

Magyar nyelvű kódexek digitális alapú feldolgozása

É. Kiss Katalin

(MTA Nyelvtudományi Intézet és PPKE)

1. Bevezetés

Írásomban röviden bemutatom az MTA Nyelvtudományi Intézetében 2009-2013 között készült ómagyar digitális adatbázist, majd azt illusztrálom két példával, hogy az adatbázis hogyan segítheti a nyelvtörténeti és azon belül a mondattörténeti kutatásokat; milyen típusú mondattani kérdések megválaszolását teszi lehetővé.

2. Az ómagyar korpusz

A bemutatandó nyelvtörténeti adatbázis a Magyar generatív történeti mondattan elnevezésű OTKA-projektum keretében készült Simon Eszter irányításával (l. Simon és Sass 2012). Bárki számára hozzáférhető az omagyarkorpusz.nytud.hu linken. Az összes fennmaradt ómagyar kori szövegelemet magában foglalja, és tartalmaz néhány kora-középmagyar kori szövegelemet is. Megtalálható benne mind a 47 magyar nyelvű kódex anyaga, továbbá 24 rövidebb szövegelem és 244 misszilis, azaz elküldött levél. A korpuszban található szövegszavak száma meghaladja a 2,2 milliót.

A szövegek digitalizálását a mérvadó szövegkiadások (általában a legutóbbi kiadás) alapján végeztük. A digitalizálás első lépését a kiválasztott szövegkiadás lapjainak képként való beolvasása, szkennelése jelentette. A szkennelt anyagot OCR (optikai karakterfelismerő) program segítségével tettük kereshetővé. A program kimenetét ellenőriztük és kézzel javítottuk. A szövegeket a mérvadó kiadások betűinek megfelelő Unicode karakterekkel tároltuk. A Huszita Biblia cs-nak hangzó L betűjét a Volf György-féle kiadásokat követve č-vel helyettesítettük.

A szövegek egy részét automatikus morfológiai elemzésnek vetettük alá. Ehhez normalizálnunk kellett az elemzendő szövegeket, azaz, az eredeti betűhű szóalakokat mai magyar helyesírású szavakká kellett alakítanunk. A morfológiai elemzés a szövegszavak szótövét és morfológiai szerkezetét adja meg.

Az elemzés különféle szintjeit a Jókai-kódex első három során mutatjuk be:

oldal	sor	eredeti	normalizált	szótó	elemzés
001	01	Mÿ	mi	mi	Pro.Nom Gen
001	01	vronknac	urunknak	úr	N.PxP1.Dat Gen
001	01	iesus	Jézus	Jézus	N:P.Nom
001	01	cristusnac	Krisztusnak	Krisztus	N:P.Dat Gen
001	01	gyczeretyre	dicséretire	dicséret	N.PxS3=i.Sub
001	02	es	és	és	C
001	02	gyczewsegere	dicsőségére	dicsőség	N.PxS3.Sub
001	02	es	és	és	C
001	02	my	mi	mi	Pro.Nom Gen
001	02	atyancnak	atyánknak	atya	N.PxP1.Dat Gen
001	03	bodog	boldog	boldog	Adj
001	03	ferencznek	Ferencnek	Ferenc	N:P.Dat Gen

Az adatbázishoz az alábbi felhasználóbarát keresőfelület tartozik:

Magyar Generatív Történeti Szintaxis

hun | eng

Kezdőlap » Keresés

kezdőlap

projektum

korpusz

keresés

szövegemlékek

kódexek

egyéb szövegek

kapcsolat

Keresés

Adja meg a keresett szó tulajdonságait ...

Betűhű (b)	:	<input type="text"/>	:		Formátum:	konkordancia
Normalizált (n)	:	<input type="text"/>	:		Megjelenítés:	minden
Szótó (e)	:	<input type="text"/>	:		Kontextus ^{BETA} :	csak a találati szakasz
Elemzés (e)	:	<input type="text"/>	:		Nyelvemlék:	mind
Értelmezés	:	<input type="text"/>	:		Szerk. mód:	új lekérdezés
Igekötő	:	<input type="text"/>	:		Kihagyás:	Min: 0
Megjegyzés	:	<input type="text"/>	:		Max: 5	<input type="button" value="OK"/>
Azonosító	:	<input type="text"/>	:			

... vagy szerkessze a lekérdezést közvetlenül. (Guide)

Megjegyzés:

v0.3.B.4 – 2013.08.29. – S. B. | Emdros – súgó

MTA Nyelvtudományi Intézet
Az ómagyar szövegek helyes megjelenítéséhez szükséges a Junicode font telepítése.

A kereséskor választhatunk, hogy mely reprezentációs szinten adjuk meg a keresendő szót: betűhíven, normalizálva, azaz mai magyar helyesírással vagy szótóként. E lehetőségeken belül is választhatunk, hogy csak a keresett elemmel teljes egészében megegyező találatokat kérjük, vagy azokat a találatokat is elfogadjuk, melyeknek vagy az eleje vagy a vége azonos a keresett elemmel. A morfológiailag elemzett nyelvemlékekben kereshetünk nyelvtani kategóriákra, kategóriakomplexumokra is. Az Értelmezés ablakban már kihalt szavak mai jelentését adhatjuk meg; például, ha arra vagyunk kíváncsiak, volt-e korábban más szavunk a *csak*-ra, rátalálhatunk a *heon*-ra.

