

A Magyar Tudományos Akadémia Elnökségének határozata alapján 1984-ben a Nyelvtudományi Intézetben számítógépek bevonásával felújították *A magyar irodalmi és köznyelv nagyszótára (1533–1990)* munkálatait, amelyet több mint száz éve indítottak meg, de a mindenkori szűkös anyagi lehetőségek és a manuális módszerek alkalmazása miatt nehezen haladt előre (lásd *Magyar számítógépes nagyszótár, CW-SZT, 89/28.*). A jelenlegi munkálat tervei szerint az adatgyűjtés során 13 millió szövegszót gépelnek be a számítógépekbe (szemben a korábbi cédulázó módszerrel) öt évszázad magyar nyomtatott szövegeinek részleteiből. A 13 millió szövegszóból 1 millió szó a XVI. századi nyomtatványokból kerül a vizsgálandó korpuszba. A szövegek számítógépes rögzítését 1986 tavaszán kezdték meg, és jelenleg mintegy 5 millió szót tartalmaz az adatbázis.

A nagyszótár korpusza nyomtatott művek részleteiből áll, de néhány, nyelvünk életében kivételesen nagy jelentőségű alkotás teljes terjedelmében számítógépbe kerül. Ilyen mű a Vizsolyi Biblia is, *Károli Gáspár* gönci protestáns prédikátor bibliafordítása. Ez az alkotás minden korábbi és későbbi könyvnel szélesebb hatású és igen kiemelkedő állomás nyelvünk egységesülésében.

Elég legyen itt utalnunk rá, hogy a Vizsolyi Biblia nyelve emelte az *e-ző* nyelvjárást a többi (az *i-ző* és az *ő-ző*) fölé, így válhatott irodalmi nyelvünk alapjává ez a nyelvjárás.

E nemzeti nyelv Bibliának a hatására és elterjedtségére mi sem jobb példa mint az, hogy mikorra az első katolikus bibliafordítás 1626-ban megjelenik *Káldi György* tollából Bécsben, akkora a „gönci prédikátor” bibliafordítását már négyszer nyomtatták ki (Vizsoly 1590, Hanou 1608, Frankfurt 1608, Oppenheim 1612). A Vizsolyi Bibliának vannak nyomtatásban is megjelent előzményei, de egyetlen korábbi mű sem tartalmazta a teljes Ószövetséget és Újszövetséget. De a Vizsolyi Biblia nemcsak teljességével tűnik ki előzményei közül, hanem nyelvezetével és a mai helyesírásunkhoz közel álló hangjelölési rendszerével is. A reformáció szellemiségében született fordításban fejeződött be az a fejlődési folyamat, amelyben kialakult azoknak a magyar beszédhangoknak többékevésbé egységes jelölése, melyeknek a latin ábécé betűivel való írása nehézséget okozott a XVI. századi nyomtatványok készítőinek (*á, é, ő, ü, cs, gy, ny, ty, sz, zs*).

A nagyszótár eddigi munkálatai során már rögzítették a Nyelvtudományi Intézetben a Vizsolyi Bibliából Mózes könyveit és a négy evangéliumot. Jelen cikkünkben bemutatjuk (mintapéldának felhasználva a Lukács evangéliuma szövegét), hogy a számítógépek, szoftverek milyen segítséget nyújthatnak az irodalomtörténészeknek, a hazai könyvnyomtatás története iránt érdeklődőknek és nem utolsósorban a nyelvészeknek. Ismertetjük azokat a lépéseket, amelyeket a számítógépes feldolgozás során tettünk meg. A nagyszótár munkálatai során a számítógépbe begépeltek (nemcsak a XVI. századi) szövegek más kutatásoknak is kincseshányajává válhatnak.

Munkánkat mindvégig IBM XT és IBM AT személyi számítógépen végeztük. A Lukács evangélium fordítását Volkwriter nevű szövegszerkesztővel vitték számítógépre. Az intézet lexikográfiai osztályán használatos „rögzítési utasítás” szerint a magyar nyomtatványokat úgynevezett kibővített Prószyka-kódban rögzítik. Ez azt jelenti, hogy az ékezetes magyar betűket és a régebbi századok magyar nyomtatványaiban előforduló speciális betűket az angol ábécé betűiből és számjegyekből alkotott kódban gépelik be a számítógépbe. Bemutatjuk Lukács evangéliuma első néhány sorának másolatát (1. ábra), valamint ugyanannak a részletnek a számítógépben lévő formáját úgy, ahogy a nagyszótár munkálatai közben gépre vitték (2. ábra).

