

Egy ötfős magyar kutatócsoport az elmúlt évek során a keresztény hebraisztika eddig ismeretlen darabjainak feltárását tűzte ki célul: 17. századi héber nyelvű köszöntőversek (*carmina gratulatoria*) vizsgálatát, amelyeket a korabeli külföldi egyetemeken tanuló (peregrinus) magyar protestáns diákok egymásnak írtak. A 17. századi versek közzététele, vizsgálata, fordítása a magyar művelődéstörténet és régi magyar irodalom számára is rendkívüli jelentőséggel bír. Az irodalomtörténet elvben tudott ezekről a versekről, de évszázadokon keresztül nem történtek kutatások e téren. Kutatócsoportunk ezt a tudománytörténeti adósságot törlesztette: hazai és külföldi levéltárakban, könyvtárakban 72 héber köszöntőverset – hebraica hungaricát – talált, digitalizált, írt át, fordított le magyarra és angolra, melyekhez a www.cgh.hu honlapon immár bárki hozzáférhet. Jelen kötetben a kutatásról, a versekről, szerzőikről és a magyar peregrinációhoz való kötődésükről szóló tanulmányok olvashatók. Az írások szóbeli változatai a kutatás zárókonferenciáján, 2023. április 28-án hangzottak el az MTA Könyvtárában.



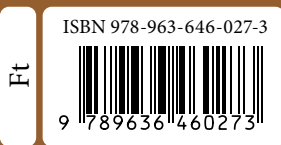
Koltai Kornélia és Zsengellér József

szerkesztette:



HÉBER ÜDVÖZLŐVERSEK ÉS A PEREGRINÁCIÓ

שיר פשוט ומחולק



L'Harmattan

STUDIA HEBRAICA HUNGARICA LX.

שיר פשוט ומחולק HÉBER ÜDVÖZLŐVERSEK ÉS A PEREGRINÁCIÓ

Szerkesztette: Koltai Kornélia és Zsengellér József

Magyar Hebraisztikai Társaság – L'Harmattan

X.
Haud vacuus patrias liquisti dotibus oras,
 Non fuit Ingenii vellus inane Tui.
 Nec tum pectus inops rerum, nec lingua Sūadæ,
 Nec fuit exiguus Religionis amor.
 At nunc, quantus eò, *Doctōr Reverende*, redibis,
 Dotibus ò quantis auctior ingenii!
 Divitiis Priami potiores exere gazas,
 Exere mox Patrio lumina tanta solo.

JOH. HENRICUS HOFMANNUS, Basil. Eccl. in
 Summ. Templ. Diac. Convictori suo appofuit.

שמחה מלבב

XI.
 אין נתון אלהים
 לבד הרשעים;
 לנצח: כי רעים
 אשר לה אויבים
 יש גואל אביונים;
 ישע עד עלמים
 יש מועד לאלהים;
 יציל מרדפים
 נלחצנו מפיוור
 גלחים כיו בור
 אמנם שרי אשר
 הוא יבא; ויבשר
 הן לנוגים נביא
 לנחם אז לבא
 זה השם יברך;
 שלאנן; בדרך
 כי הן זה מלמד
 גם מכל עם נכבד
 הכבוד לאלהים;
 אמרו מרמו עמים;
 האיר את חרם
 ואל אין הרש
 על יונה צהלו
 כל רעה פעלו
 אשר לא ישן
 מצמים דשן
 מתי חסי בו
 אם הפן לבו
 ושטן הגיל
 והוגים החיל
 עתה ברחוקים
 לעמו תנחומים
 ישלח עמים;
 יפול הצמים
 וישב עמו
 לחכם לבו
 יעטר ביקר;
 כי הוא רב נכר
 אותו הלו יה
 אותו הלו יה

Ita Clavissimo, atq; Doctissimo D. Candidato, in honoribus justiffimis
 Doctōr abbas, ex corde sincero cecimir latus, & animo ag-
 gratulante impulsus vota sua fecit
 JOHANNES BEKES, Horvāti Ungarus;