A keresést a Szerkesztésmód ablakban lehulló menü segítségével lépésről lépésre újabb szavakkal lehet bővíteni, s ezáltal nemcsak szavakra, hanem kifejezésekre, sőt, a Kihagyás opció segítségével nem szomszédos szavakból álló szókapcsolatokra is ki lehet terjeszteni. Az igekötős igék esetében ezt egyszerűbben is megtehetjük: ha az Igekötő ablakban megadunk egy igekötőt, megkapjuk a hozzá tartozó igéket is. A kereső felület jobb oldalán kiválaszthatjuk, hogy találatként szövegtöredékeket (konkordanciát) vagy gyakorisági adatokat kérünk. Meg kell adnunk a megjelenítés szintjét vagy szintjeit is. A megjelenítés szintjének nem kell egybeesnie a keresési szinttel. Arról is döntenünk kell, hogy mely nyelvemlékekben keresünk. Normalizált szavakra csak a normalizált, n-nel jelölt nyelvemlékekben, elemzésekre csak az elemzett, e-vel jelölt nyelvemlékekben kereshetünk. Választhatunk, hogy a találatokat mekkora kontextussal együtt kérjük megjeleníteni. A keresési opciók kiválasztása után mind a keresésre, mind a megjelenítésre vonatkozó választásainkat OK-val kell jóváhagynunk. A keresést a Mehet ablakban indíthatjuk el.

A keresés eredményeként könnyen olvasható vízszintes szóláncokat kapunk. Ha például a betűhű *fföld* vagy a normalizált *föld* szónak a Könyvecskében található előfordulásait keressük, és a megadott opciók közül azt választjuk, mely megengedi, hogy a találatoknak csak az eleje legyen *fföld* vagy *föld*, továbbá, ha a találatokat minden reprezentációs szintjét látni akarjuk, és csak a találati szakaszt (azaz, az adott tagmondatot) kérjük, akkor a kereső így mutatja az adatokat:

[1] Konyvecse - 6r - 1/92452

mert	ők	bírják	az	fföldeth:
mert	ők	bírják	az	földet.
mert	ők	bír	az	föld
C	N Pro.P3	V.P3.Def	Det	N.Acc

[2] Konyvecse - 6r - 1/92597

Myth	foglallja	hiába	az	fföldeth
mit	foglalja	hiába	az	földet?
mi	foglal	hiába	az	föld
N Pro Int.Acc	V.S3.Def	Adv	Det	N.Acc

[3] Konyvecse - 12r - 1/94384

kýnek	zerelmetől	yth	ez	fföldön	kýnokkal	sem	zakasztathatának	el:
kinek	szerelmétől	itt	ez	földön	kínokkal	sem	szakasztathatának	el.
akinek;cmt:DIFFANA	szerelme	itt	ez	föld	kín	sem	szakaszt	el
aki	N.Abl	Adv Pro	Det Pro	N.Sup	N.Pl.Ins	Adv	V.Fact.Mod.Ipf.P3	VPfx
N Pro Rel.Dat								

Tervezzük a korpusz továbbfejlesztését. Idővel valamennyi ómagyar szöveget szeretnénk normalizálni és morfológiailag elemezni, és az ómagyar anyagot egyre több középmagyar anyaggal szeretnénk kibővíteni. Az észlelt hibákat folyamatosan javítjuk; a korpusz honlapján megadott címen várjuk a felhasználók észrevételeit.

3. Mire jó a kereshető korpusz?

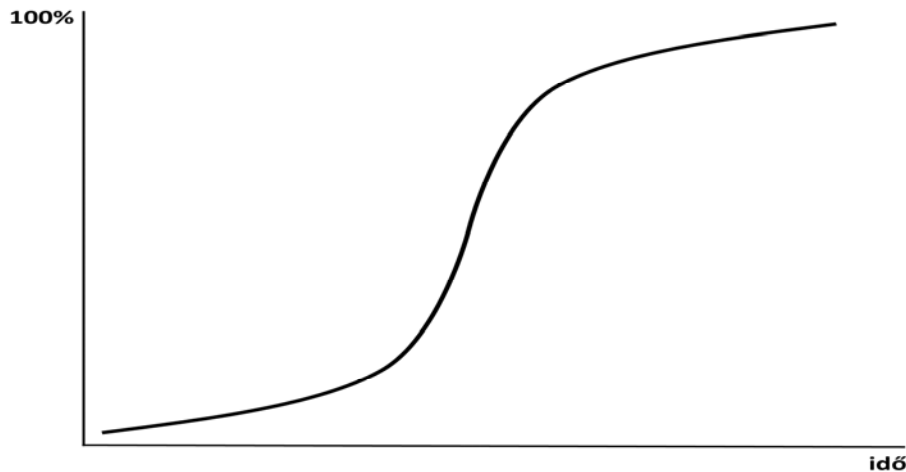
A digitális nyelvtörténeti korpusz új lehetőségeket nyitott meg a korábbi nyelvállapotok és a nyelvi változások rekonstruálásában. A Magyar generatív történeti mondatlan projektum keretében végzett kutatásaink során már élünk e lehetőségekkel – l. a *Magyar történeti mondatlan*, Akadémiai Kiadó (2014) és a *The Evolution of Functional Left Peripheries in Hungarian Syntax*, Oxford University Press (2014) tanulmányköteteket (É. Kiss 2014a,b). Itt két példával illusztrálom a korpuszadatokra épülő elemzési lehetőségeket.

3.1. Mondattani rekonstrukció

A nyelvtörténeti kutatásokban sokféle adattal alátámasztott, elfogadott feltételezés, hogy a nyelvi változások során a kialakuló új jelenségek terjedése *f*-görbét ír le (l. pl. Weinreich, Labov és Herzog 1964). Azaz, az új változat térnyerése és a régi visszaszorulása lassan indul, majd hosszú idő után hirtelen felgyorsul, és viszonylag rövid idő alatt az új változat válik

uralkodóvá. A lassú kezdés után gyorsan felívelő szakaszt lassú zárószakasz követi; az új változatnak gyakran évszázadokon át sem sikerül a régit teljesen kiszorítania.