Feldolgozásunk első lépéseként, a könnyebb olvashatóság érdekében a 2. ábrán

A 400 éves Vizsolyi Biblia számítógépes vizsgálata

Kiefer Ferenc akadémikus, a Nyelvtudományi Intézet igazgatóhelyettese az MTA rendes közgyűlésén a „Paradigmaváltás a nyelvtudományban is?” című előadásában bemutatta, hogy a számítógépek felhasználása az általános és a magyar nyelvtudományban paradigmaváltáshoz vezethet. Újszerű szempontok alapján nagy tömegű információt rendszerezhet a gép, s a számítógépes nyelvelemzés eddig nem ismert adatokat szolgáltat a kutató számára. Tisztelegve a Kárpát-medence legnagyobb hatású XVI. századi könyve, a *Vizsolyi Biblia* fordítóinak emléke előtt, s egyben a fenti gondolatok alátámasztására bemutatjuk a Biblia számítógépes vizsgálatának menetét.

SZENT LVCATS IRASA SZERINT VALO EVANGELIVM.

ELSŐ RESZ.



MIVEL hogy fokban kezdtec rend szerint meg irni, azoknac a dolgoknak meg belíelléséc, mellyec mi nálunc mindeneftől foguan bizonyofoc.

22 Az mint mi nekünc elönkbe adtác, az kic eleitől fogua semmel látói, és siolgái voltanac az igénc.

3 Azért tettéc énnékemis, ki mindeneftől fogua, mindeneknecc iól végere mentem eleitől fogua, hogy azokról tenéked rend szerint irnéc ió Theophile.

4 Hogy meg tudhassád azoknac az dolgoknak valóságos igallágát, mellyeket hallottál.

5 Vala az Iudca tartományának királyánac, a Herodesnec ideiben egy Pap, kinec Zachariás vala neuc, Abiánac rendéből való: Annac pedig felelége, Aaron Papnac leáni közzül való vala: kinec neuc Érsébet vala.

Zachariás Pap és Érsébet felelége, kinec felelége és Áron papnak leáni közzül való vala.

1. ábra. Lukács evangéliumának kezdete a Vizsolyi Bibliában

látható szöveget — a saját számítógépünk (a billentyűzet, a képernyő, a nyomtató) karakterkészletének figyelembevételével — „visszatranszformáltuk” az eredeti szöveg betűképeihez hasonló formátumra úgy, hogy az ASCII karakterkészlet elemeit használtuk fel. A kódok visszaalakítását a WordPerfect 5.0 szövegszerkesztő program makroprogramozási lehetőségeit kihasználva hajtottuk végre. A bemutatott szövegrészlet visszatranszformálás utáni képe a 3. ábrán látható.

A lexikográfiai vizsgálat

Az evangélium lexikográfiai vizsgálatát a WordCruncher általános célú szövegfeldolgozó, konkordancia- és szótárkészítő programcsomag segítségével végeztük el. (A WordCruncher az amerikai Electronic Text Corporation cég által kifejlesztett szoftver.) A feldolgozandó szövegrészt a programcsomag IndexETC programjával készítettük elő (indexeltük): a szövegben jeleket helyeztünk el, hogy a további vizsgálatok során az egyes szavak (kikeresett jelenségek) helye azonosítható legyen. A lokalizációs jeleket az evangéliumban lévő betűkódok visszatranszformálását követően szintén a WordPerfect programmal helyeztük el. A WordCruncher hármas mélységű beosztást engedélyez, így az evangélium eredeti felosztását: 1) evangélium, 2) rész, 3) vers követve helyeztük el jeleinket. A 3. ábrán ilyen jelek a /bLc, /c, /p.