XII.
 Στ' δέχρα ὦ Μοισῶν παῖ γλώσσας αἰ
 Τέθμισεν ἔστ λαβῆεν, μοὶ δὲ εὐο
 ὦ, ἐπεὶ ἀθανάτης γραφῆς πη' ἵχνια
 Ἦ' δὲ ἀδαρβήτων ἔθνεα συγγε
 Νῦν ῥα θεοῦ Σ πέρ πτωι σάζει χάρε
 Ὅτι πέλεισ αὐτῶ μάρτυς αἰσ
 Γεῖμμα αὐτῶ σὺ ἀρκίῃ πησῶν παρτίπυσι
 Ὅϊόν σὺ ἀρκίῃ, ὅ, πη σελῆ, κ
 Σάμερον ἔοσι κάρην λδοκωθῆς ἔχ
 ΘΕΙΟΛΟΓΩΝ κλυτῶν αἰθῆ
 Πολλὰ παθῶν ῥα θυμῶ πηλλὰς δὲ πῶ
 Γ' αὐτομῆην ὁρῆεις δὲ λογίαισι
 Πανύχιος μετ' ἰγῶ λακίσιω αἰθῆρῶ
 Ὡς ἴσθον κῦδῆς σὺ μετ' αὐτῶ
 Ἐκρῶνε, ΧΡΙΣΤΕ, ΘΕΟΤ' αὐ δὲ
 Ὅ φρε μετ' αὐ κῦσιν αἰθῆρῶ χ
 Πλένεα ἀρλεπῆς νυ ἔχεις τοξόματ
 Ταῦτα δὲ μελλῶν ὡς σῶπ
 Βραχέων ἐπίων εἴπερ δὲ ἐκ ἔφθιν
 Ἀμμιν χεῖρες αἰετῆ καὶ ἴπ
 Ad Verbum

EXcipe Mufarum Fili pia pascua
 Accipere Tuum est; Exhi
 Cui post emensas Divina Oracu
 Hac repetita tenus post mo
 Dia decus pleno nunc spargit g
 Teflis cui fidus, Praeco bo
 Cuius Apostolicae solae fundam
 Hinc redolens hodie gloria
 Haec est illa dies Frontem qua
 THEOLOGICI tanti qua Tu

שיר פשוט ומחולק
Héber üdvözlőversek
és a peregrináció

Studia Hebraica Hungarica, 9

Sorozatszerkesztő: Zsengellér József

שיר פשוט ומחולק
Héber üdvözlőversek
és a peregrináció

Szerkesztette: Koltai Kornélia és Zsengellér József

Magyar Hebraisztikai Társaság – L'Harmattan Kiadó
Budapest, 2023

© Tanulmányok szerzői, 2023

© Szerkesztők

© Magyar Hebraisztikai Társaság

© L'Harmattan Kiadó

ISSN 2061-1781

ISBN 978-963-646-027-3

A kiadásért felel: Gyenes Ádám és Bányai Viktória



A projekt száma: NKFI K-125486/142453

A kötet borítóján elől Horváti Békés János héber köszöntőverse Veresegyházi Tamás bázeli doktorálása alkalmából (RMK III. 2647), hátul a magyar peregrinusok domborműve az Utrechti Egyetem főépületében (Academiegebouw).

Tördelő: Kovácsné Daróczy Annamária

Nyomda:

TARTALOMJEGYZÉK

Bevezető: A 17. századi magyar peregrinusok héber köszöntőversei – projektzáró értékelés (Zsengellér József)	5
Zsengellér József: Miért éppen a 17. század?	31
Dienes Dénes: Kényszerű változások és következmények a 17. század első felének peregrinációjában	49
Juhász Szandra: A héber nyelvű <i>carmina gratulatoria</i> szerzőinek, címzettjeinek relációi	63
Bartók István: Héber üdvözlővers-írók szerepe a 17. század irodalmi gondolkodásának alakításában	95
Koltai Kornélia: Túlások és töréspontok – stilisztikai trendek a 17. századi héber üdvözlőversekben	105
Biró Tamás: PeregrXML és egy középkori keresztény héber korpusz. A köszöntőversek a digitális bölcsészet szemszögéből	131
Götz Andrea: What can Hebrew poems tell us about 17 th -century Hungarian peregrination? Using metadata and frequency analyses for sociopragmatic research	153

Ferenc Postma – Zsengellér József: Két eddig ismeretlen héber nyelvű köszöntővers Csedregi Péter <i>Disputatio theologica de miraculis</i> című művéhez (Franeker, 1659)	177
A kötet szerzői	209

PEREGRXML ÉS EGY KÖZÉPKORI KERESZTÉNY
HÉBER KORPUSZ
A KÖSZÖNTŐVERSEK A DIGITÁLIS BÖLCSÉSZET
SZEMSZÖGÉBŐL

Biró Tamás

KÉPEKTŐL A PUBLIKÁCIÓIG – A 21. SZÁZADBAN

Projektünk a kora újkori – elsősorban 17. századi¹ – „hallgatói mobilitással” foglalkozik. Magyarországi és erdélyi protestáns teológusokra fókuszáltunk, akik főként németalföldi és németországi egyetemekre peregrináltak. Az általuk írt héber nyelvű köszöntőverseket gyűjtöttük össze, hogy feldolgozzuk és publikáljuk őket.