(2) A nyelvi változások f -görbéje



Az a tény, hogy a visszaszoruló régi változatok még évszázadokig kimutathatók, lehetővé teszi, hogy kifutó f -görbék visszafejtésével korábbi, dokumentálatlan nyelvállapotokról is képet kapjunk. Így valószínűsíthető például, hogy az ősmagyar mondat – a mai hantihoz és manysihoz hasonlóan, de az SVO alapszórendű ómagyartól és mai magyartól eltérően – SOV szórendű volt, és tárgya (hacsak nem topikként funkcionált – l. É. Kiss (2013)) nem viselt tárgyragot. E hipotézist azzal támaszthatjuk alá, ha megmutatjuk, hogy az ómagyarban még előforduló ragtalan tárgyú SOV mondatok arányának csökkenése egy kifutó f -görbe végső szakaszának felel meg. Tipológiai megfigyelések szerint az SOV szórend a grammatika egyéb területein is meghatározott szerkezet típusokkal jár együtt. Ha e szerkezet típusok ómagyar kori csökkenése is ellaposodó f -görbét rajzol ki, szintén az ősmagyar SOV szórend mellett szolgáltatunk érveket. A digitális adatbázis lehetővé teszi, hogy pontosan számszerűsítsük e szerkezetek előfordulási gyakoriságának változását.

3.1.1. Eltűnő ragtalan tárgyú SOV tagmondatok

Legkorábbi kódexeink, az 1370 körül keletkezett szöveget tartalmazó Jókai-kódex, valamint az 1416 körül fordított Huszita Biblia részleteit tartalmazó Bécsi, Münchener és Apocryphus-kódex még bővelkedik ragtalan tárgyú tagmondatokban. A ragtalan tárgy kivétel nélkül mindig megelőzi az igét, és ha a mondatnak hangzó alanya van, az megelőzi a tárgyat, tehát a ragtalan

tárgyú mondat típus SOV szórendű. Minden ragtalan tárgyú tagmondat igeneves alárendelő mondat. Lightfoot (1991) elmélete értelmében a nyelvi változások mindig a főmondatokban kezdődnek, ebből következően az alárendelő mondatok konzervatívabbak – nem meglepő tehát, hogy a ragtalan tárgy az alárendelő mondatokban élt legtovább. Az ugor alapnyelv feltehetőleg csak igeneves alárendelést használt (az obi-ugor nyelvekben ma is az az általános), tehát a ragtalan tárgyú SOV szórendű alárendelő mondatokban az igenévi alaptag is megőrzött régiség. Pusztai főnév ma is előfordulhat főnévi igenév inkorporált ragtalan tárgyaként, főként nyelvjárásokban, például: *Megyek szőlő kötni*. Azonban a korai ómagyarban a ragtalan tárgy határozott főnévi kifejezés, így birtokos szerkezet vagy mutató névmás is lehetett. Például:

- (1) ne fordo'l'lon m̄g [ø k̄ntosø feluènni] (MünchK 30ra)
- (2) **mynden** thehetew atya ysten (AporK 134)
- (3) Agyad meg ymmar [**bewne** zantnak] (JókK 158)
- (4) ky zent fferenczet lewlteuala [**egy haz** seprette] (JókK 97)
- (5) [ø è gondoluan] yme vrnac angala ièlenec nèki (MünchK 8va)

Ahhoz, hogy megállapíthassuk, hogy a ragtalan tárgyú mellékmondatok előfordulása csökkenő *f*-görbét ír-e le, meg kell számolnunk, hogy korai kódexeink hány ragtalan tárgyú igeneves kifejezést tartalmaznak, és ezek száma hogyan aránylik a kódexek szószámához. A ragtalan tárgyak kigyűjtése az ómagyar adatbázis keresőprogramja segítségével sem volna könnyű feladat – de szerencsére Károly Sándor ezt a munkát már 60 éve kézi erővel elvégezte (Károly 1956), a digitális adatbázisból csak a kódexek szószámát kell megtudnunk.

A ragtalan tárgyak száma és aránya első kódexeinkben:

Kódexek:	keletkezés ideje	szószám	ragtalan tárgy	szó/ragtalan tárgy
Jókai-kódex	(1370 után):	22 733	42	1: 540
Müncheni kódex	(1416 után):	69 589	78	1: 892
Apor-kódex	(1416 után):	22 118	18	1: 1 382
Bécsi kódex	(1416 után):	54 423	24	1: 2 268
Jordánszky-kódex	(1516-19):	200 185	16	1: 12 511

A ragtalan tárgyak aránya a Jókai-kódexben a legnagyobb: körülbelül minden 500. szó ragtalan tárgy. A mintegy fél évszázaddal később fordított Huszita Biblia másolatait

tartalmazó kódexekben körülbelül 1000-2000 szóra jut egy ragtalan tárgy. Egy évszázaddal később, a Jordánszky-kódexben a ragtalan tárgyak aránya már elenyésző. A csökkenés megfelel egy kifutó (fordított) *f*-görbe végső szakaszának.

Ha visszafelé követjük e fordított *f*-görbét, és rekonstruáljuk a korábbi, dokumentálatlan szakaszait, arra a következtetésre jutunk, hogy az ősmagyar korban az SOV szórendű, ragtalan tárgyat tartalmazó mondattípus volt az uralkodó. Minthogy az obi-ugor nyelvekben is az SOV szórendű, ragtalan tárgyú mondattípus az általános,¹ feltehetőleg mindhárom leánynyelv ezt a szerkezetet örökölte az ugor alapnyelvből. A tárgyrag használata és ezzel összefüggésben az SOV szórend felbomlása a magyar nyelv külön életében, az ősmagyar korban vagy az ősmagyar és az ómagyar kor fordulóján válhatott általánossá.