A 4 perc 12 másodpercig tartó indexálás után az IndexETC program a következő jellemző adatokat szolgáltatott a Vizsolyi Biblia Lukács evangéliumáról:

a szövegszavak száma	25 661
a szóalakok száma	6 513
karakterek száma	126 733
a részek száma	24
a versek száma	1 150

Ezzel párhuzamosan az IndexETC előállított egy szólistát is, amelyben az előforduló szövegszavak ábécérendben követik egymást, feltüntetve az előfordulások számát is. (Itt jegyezzük meg, hogy a program rugalmasságára mi sem jobb bizonyíték, mint az, hogy a felhasználó definiálhatja az ábécé elemeinek sorrendjét, és így XVI. századi magyar szövegek esetében is hatásosan lehet használni.) A 4. ábrán ennek a szólistának az *ü* és a *ý* betűkkel kezdődő elemeit mutatjuk be. A *ý* jel a Bibliában szókezdő pozícióban (hosszú) *ú* hangértékű.

Ez az ábécébe rendezett szólista alapjául szolgálhat egy szövegszó-gyakorisági listának. Ennek a gyakorisági listának az elkészítéséhez újabb programokat kell használnunk. Mi a listát a dBASE IV adatbázis-kezelő programba emeltük be, egy alkalmas rekordkép meghatározása után: az APPEND from LISTA type SDF parancs segítségével. Ezt követően a dBASE program SORT parancs segítségével, az előfordulást jelölő mező alapján rendeztük csökkenő gyakorisági sorrendbe a listát, majd a COPY TO GYAKLIST type SDF parancsral másoltuk át szöveges állományba. Ennek a szóalak-gyakorisági listának az elejét mutatjuk be, a legalább 75-ször előforduló szövegszavak felsorolásával az 5. ábrán.

A magyar nyelv toldalékoló (agglutinatív) jellegéből adódóan a szavak vége igencsak izgalmas a kutató számára, ezért úgy készítettünk egy, a szavak vége felől ábécébe szedett és rendezett atergólistát is a szövegszólistából kiindulva, mégpedig a következő lépésekben: a) a szólista magyar betűit (például: *é, ő, gy, ty*) a WordPerfect 5.0 szövegszerkesztő programmal kicseréltük az angol ábécé betűinek és számoknak alkalmas kombinációjára (a szöveg felől kiinduló ábécébe rendezés miatt); b) ezeket a szavakat az erre a célra elkészített Tükröz nevű, Turbo Pascal 5.0 nyelven megírt programmal tükröztük (például „kezeiteket” → „teketiezek”);

c) a tükrözött szavak listáját beemeltük a dBASE IV programba, a SORT paranccsal ábécébe rendeztük a listát, majd a rendezett listát visszatettük egy szöveges állományba;
d) ennek az ábécébe rendezett listának a szavait visszatükröztük („tekietezek” → „kezeiteket”) ismét az előbb már használt Pascal program segítségével.

Nem állítjuk, hogy ez egyszerű művelet volt így, de úgy gondoljuk, hogy a kapott atergólista igen hasznos információkat szolgáltat a szakember számára. A 6. ábrán ennek az atergólistának azt a részletét mutatjuk be, amely a -tanac, -tenec, -and, -end végződésű szavakat tartalmazza, kiegészítve azzal a 18 szóalakkal, melyben a „~” jellel a magánhangzón jelölték a szóvégi nazális (n, m, ny) hangok meglétét. A -tanac és a -tenec végződésű szavak esetében megfigyelhető, hogy szó végén a k hangot c-vel jelölték, valamint figyelmet érdemelnek a ma már kihalt -and, -end ragos jövő idejű igealakok.

A Lukács evangéliuma szólistái (1. szöveg-szólista ábécébe, 2. gyakorisági szóalakkalista, 3. atergólista — hátulról — ábécébe szedve) után bemutatjuk a WordCruncher programcsomag ViewETC programja segítségével előállítható konkordancialistát. A számítógéppel előállított konkordancialisták célszerűségére és nyelvészeti célú felhasználhatóságára Papp Ferenc már tizenöt évvel ezelőtt felhívta a figyelmet. A konkordancalista előállításához segítség, hogy a korábbiakban már említett, az IndexETC program által generált szövegszólista a képernyőn görgethető és tetszőleges elemek kijelölhető. A WiewETC program megmutatja a szólistából kiválasztott szó szövegvörnyezetét. Amennyiben egy szövegszó több helyen is előfordul, az előfordulások környezetét egymás alatt sorolja fel. Sőt a ViewETC lehetőséget ad — sok egyéb szolgáltatás mellett — arra, hogy kikeressük akár több szó együttes előfordulását is (adott karaktertávolságon belül) a vizsgálendő szövegvörnyezetből. Mi most a Biblia rá-, reá-, reia- kezdetű szavainak konkordanciáját mutatjuk be a 7. ábrán.