Akkoriban ilyen költeményeket nyilvános előadások, védések, viták alkalmából írtak hallgatótársak és professzorok, akik nem csupán üdvözölték az adott eseményt, és kommentálták az *oratio*, a *dissertatio* vagy a *disputa* tárgyát, de egyúttal szórakoztatták is a jelenlévőket szellemi teljesítményükkel.² A versek többsége a korabeli tudomány nyelvén, latinul született, de találkozhatunk görögül, héberül, sőt arámi, szír, arab, valamint a diákok anyanyelvein (hollandul, németül, magyarul...) írt költeményekkel is.³ A disszertációval, előadással egy kötetben, annak függelékeként meg is jelentették ezeket a műveket. Projektünk fókuszában azok a köszöntőversek állnak, amelyeket (*H*)*Ungarus-Transylvanus* diáktársak írtak *Ungarus-Transylvanus* diákok köteteibe. Előbbiekre mint a költeményeik *szerzőire*

¹ Lásd Zsengellér József „Miért éppen a 17. század?” című cikkét jelen kötetünkben, Zsengellér 2023.

² Habár a folytonosság nem bizonyítható, de hasonló funkciót betöltő műfajjal a 21. század elején is találkoztam Hollandiában: az egyik tanszéki kolléga – a történelmi adatbázisok kutatásának egyik úttörője – személyre szabott humoros dalt komponált gitárra minden doktori védés alkalmából.

³ Postma–Van Sluis 1995: 705–706. Legalább tizenöt nyelv csak Franekerben.

hivatkozom az alábbiakban, utóbbiak pedig a költemények *címzettjei*, azaz a *kötetek szerzői*.

Ezeknek a kora újkori költeményeknek a nyelvét a *középkori keresztény héber* (*medieval Christian Hebrew*) kifejezéssel adhatjuk vissza.⁴ A héber nyelv hagyományosan négy nagy korszaka közül – bibliai héber, rabbinikus/misnai héber, középkori héber, modern héber – a harmadikba sorolható, annak sajátos változata. Nem csupán a szövegek teológiai aspektusai különböznek a korabeli zsidó szerzők műveitől, de nyelvészeti értelemben is sajátosak. Míg a középkori zsidó héber a misnai héberhez, addig a középkori keresztény héber a bibliai héberhez áll közelebb – evidens módon, gondolnánk, hiszen keresztény teológiákon a Héber Bibliát, a *jesivákban* pedig a Talmudot tanulták a hallgatók. Jóval kevésbé evidensek a rabbinikus hatások a középkori keresztény héberben.⁵ Jelen tanulmány szempontjából viszont a legfontosabb azt kiemelni, hogy sem a középkori zsidó hébernek, sem a középkori keresztény hébernek nem voltak anyanyelvi beszélői, így a vizsgált művek néha magasabb, néha alacsonyabb nyelvi kompetenciával rendelkező szerzők tollából születtek, és úgyszintén változó hébertudással rendelkező nyomdászok munkája révén jutottak el hozzánk. Egyszer, kis példányszámban kiadott szövegekkel dolgozunk. Az esetleges szerzői és nyomdahibákat egy későbbi kiadásban nem javította a szerző. A szövegek feldolgozása során nem tudhatjuk, hogy ha valamit nem értünk, az a mi hibánk, vagy a szerzőé, esetleg a nyomdászé, talán a versforma kényszeréből agrammatikus egy kifejezés, avagy a szerző és a címzett közötti teológiai utalások, netán egy bennfentes vicc nem világos-e számunkra.

A projekt során képfájlként beszerzett költeményeket fel kellett dolgozni, hogy azokat publikálni, valamint kutatási kérdésekre válaszolni lehessen. A kutatási kérdések számos diszciplínából érkehetnek: kultúrtörténeti, filológiai, irodalomtudományi, nyelvészeti, nyomdászattörténeti és teológiai problémákat is

⁴ Biró 2018.

⁵ Koltai 2021.

felvetnek ezek a költemények. Tanulmányomban a digitális bölcsészet nézőpontjából reflektálok a kutatási projektekre.

A DIGITÁLIS BÖLCSÉSZETRŐL RÖVIDEN

A digitális bölcsészet elődje, a *humanities computing* („bölcsészeti informatika”) már a 20. század második felében jelentős eredményeket mutatott fel.