3.1.2. Visszaszoruló prenominális igenévi jelzői mellékmondatok

A világ nyelveinek tipológiai atlasza, a *The World Atlas of Language Structures* (Dryer és Haspelmath 2013) 96. térképlapjának tanúsága szerint a szigorúan SOV szórendű nyelvekre a prenominális jelzői mellékmondat a jellemző. E mellékmondattípusban nincs vonatkozó névmás; a mellékmondatnak a jelzett főnévi kifejezéssel koreferens eleme üres. Az ómagyarban is megvan még e mondattípus. Mint az ómagyar valamennyi archaikus, ősmagyar örökséget képviselő alárendelő mondattípusa, e mondattípus is igenevet tartalmaz, nem véges igét. Például:

(6) ot vala eg [keze meg azott] ember (MüncK 38ra)

(7) Es adom the neked es te nemednek te vtannad ez [te laktad] feldet (JordK VIIb)

Ha az ómagyar korpuszban e mondattípus aránya is a kifutó *f*-görbékre jellemző módon fokozatosan csökken, joggal következtethetünk rá, hogy az ősmagyar korban e prenominális, igeneves, vonatkozó névmás nélküli jelzői alárendelés volt az uralkodó. Minthogy a hanti és a manysi is ilyen jelzői mellékmondatokat használ, e mondattípust a közös ugor alapnyelvig is visszavetíthetjük.

Automatikus kereséssel könnyebb a jelzői mellékmondatok újabb, vonatkozó névmással kezdődő típusát megtalálni, mint az általa kiszorított archaikus, vonatkozó névmás nélküli típust. Ezért a két jelzői mondattípus arányának változását nem a vonatkozó névmás nélküli mondatok arányának csökkenésével, hanem a vonatkozó névmással bevezetett posztinominális

¹ Ez már nem teljesen igaz a hantinak és manysinak az orosz-domináns kétnyelvűek által beszélt újabb kontaktusváltozataira.

jelzői mellékmondatok arányának növekedésével mértük. Kigyűjtöttük Máté Evangéliumának három egymást követő fordításából, a *ki*, *mi* és *mely* névmásokat. Minthogy az adatbázis e szövegeknek még nem tartalmazza a morfológiai elemzését, a kérdő névmásokat kézzel töröltük közülük. Az adatok a vonatkozó névmással bevezetett jelzői mellékmondatok számának fokozatos növekedését mutatják.

A vonatkozó kötőszói szerepű ki, mi, mely(ik) névmások száma Máté Evangéliumában:

Müncheneri kódex (1416 k.): 225

Jordánszky-kódex (1516 k.): 314

Károli Biblia (1590): 330

Minthogy a növekedés zárt korpuszban történt, melyben a jelzői mellékmondatok száma lényegében változatlan, amennyivel nőtt a vonatkozó jelzői mellékmondatok száma, lényegében annyival csökkent a vonatkozó névmás nélküli preminális igeneves jelzői mellékmondatoké. A csökkenés 1416-1516 között gyorsabb volt, mint 1516-1590 között, ami arra enged következtetni, hogy az 1500-as években az *f*-görbe már lassuló, kifutó szakaszába lépett. Ha rekonstruáljuk a teret vesztő preminális igeneves jelzői mellékmondatok és a térhódító vonatkozó mellékmondatok arányának változását leképező *f*-görbe korábbi szakaszait, arra kell következtetnünk, hogy az ősmagyarban e jelzői mellékmondattípus volt az uralkodó. A preminális igeneves jelzői mellékmondatok nem tűntek el teljesen, de a korábbiakkal ellentétben csak nagyon szűk korlátok között használhatók; így a (6)-(7) alatti példákkal ellentétben nem lehet saját külön hangzó alanyuk, és igenevük nem viselhet személyragot.

3.1.3. Visszaszoruló igeneves határozói mellékmondatok

A The World Atlas of Language Structures (Dryer és Haspelmath 2013) 94. lapja szerint az SOV alapszórendű nyelvekre az igeneves határozói mellékmondatok a jellemzők. Ha az ősmagyar valóban SOV szórendű volt, azt várjuk, hogy az igeneves határozói mellékmondatok aránya az SVO alapszórendű ómagyarban *f*-görbével modellálható csökkenést mutat. A jelzői mellékmondatok esetében alkalmazott módszert követve három egymást követő Biblia-fordítás Máté Evangéliumában kikerestük és megszámláltuk a *-ván/vén* ragos határozói igenevet tartalmazó mellékmondatokat:

-ván/vén *igeneves határozói mondatok Máté Evangéliumában:*

Müncheni kódex (1416 k.): 486

Jordánszky-kódex (1516 k.): 322

Károli Biblia (1590): 286

A *-ván/vén* igeneves határozói mellékmondatok száma egyre kisebb; helyüket kötőszóval vagy vonatkozó névmással bevezetett véges határozói mellékmondatok veszik át. Például az alábbi bibliai mondat (Máté 4,18) a Müncheni kódexben és a Jordánszky-kódexben még két határozói igeneves alárendelést tartalmaz; a Károli Bibliában viszont már mindkét ige véges alakban áll:

(8) a. [**Iaruan** kedig Iesus galileabèli tenger mèllet] lata ket at'afiat Simont ki hiuatic pètern^e es Andoriaft \varnothing at't'afiat [**èrèztuen** \varnothing haloioat a· tengerbè] (MünchK 10rb)

b. [**Jarwan** kedyglen Iesus galyleabely tengher mellet], lata ket atyaffyakat, Simont, ky nevezetyk peternek , es Andrasth hew atyaffyat, [**vetthveen** haloyokat az tengherben] (JordK 364)

c. [Mikoron pedig **iárna** Iesus az Galiléánac tengere mellett], láta két atyafiakat, Simont, ki mondati Peternek, és Andrást, (Simonnac) atyafiát, [kic az ő hálóiokat akkor **vetic vala** az tengerbe] (Károli II/II 4v)

A határozói igeneves alárendelő mellékmondatok visszaszorulása is a mondat SOV alapszórendjének megváltozására enged következtetni.