Hangtani vizsgálat

A Vizsolyi Biblia hangjelölési rendszerét korábban már többen is behatóan vizsgálták. Annak ellenére, hogy e hangjelölési rendszer példamutató volt a későbbi nyomtatványok számára, néhány hangot, így a ny-t és a gy-t négyféleképpen is jelölték a Bibliában. Igaz, hogy a ny hangjelölése nyelvünkben igen problematikus volt minden időben. Kniezsa István helyesírásunk történetét a könyvnyomtatás koráig tárgyaló műve 15 betű- és jelkombinációt mutat be az ny jelölésre a könyvnyomtatás előtti korból.

Jelen feldolgozásunkban célul tűztük ki, hogy a számítógép adta lehetőségek felhasználásával szemügyre vegyük az ny hangjelölésének variánsait, a Vizsolyi Biblia Lukács evangéliumában, és így számszerű adatokkal szolgálhassunk az egyes hangjelölési variánsok elterjedtségéről a korábbi szubjektív ítéletekkel szemben. Vizsgálódásunk alapját a korábbiakban bemutatott listák alkották. Ezekből gyűjtöttük össze a WordPerfect program makrójának segítségével azokat a szavakat, melyekben szerepel az ny hang; ny, nj, ni, n, ny, nny betűkkel, jelekkel jelölve. Manuális beavatkozásra volt szükség az ni-t tartalmazó szavak esetében. Ugyanis a kigyűjtött szavak listájából töröltük azokat az elemeket, amelyekben a ni betűkombináció nem jelölhet ny hangot (például „tanítani”). Az összegyűjtött szavakban lévő ny-jelöléseket a Hangtat nevű Turbo Pascal programmal számoltattuk össze (8. ábra).

Tervezzük a szövegszólista bokrosítását (lemmatizálását) olyan formában, hogy egy-egy szótári címszó alatt legyenek felsorolva az adott címszónak a Bibliában található toldalékos alakjai. Ehhez segítséget nyújthat egy morfológiai analízisát végző, a XVIII. századi magyar nyelvet és annak nyomtatott változatát „ismerő” program.

SZENT LVCATS IRASA SZERINT VALO EVANGELIVM. # (1 ELSO20 RESZ.)1 #/ 1 (1 MIVEL)1 hogy s43okan kezdettec rend s43z40errint meg irni, azoknac a4 dolgoknac meg bes43z40e111e1s43e1t, mellyec mi nallunc mindenes43to20l foguan bizonyos43oc. #/ 22 (!) Az mint mi nelku20nc elo20nke adtalc, az kic eleito20l fogua s43z40emmel laltoli, els s43z40olgalil voltlanac az igelne. #/ 3 Aze1rt tettzeic elnnelkemis, ki mindenes43to20l fogua, mindenekne iol1 velge1re mentem eleito20l fogua, hogy azokro1l te nelked rend s43z40erint irnelc iol1 Theophile. #/ 3 Hogy meg tudhas43s43ad azoknac az dolgoknac valo1s43algos igas43s43algal1t, mellyeket hallottall. #/ 5 (9 Zacharias Pap, els Ers43ebeth as43z40s43z40on, Kereztelo20 Ianos s43z40u20leij.)9 Vala az lude33a tartomalyalnac kirallyalnac, a4 Herodes43nec ideielben egy Pap, kinec Zachariás vala neue. Abialnac rendelbo20l valo: Annac pedig feles43e1ge, Aaron Papnac lealni ko20zzu20l valo1 vala: kinec neue Ers43ebet vala.

