltással létrejött *mező* vezeti, 88 előfordulással. Két szláv jövevény követi: a *király* (70) és a *tiszt* (42). Három származék áll utánuk a sorban: *járás* (33), *lovás* (31), *talpas* (30). Az ősi örökség képviselői közül a tisztázatlan rétegből való *had* (22) a leggyakoribb, a tisztázatlan eredetűek főkategóriájából pedig a bizonytalan származatúsú *fehér* (26). Rajtuk kívül még négy terminusnak van 20-nál több adata: *veszt* főnév (származék, 26), *torony* (német, 24), *nyilas* (származék, 23), *hadnagy* (összetétel, 21). Érdeemes megjegyezni, de nem meglepő, hogy minden sakkfigurának van 20-nál többször előforduló megnevezése.

5. Összességében az etimológiai statisztika és a diakrón egybevetés a Sächot belső keletkezésű szavakban gazdag könyvként mutatja be nekünk. Az ősi örökség képviselői a szóanyagában ritkábbak a korábbi vizsgálatokban megszokottnál, a tövek állományát és előfordulásait tekintve viszont tiszteletet parancsol ennek a főkategóriának a dominanciája. Érdeemes ezt a tényt mindazoknak a figyelmébe ajánlani, akik a magyar nyelv uráli-finnugor-ugor származásában azért kétkednek, mert keveslik az ilyen eredetű szavak számát a teljes magyar szókészlethez viszonyítva.

A jövevények képviselete a Säch egészét tekintve nemigen tér el a máshol megszokottól. A sakkterminológia viszont feltűnően nagy számban és gyakran él velük. Ennek ellenére az uralkodó főkategória a terminológiában is a belső keletkezésűeké.

Ilyen is lehetett tehát Kazinczy születésének idején egy ismeretterjesztő szakkönyv szó- és tökélszétének eredetbeli összetétele és benne az elemek előfordulási aránya. Hogy az itt látható kép mennyire jellemző a nyelvtörténeti periódusra és a műfajra, annak a megállapításához természetesen szükség volna a Säch néhány hasonló műfajú, de más témájú kortársának etimológiai statisztikájára is.

A Kazinczy-évfordulótól függetlenül is időszerűnek és fontosnak tartottam, hogy összeállít-sam egy teljes könyv etimológiai statisztikáját, remélve, hogy megfelelő összehasonlítási alapul szolgálhat más hasonló elemzésekhez. Ezzel együtt abban szintén bízom, hogy a vizsgálat hozzájárult a magyar sakknyelv történetének alaposabb, újabb szempontokra kiterjedő megismeréséhez.

SHAKIRODALOM

- Benkő Loránd 1962. Adatok a magyar szókincs szerkezetének változásához. *Nyelvtudományi Közlemények* 64: 116–36.
- Csiszár Gábor 2002. *Kosztolányi első verseskötetének etimológiai vizsgálata*. Szakdolgozat. Kézirat. ELTE, Budapest.
- EWUng. = Benkő, Loránd (Hrsg.) 1993–1997. *Etimológisches Wörterbuch des Ungarischen I–III*. Akadémiai Kiadó, Budapest.
- Földi József 1968. *A magyar sakkszókincs eredete és kialakulása*. Bölcsészdoktori disszertáció. Kézirat. ELTE, Budapest.
- Hajdú Mihály (szerk.) 1997. *Hét bibliafordítás összehasonlító etimológiai vizsgálata*. Károli Gáspár Református Egyetem, Budapest.
- 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, február 26.

Horváth László 2010. A magyar sakknyelv történetéből. *Magyar Nyelv* 106: 59–78.

Tolnai Vilmos 1924. Halhatatlan magyar nyelv. *Magyar Nyelv* 20: 50–9.

Horváth László
MTA
Nyelvtudományi Intézet

SUMMARY

Horváth, László

An etymological statistics of the earliest Hungarian chess manual

After similar analyses of twentieth-century short stories and Old Hungarian minor documents, the author has now compiled an etymological statistics of a whole book, the earliest Hungarian chess manual, published in 1758. He not only explores the directly occurring word forms in terms of their categories of origin; he establishes those proportions after lemmatisation, too. Furthermore, he also produces an etymological statistics of the stems that the words of the text can be traced back to. He compares the results, by compiling a diachronic statistics, with those obtained in his earlier studies. He makes a separate statistics of the chess terms occurring in the book and compares it to the statistics of the full stock of lemmas in the same book. – The aim of the analysis was to find out the etymological make-up of a non-fiction text coming from the mid-eighteenth century, right before the beginning of the language reform. The study will also have contributed to a more detailed and up-to-date description of the history of Hungarian chess terminology.

Keywords: etymological statistics, word statistics, lemma statistics, stem statistics, diachronic statistics, chess terminology