3.2. Nyelvtörténeti hipotézisek ellenőrzése

3.2.1. Alternatív magyarázatok a tárgyas ragozás eredeti funkciójáról

Az ómagyar adatbázis adatai alapján a magyar nyelv dokumentált korszakairól is pontosabb képet kaphatunk. Például kimutatható, hogy a tárgyi névmások elhagyása az ómagyar korban még szórványos volt, csak az újmagyar korban vált általánossá. Ezzel cáfolható az a közkeletű vélekedés, hogy a tárgyas ragozás a névmási tárgynak az igébe való inkorporálásával jött létre, az önálló névmási tárgy hiányát eredményezve (l. pl. Kulonen 1999).

Az kétségtelen ténynek látszik, hogy a tárgyas ragozás egyes szám 3. személyű és többes számú alakjaiban megjelenő *-ja/i* morféma a 3. személyű uráli névmás leszármazottja (Honti 2009):

(9) ír-ja	kér-i
ír-j-uk	kér-j-ük
ír-já-tok	kér-i-tek
ír-já-k	kér-i-k

A *-ja/i* morféma névmási eredeztetése vezetett ahhoz a feltételezéshez, hogy a tárgyas ragozásban az igéhez simuló tárgyi névmást nem kellett újra kitenni a mondatban, tehát a tárgyas ragozás kialakulása óta általános a magyarban az ún. rejtett tárgyi névmás.

Ugyanakkor ez az elmélet nem magyarázza meg, hogy miért az 1. számú tárgyi névmás inkorporálódott az igébe többes szám 3. személyű névmási tárgy esetén is, és miért nem inkorporálódtak az 1. és 2. személyű tárgyak.

A tárgyas ragozás alternatív magyarázata Givón (1975) azon elméletén alapszik, hogy a tárgyas ragozásban megjelenő morféma topikisméltő névmásból kialakult tárgyi egyeztető toldalék. (A topikisméltő névmás a topikalizált tárgyra utal a mondat komment részében (mint az *azt* névmás a *Jánost, azt nem ismerem* példában). E szerint az elmélet szerint a tárgyas ragozás névmási eredetű eleme nem a tárgy helyett áll, hanem csupán egyezik vele, azaz, nincs direkt összefüggés a tárgyas ragozás és a hiányzó névmási tárgy között.

Givón elméletét először Marcantonio (1985) alkalmazta a magyarra és testvérnnyelveire. Marcantonio elméletét azóta többen is továbbgondoltuk és pontosítottuk (É. Kiss 2003, 2013, Dalrymple és Nikolaeva 2011). Mint Nikolaeva (2001) és Dalrymple és Nikolaeva (2011) kimutatta, a hantiban és a manysiban a tárgyas ragozás szerepe ma is egyértelműen a topikalizált tárggyal való egyeztetés, azaz, a tárgy topik szerepének kódolása. Az SOV szórendű hanti és manysi mondatban az alany kötelezően topik, a tárgy pedig vagy másodlagos topik, vagy új információt közlő fókusz. Topik voltát az jelzi, ha egyezik vele az ige, tehát tárgyas ragozásban áll. Bizonyítékunk van rá, hogy a tárgyas ragozás eredetileg a magyarban is a tárgy topik voltának kódolására szolgált. A bizonyíték a következő:

A tipológiai irodalomból közismert, hogy a kifejezések topikálisuk, topik szerepre való elérhetőségük szerint hierarchiába rendeződnek (Moravcsik 1974). A topikális hierarchia élén az egyes szám 1. személyű névmás áll, ezt a többes szám 1. személyű névmás és a 2. személyű névmások követik. Több nyelvben is megfigyelhető, hogy a másodlagos topik nem

állhat magasabban a topikális hierarchiában mint az elsődleges topik (Dalrymple és Nikolaeva 2011). Ha például a szintén SOV szórendű szamojéd nyelvekben a mindig elsődleges topikként funkcionáló alany 3. személyű és a tárgy 1. vagy 2. személyű, a tárgy nem lehet másodlagos topik, csak fókusz – azaz, az ige nem állhat tárgyas ragozásban. Ez a megkötés a magyarban is megőrződött a tárgyas ragozás korábbi topikegyeztető szerepének maradványaként, hiszen 3. személyű alany esetén 1. vagy 2. személyű tárgy mellett alanyi ragozásban áll az ige: *Ő lát engem, téged, minket, titeket.*

3.2.2. A tárgyas ragozás funkcióváltása

Okunk van tehát feltételezni, hogy a tárgyas ragozás a magyarban is topikisméltó névmás inkorporálódásával keletkezett, és eredetileg a tárgy topik szerepének kódolására szolgált. Funkciója azonban idővel megváltozott; a tárgy topik szerepe helyett a tárgy határozottságát kezdte jelölni. A funkcionális újraértelmezés magyarázata, hogy a magyarban valamikor az ősmagyar kor folyamán felbomlott a kötött SOV szórend, kialakult a topikmozgatás, minek következtében a topik funkciót már a kifejezések szerkezeti pozíciójából is le lehetett olvasni. A tárgy topik szerepét jelölő tárgyas ragozás így funkcióját veszítette; de új szerepet talált magának – azaz, az evolúciós biológiából kölcsönvett nyelvtörténeti szakszóval: exaptáció történt. A 'topikjelölőből határozottság-jelölő' funkcióváltást megkönnyítette, hogy a topik többnyire határozott főnévi kifejezés (bár határozatlan főnévi kifejezés is lehet topik, ha specifikus, azaz, egy már ismert halmaz valamely elemére utal). Az uráli és az ugor alapnyelv, sőt az ősmagyar sem használt határozott névelőt, így a tárgyas ragozás feladata a tárgy határozottságának jelölése lett. Lényegében ezt a nyelvállapotot rögzítik első nyelvemlékeink. A Halotti beszéd és könyörgés és az Ómagyar Mária-siralom alábbi mondataiban a tárgyas ragozás kódolja formailag a névelő nélküli tárgy határozottságát:

- (10) a. **odutta** vola neki **paradifumut** hazóá.
b. **anyath** ezes fyaal /
egembelu **ullyetuk**.