2. ábra. Az evangélium a számítógépen — a nagyszótár munkálatai közben

Előfordulás	Szövegszó	Előfordulás	Szövegszó	Előfordulás	Szövegszó
5	ül	5	ülün	4	üzöc
5	üle	1	ülvén	1	úgj
1	ülénec	2	ünnepnec	22	úgy
1	üléj	1	üres	1	úntajson
2	üléjt	3	üreflen	1	út
2	üll	1	üljed	1	útac
5	üllly	1	üté	1	útadat
1	ülllyetec	1	ültejeket	1	útait
1	ülne	1	üti	1	útakra
2	ülne	1	ütökzic	1	útára
1	ülne	1	ütözött	2	útát
1	ülni	1	ütözzéc	1	úti
1	ülő	3	ütött	10	úton
1	ülött	3	üz	3	útra
2	ült	1	üzéc	1	útról
2	ültec	2	üzne	1	útzáinkon
1	ültejtetec	1	üzne	2	úzáira
1	ülteti	1	üzni		

4. ábra. A Lukács evangéliumának szövegszólistájából az ü és (az ú hangértékkel bíró) v kezdetű szavak

Előfordulás	Szövegszó	Előfordulás	Szövegszó	Előfordulás	Szövegszó
1	adtanac	1	tagadand	1	tisztátalá
1	hadtanac	1	mondand	1	tisztátalá
1	alitanac	2	haland	1	óthatatlá
1	Boritanac	1	valland	1	Babadóluá
1	raktanac	1	Bóland	1	gyönyörü/égbe
3	kiáltanac	1	Bóland	1	törüenibé
1	hóltanac	1	kérdend	1	öltözetbe
1	Bóltanac	6	lelend	4	mindé
4	vóltanac	1	kelend	2	ellé
1	eppitenec	3	ménend	13	né
2	tóltanec	2	rend	3	hané
1	Bültanec	1	térend	2	migné
2	teremtanec	2	veftend	1	lébé
4	mentenec	5	iöüend	1	mégyé
2	intanec	1	yöüend	1	mondó
2	téranec			1	vagjó
1	vetetanec			1	azokó
3	lötanec			1	mikoró
1	töttenec				

6. ábra. Az atergólista -tanac, -tenec, és and, end végződésű szavai, valamint a szóvégi magánhangzón a nazális (~) szavak listája

Lukács:
23:28 köueti. Iefus pedig ö ráioc fordulóán, monda: Ierufálemnec
21:34 miatta, és nagy hértelen réioc iöijön az az nap. Mert mint
10:35 erre, és valamit ezen fellül reá iöit, én mikor meg térendec
19:35 vehmére vetuén, Iefu/t reá helhezettec. Es mikoron ö
22:56 hogj az tüznél ünle, Bemeit reá vetuén monda: Ezis ö véle vala
1:35 nékie: Az Bent lélec Bál te reá, és az magaffágban lakozó
1:12 háborodéc, és félelem eféc ö reáia. Monda pedig az Istenec
7:13 vala ö véle. Es az Wr ö reáia tekintuén, könyörüle rayta, és
11:22 valaki ö nálanál erőffeb ö reáia iöüén, ötet meg gjözendü,
13:13 az te betegfégedből. Es reáia veté az ö kezéit, és azonnal
20:19 az irás tudóc az ö kezéket ö reáia vetni azon órában, de félelenc,
9:55 czelekedett? De az Iefus reáioc fordulóán meg dorgálá öket,
1:25 napokban, mellyekben Bemeit én reám veté, hogy el venné én róllam
22:53 ti kezeiteket nem vetettétec én reám. De ez amaz ti óratoc, és az
10:11 Még az portis az melly mi reánc ragadott az ti vároftokból itt
23:30 az hegyeknec, effetec mi reánc, és az halmoknac, buritfatoc el
10:6 marad, hogj ha nem lelend, ti reátoc tér. Vgyan azon házban
16:11 nem vóltatoc, kitfoda bizná reátoc igaz gazdagfágot?
16:12 mi tijeitc kitfoda bizná azt reátoc? Egy Bólgais nem Bólgálhat
24:49 Es imé én el küldöm ti reátoc az én atyámnac igéretit, ti
1:48 nem vetett el engemet hanem reiam tekintet. Mert meg tekinté
4:23 azt mondhattac volna. Te mi reianc haragból hogj tegedet be nem
1:9 Az Papi tiñtne rendi Bérin, reia iutá az Bér, hogj az Wrnac
3:22 ábratban mint egy galamb, ö reia, és Bózat lenne menyből, ezt
4:20 Bemeit az Synagóában ö reia fúggettettec vala. Es
5:1 ES lön hogj mikor az foka/fág reia Borúlna az Iften beBédénec
23:34 pedig az ö ruháit veténe reia forfot. Es az köffég áll
18:5 Babaditom ötet, hogj Büntelen reiam iáruán ne vntajson engemet.
3:7 mutatta meg néktec, hogj á reiatoc köuetkezendü haragot el
21:12 De mind ezeknec előtte kezéket reiatoc vetic, és titeket kergetnec,