A tárgyas ragozás használata, disztribúciója lényegében első nyelvemlékeinktől kezdve változatlan – ugyanakkor a tárgyas ragozás a korai szövegekben csak nagyon szórványosan jár együtt a névmási tárgy hiányával. A Halotti beszéd és könyörgés hat kitett névmási tárgyat tartalmaz:

- (11) Heon tilutoa **wt** ig fa gimilcetvl.
ef veteve **wt** ez muncal vilagbele.
Ef zoboducha **wt** urdung ildetuitvl.
ef vezeffe **wt** paradifū nugulmabeli.
ef vezeffe **wt** paradifū nugulmabeli.
ioy felevl iochtotnia ilezie **wt**.

Kérdéses, hogy a Halotti beszéd és könyörgésben rejtett tárgyi névmásra találunk-e egyetlen példát is. Az alábbi mondat jöhet szóba:

- (12) Hug ef tiv latiatuc fzumtuchel. ifa ef num igg ember mulchotia ez vermut.

A *latiatuc* igének tulajdoníthatunk rejtett névmási tárgyat – bár legalábbis értelmileg az *ifa ef num igg ember mulchotia ez vermut* mondatot értjük az ige tárgyaként..

Az első kétségtelen 2. személyű rejtett tárgyi névmás az 1300 körül másolt Ómagyar Mária-siralomban, az első kétségtelen 3. személyű rejtett tárgyi névmási pedig az 1350 körül másolt Königsbergi töredékben szerepel:

- (13) Walasth vylagum tul sydou fyodumtul

- (14) vleben tart chudaltus fiot.
furiscte musia. etety ymletí

A Königsbergi töredék alábbi sorában a tárgyi névmást utólag szűrték be. Ez arra utalhat, hogy a szöveg keletkezésekor már elbizonytalanodott a nyelvhasználat a tekintetben, hogy a 3. személyű tárgyi névmást ki kell-e tenni. Talán a scriptor nyelvérzéke már megengedte a rejtett névmást, és csak az eredetivel való összeolvasáskor derült ki, hogy a másolandó eredetiben kitett névmás van.

- (14) Tudyuc latíuc ^{evt} scuz lean nac

Az elbizonytalanodás oka, hogy a 14. században már kibontakozóban van a határozottság névelővel való következetes jelölése (l. Egedi 2014), így a tárgyas ragozás ismét

funkciótlanná válik, és újabb funkcióját az egyes szám 3. személyű rejtett tárgyi névmás engedélyezésében fogja megtalálni.

3.2.3. A tárgyas ragozás másodszori funkcióváltása

A tárgyas ragozás másodszori exaptációja fokozatosan, több évszázad alatt ment végbe. Az 1370 körül keletkezett szöveget tartalmazó, 1448 körül másolt Jókai-kódexben még ritkaság az efféle, itt [*pro*]-val jelölt rejtett névmás:

(15) Es azert ewele kyuanuala zolnya: Azert nemy haladek leuen: legottan masoczor es harmaczwr hyua [*pro*]:

Hogy milyen ritka a rejtett tárgyi névmás, arra az alábbi statisztikai adatokból következtethetünk: A kódex 98 kitett *ötet* névmást tartalmaz. A többes számú *őket* 19-szer fordul elő, azaz, az egyes számú névmási tárgyak száma kb. ötszöröse a többes számúaknak. A dativusban álló 3. személyű névmások esetében az egyes számúak száma (107) két és félszerese a többes számúaknak (42). A nominativusi névmások között az egyes számúak négyszer annyian vannak (36) mint a többes számúak (9). Ha az egyes szám 3. személyű tárgyi névmások esetében számottevő volna a rejtett névmások száma, akkor azt várnánk, hogy a hangzó egyes számú –többes számú névmások aránya közelebb volna egymáshoz, mint az egyes és többes számban is mindig kiejtett dativusi névmások esetében és az egyes és többes számban egyaránt rejtve maradható nominativusi névmások esetében. Valójában az egyes számú névmásoknak a többes számúakhoz képesti aránya a tárgyi névmások esetében a legnagyobb; nem valószínű tehát, hogy sok volna a rejtett egyes számú tárgyi névmás.

Az alanyi és tárgyi névmások arányából is ugyanezre következtethetünk. A többes számú 3. személyű névmások között, ahol csak a nominativusi névmások maradhatnak rejtve, a nominativusi–accusativusi arány 9:29, azaz, 0,31. Az egyes számú névmások között, ahol elvileg a tárgyi névmások is rejtve maradhatnak, ez az arány 36:98, azaz, 0,367. Ha pontosan ugyanazt az arányt várnánk mint a többes számban, akkor 116 egyes számú tárgyi névmással kellene számolnunk, azaz, az egyes számú tárgyi névmásoknak kb. 20%-a volna nem hangzó névmás.

Máté Evangéliumának első fennmaradt magyar fordításában, a kevéssel 1416 után készült Müncheneri kódex-beli szövegben a hangzó egyes szám 3. személyű névmások száma nem sokkal kisebb, mint Luther német fordításában. Míg Luther fordítása 156 egyes szám 3. személyű névmást tartalmaz (134 *ihn*-t és 22 egyes szám accusativusi *sie*-t), a magyar

fordításban 137 *őt* névmást találunk. Az ezt követő Máté Evangélium fordításokban a hangzó *őt* névmások száma lassú csökkenést mutat.² Pontosabban Káldi György 1605-1607 között készült, 1626-ban megjelent Máté Evangéliumában kis visszarendeződést tapasztalunk a vele csaknem egykorú Károli Bibliához képest. A döntő fordulatot az egyes szám 3. személyű névmások rejtetté válásában Békés Gellért és Dalos Patrik 20. századi fordítása dokumentálja.