7. ábra. A Lukács evangéliumában található rá, reá, reia kezdetű szavak konkordanciája

/bLc/c1/p1 MIVEL hogj fokan kezdettec rend Rerint meg irni, azoknac a dolgoknac meg beRelléféte, mellyec mi nálunc mindeneftől foguan bizonyofoc. /p2 Az mint mi nekünc elönkbe adtác, az kic eleitől fogua Rimmel látói, és Rólgái vóltanac az igéne. /p3 Azért tettzec énnékemis, ki mindeneftől fogua, mindenekne iol1 végére mentem eleitől fogua, hogj azokról te néked rend Rerint irnéc ió Theophile. /p4 Hogj meg tudhafad azoknac az dolgoknac valófágos igaffágát, mellyeket hallottál. /p5 (9 Zacharias Pap, és Erfebeth arRon, Kereztelo Ianos Rüleij.)9 Vala az ludea tartományánac királyánac, a Herodéfne ideielben egy Pap, kinec Zachariás vala neue. Abialnac rendéből való: Annac pedig felefége, Aaron Papnac leáni közzül való vala: kinec neue Erfebet vala.

3. ábra. A betűkódok visszatranszformálásával nyert szövegforma

előfor- dulás	szöveg- szám	előfor- dulás	szöveg- szám
25.	119 fel	25.	119 fel
26.	110 nékiec	26.	110 nékiec
27.	110 mi	27.	110 mi
28.	104 bé	28.	104 bé
29.	103 ti	29.	103 ti
30.	97 ha	30.	97 ha
31.	92 vagon	31.	92 vagon
32.	91 ezt	32.	91 ezt
33.	84 monduán	33.	84 monduán
34.	84 vólt	34.	84 vólt
35.	78 néktec	35.	78 néktec
36.	75 Iften	36.	75 Iften
37.	75 ne	37.	75 ne
38.	73 de	38.	73 de
39.	72 azért	39.	72 azért
40.	72 annac	40.	72 annac
41.	72 Iefus	41.	72 Iefus
42.	69 néki	42.	69 néki

5. ábra. Részlet a Lukács evangéliumában legalább 75-ször előforduló szövegszavak listájából

A hajdani bibliafordítók elhivatottsága

A Vizsolyi Biblia, melynek nyelve Arany Jánosra, Ady Endrere és Móricz Zsigmondra is hatott, országunk történetének nehéz szakaszában született meg. Mindnyájunk számára példamutató az az elhivatottság és tenniakarás, mellyel Károli Gáspár és néhány segítőtársa az akkor három részre szakadt ország egy eldugott szegletében megalkotta az első, teljes, magyar nyelvű nyomtatott Bibliát.

Befejezésül idézzük Károli Gáspár szavait a Biblia előszavából: „... a nagy Istennek tisztességére, ő házának épülésére néztem e dologban... szabad mindenek az Isten házába ajánlékot vinni. Egyebek vigyenen aranyat, ezüstöt, drágaköveket, amit (én) vihetek, tudniillik, magyar nyelven az egész Bibliát.”

Kiss Gábor
MTA Nyelvtudományi Intézet

Az ny hang jele	A szóban elfoglalt pozíció			Összesen
	szó elején	szó belsejében	szó végén	
ny	36/43	110/188	14/36	160/266
nj	1/1	4/5	3/3	8/9
ni	9/12	29/49	12/48	50/109
ñ	0/0	25/45	4/7	29/52
ñy	0/0	3/4	0/0	3/4
nny	0/0	30/45	0/0	30/45
Összesen	46/56	201/336	33/94	280/486

8. ábra. A Vizsolyi Biblia Lukács evangéliumában az ny hang jelölésének változatai és az előfordulások száma (szóalak/előfordulás)