Müncheni kódex	(1416/1466):	137
Pesti Gábor: Novum Testamentum (1536):		124
Károli Gáspár	(1590):	108
Káldi György: Vulgata	(1626):	117
Békés Gellért – Dalos Patrik ford. (1950):		19

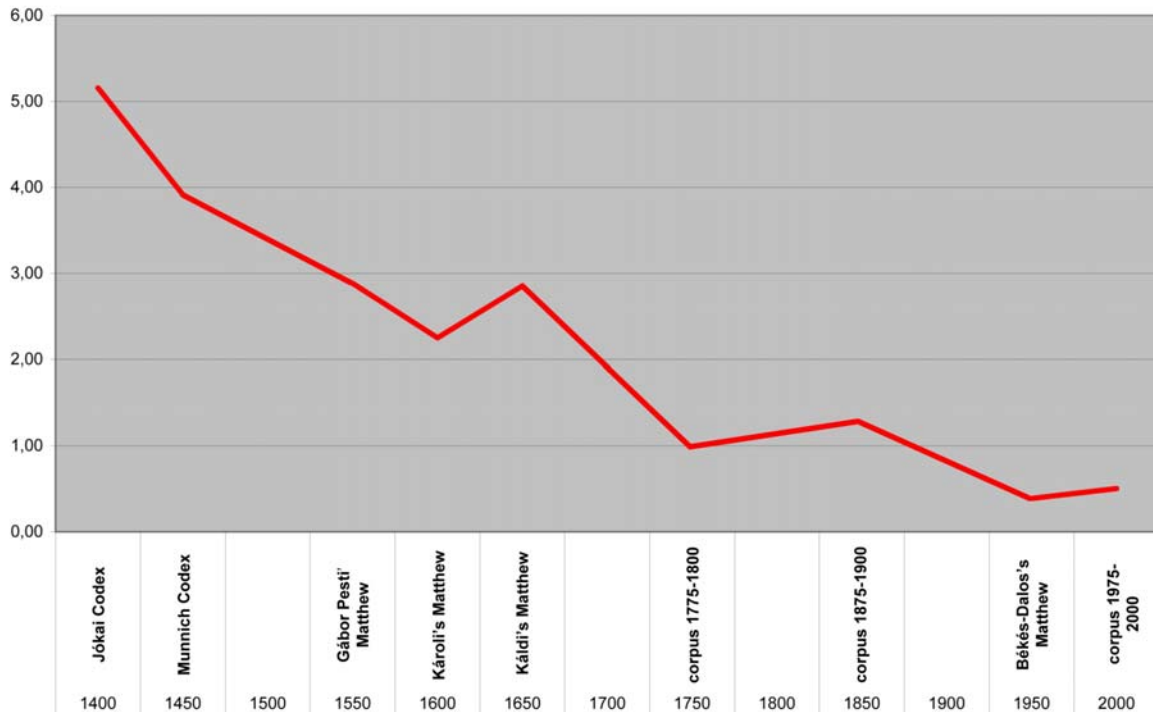
Ezek az adatok azt mutatják, hogy a hangzó egyes szám 3. személyű tárgyi névmások arányának csökkenése és a rejtett névmások terjedése a 20. század közepére gyorsult fel. Hogy arról is képet kapjunk, hogy mi történt 1616 és 1950 között, a kitett *őt* és *őket* névmások arányát az újmagyar szövegek adatbázisában, a Magyar Irodalmi és Köznyelv Nagyszótárának korpusza/Magyar Történeti Korpuszban is megvizsgáltam <http://www.nytud.hu/hhc/>. Az 1775-1800, 1875-1900 és 1975-2000 közötti időszakokból kigyűjtöttem az *őt* és az *őket* névmások valamennyi előfordulását, és a megfigyelt arányokat a Jókai-kódexben és a Máté Evangélium fordításaiban tapasztalható arányokkal is kiegészítettem. Az adatokat számszerűen és grafikus formában is bemutatom.

A kitett őt és őket arányának változása:

Jókai-kódex	(1370):	98/19	5 :1
Müncheni kódex	(1416/1466):	137/35	3,9:1
Pesti G.:Novum Testamentum (1536):		124/43	2,8:1
Károli Máté Evangéliuma (1590):		108/48	2,2:1
Káldi Máté Evangéliuma (1626):		117/41	2,8:1
Szövegtár	1775-1800:	116/108	1,01:1
Szövegtár	1875-1900:	1171/915	1,2:1
Szövegtár	1975-2000:	1317/2621	0,5:1

² Az időben egymáshoz közel eső Biblia-fordítások közül a legismertebbet vizsgáltam. A 20. századi fordítások közül azért Békés és Dalos munkáját választottam, mert ez az a fordítás, mely nem valamely korábbi fordításon alapul, hanem teljesen új.

A kitett *őt* és *őket* arányának változása:



Az adatok azt mutatják, hogy az a feltételezés, miszerint a tárgyas ragozás mindig a rejtett tárgyi névmások engedélyezésére szolgált, nem tartható. Míg a tárgyas ragozás az ugor egység koráig, sőt tovább is visszavezethető, a rejtett tárgyi névmások használata a magyar nyelv önálló életének korai szakaszaiban még szórványos, sőt, a középmagyar korban sem vált általánossá.³ Csaknem kizárólagossá csak az újmagyar korban, a 20. század első felében lett.

4. Összefoglalás

Írásomban amellet érveltem, hogy érdemes a Nyelvtudományi Intézetben készült ómagyar korpusszal megismerkedni, keresőjének használatát megtanulni, és a digitális adatbázis nyújtotta keresési lehetőségeket, a korpuszban rejlő statisztikai mintázatokat kihasználni. Az ily módon feltárható összefüggésekre olyan következtetéseket építhetünk, melyek egyébként

³ Az a tény, hogy a korpusz csak írott szövegeket tartalmaz, torzíthatja az eredményt. Teljesebb képet fogunk kapni, ha az élőbeszédhez közelebb álló középkori levelekben is (Hegedüs és Papp 1991) megvizsgáljuk a rejtett tárgyi névmások arányát. A formálisabb egyházi szövegek feltehetően valamivel korábbi nyelvállapotot tükröznek, mint a gyakran diktált levelek.

rejtve maradnának előttünk, vagy kézi munkával csak rendkívüli erőfeszítéssel volnának kibányászhatók. Írásom első részében azt mutattam be, hogy a korpusz hogyan segíti elő a nyelvemlékes kor előtti nyelvállapot mondattani sajátosságainak megismerését. Minthogy a nyelvi változások során a régi változatnak az új által való kiszorítása, a régi és új változat arányának módosulása egy ún. *f*-görbével modellálható, ha dokumentálni tudjuk az *f*-görbe végső szakaszát, az *f*-görbe egészét, azaz, a változás megelőző szakaszait és kiinduló állapotát is rekonstruálni tudjuk. E módszerrel valószínűsíthető, hogy az ősmagyar alapszórendje SOV volt, a tárgy nem viselt tárgyragot (csak a tárgy topik szerepét jelölte morféma), a határozói és jelzői alárendelő mondatok igeneves kifejezések voltak, és a jelzői mellékmondatban a jelzett kifejezéssel koreferens elem rejtve maradt.

Írásom második részében azt illusztráltam, hogy a korpuszból kinyerhető statisztikai adatokból a mondattani változások funkciójának megértéséhez, a mondattani jelenségek magyarázatához is közelebb juthatunk. Vitatott, hogy a tárgyas ragozás a névmási tárgy inkorporációjával jött létre, az önálló névmási tárgy hiányát eredményezve, vagy topikismétlő névmásból keletkezett, és a topikalizált tárggyal való egyeztetést szolgálta. Az előbbi hipotézist az támasztaná alá, ha a tárgyas ragozás megléte minden korban rejtett tárgyi névmással járna együtt. Azonban, mint kimutattuk, a rejtett tárgyi névmás a korai ősmagyar korban csak szórványosan fordul elő; írott szövegekben csak a 20. századra válik általánossá. Ez azt valószínűsíti, hogy a tárgyas ragozás először nem a tárgyi névmás kiváltását szolgálta, hanem a tárgy topik szerepének kódolását. Amikor ezt a feladatot a topik pozícióba való mozgatás vette át, exaptáció, funkcióváltás történt: a tárgyas ragozás a tárgy határozottságát kezdte jelölni. Amikor ezt a feladatot a határozott névelő vállalta magára, a tárgyas ragozás ismét funkcióváltásra kényszerült: ekkor lett általános az egyes szám 3. személyű tárgyi névmást helyettesítő szerepe; ekkor vált általánossá a 3. személyű rejtett tárgyi névmás.

Hivatkozások

Dalrymple, Mary – Irina Nikolaeva 2011. *Objects and Information Structure*. Cambridge: Cambridge University Press.

Egedi, Barbara 2014. Főnévi kifejezések: Határozottság, névelőhasználat, birtokos szerkezetek. In: É. Kiss Katalin (szerk.) *Magyar generatív történeti mondattan*. Budapest: Akadémiai Kiadó, 96-127.

É. Kiss, Katalin 2003. A szibériai kapcsolat – avagy miért nem tárgyasán ragozzuk az igét 1. és 2. személyű tárgy esetén. *Magyar Nyelvjárások* 41: 321-326.2013

- Givón, Talmy 1976. Topic, pronoun, and grammatical agreement. In: Charles Li – Sandra Thompson (szerk.): *Subject and Topic*. 149-188. New York: Academic Press.
- Dryer, Matthew S. – Martin Haspelmath (szerk.) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Hegedűs Attila – Papp Lajos 1991. *Középkori leveleink 1541-ig*. Budapest: Akadémiai Kiadó.
- Honti, László. 2009. Tűnődések tárgyas igeragozásunk kialakulásáról. *Nyelvtudományi Közlemények* 106:132-146.
- Károly, Sándor 1956. *Igenérendszerünk a kódexirodalom első szakaszában*. (Nyelvtudományi Értekezések 10.) Budapest: Akadémiai Kiadó.
- Kulonen, Ulla-Maia 1999. Object marking in the Ugric languages. In: Bakró-Nagy Marianne – Molnár Zoltán – Salánki Zsuzsanna – Sipos Mária (szerk.), Budapesti Uráli Műhely I. Ugor Műhely. Budapest: MTA Nyelvtudományi Intézet, 63–71.
- Lightfoot, David 1991. *How to set Parameters: Arguments from Language Change*. Cambridge, MA: MIT Press/Bradford Books.
- Marcantonio, Angela 1985. On the definite vs. indefinite conjugation in Hungarian: A typological and diachronic analysis. *Acta Linguistica Hungarica* 35: 267-298.
- Moravcsik Edith 1974. Object–verb agreement. *Working Papers in Language Universals*. 15, 25–140. Stanford: Stanford University.
- Nikolaeva, Irina 2001. Secondary topic as a relation in information structure. *Linguistics* 39: 1-49.
- Simon Eszter, Sass Bálint 2012. Nyelvtechnológia és kulturális örökség, avagy korpuszépítés ómagyar kódexekből. *Általános Nyelvészeti Tanulmányok* 24: 243-264.
- Weinreich, Uriel, William Labov, Marvin I. Herzog (1968): Empirical Foundations for a Theory of Language Change. In: Winifred P. Lehman (szerk.): *Directions for Historical Linguistics*. Austin: University of Texas Press.