Alig jelentek meg az első digitális számítógépek, amikor 1949-ben, amerikai útja során Roberto Busa olasz jezsuita szerzetes (1913–2011) rávette az IBM vezetőjét, Thomas J. Watsont, hogy a három évvel korábban elkezdett projektjét, Aquinói Szent Tamás műveihez készülő konkordanciát, lyukkártyás gépek segítségével valósítsák meg.⁶ Az *Index Thomisticus* projekt honlapján napjainkban is kereshetünk Aquinói Tamás műveiben.⁷ Egy kevésbé sikeres projektben, vatikáni kezdeményezésre, 1957 és 1960 között Busa a módszerét az alig egy évtizeddel korábban felfedezett holt-tengeri tekercsek digitalizálására alkalmazta volna.⁸ Hebraisztikai vonatkozása miatt jelentős számunkra az 1963 óta futó *Bar Ilan Responsa Project*,⁹ amelynek ma már egy pendrive-on is elérhető termékét mi magunk is használtuk a köszöntőversek bibliai allúziói feltárása érdekében.¹⁰

A huszadik század második fele tehát, elsősorban, szövegek digitalizációjáról szólt; továbbá ezen szövegek (valamint a század végén már eredetileg is digitálisan létrehozott – ún. *born-digital* – szövegek) digitális feldolgozásáról. Az ezredfordulót követően, 2002-ben, egy elsőéves bölcsészkarai bevezető informatikai kurzuson már könnyedén összedobtunk olyan programcskákat, amelyek pár

⁶ Fantoli 2013.

⁷ <https://www.corpusthomaticum.org/>.

⁸ Fantoli 2013: 678.

⁹ Kopel 2014. Ladon 2022: 60.

¹⁰ Az egyelőre fejlesztés alatt álló *Dicta Analytical Tools for Hebrew Texts* (<https://dicta.org.il/>) olyan eszköz lehet a közeljövőben, amelynek segítségével részben automatizálható a bibliai utalások beazonosítása.

másodperc alatt összeállítottak tetszőleges szöveghez szógyakorisági listát, szóvégmutato szótárt vagy konkordanciát.

Ezzel a folyamattal párhuzamosan, szintén az 1950-es évektől kezdődően, a számítógépes nyelvészet is egyre hatékonyabb módszereket fejlesztett ki mondattani és alaktani elemzésekre, információkinyerésre, szövegek kategorizálására, gépi fordításra, a szövegekben történő keresésekre és így tovább.¹¹ De a digitális bölcsészettudomány már ekkor sem korlátozódott a szövegekre és a nyelvre. Végiglapozva az 1966 és 2004 között megjelent *Computers and the Humanities* című szaklapot, a legkorábbi évektől kezdve találkozunk irodalomtudományi, történelemtudományi, muzikológiai, antropológiai témájú cikkekkkel is.¹²

A fordulatot és az új elnevezést a 2004-ben megjelent *A Companion to Digital Humanities* című kötethez köthetjük.¹³ Az új elnevezés a digitalizáción túlmutató, szélesebb perspektívát nyitott meg, és azóta a *digital humanities* (DH) divatszóvá vált. Sorra nyíltak meg a digitális bölcsészettel foglalkozó intézetek és szakok a világ nagy egyetemlein, a sikeres kutatási pályázatoknak szinte kötelező elemévé vált egy digitális alprojekt beiktatása, és ma már több bölcsészkaron is kötelező alapozó tárgyként oktatják a DH-t. A *European Association for Jewish Studies* (EAJS) keretén belül 2018 óta működik a *Digital Forum*, amelynek célja az, hogy elősegítse a digitális bölcsészet elterjedését a judaisztikában.¹⁴

A digitális bölcsészet tárgya nagyon sokféle: többek közt irodalmi és történelmi adatbázisok, digitalizált szövegek, képek, térképek és zene is belefér. A terület intenzív módon együttműködik a nyelvtechnológiával (természetes nyelvfeldolgozás, NLP¹⁵), amely fontos eszközöket biztosít a DH számára, valamint

¹¹ Jurafsky és Martin 2008, Manning és Schütze 2000.

¹² <https://www.jstor.org/journal/comphuma>.

¹³ Schreibman et al. 2004.

¹⁴ <https://www.eurojewishstudies.org/digital-forum-about-us/>. A judaisztikai vonatkozású digitális kutatások keresztmetszetét nyújtják az EAJS Digital Forum által publikált *showcases*, akárcsak az OR-ZSE Digitális Bölcsészet Műhelynek a Facebook-oldala (<https://www.facebook.com/profile.php?id=100063683971184>).

¹⁵ A számítógépes nyelvészet (*computational linguistics*) egyaránt lefedi a nyelvészeti elméletek formális, matematizálható aspektusait és ezek mérnöki alkalmazásait. Ahogy arra egykori tanárom, John Nerbonne rámutatott, az angol elnevezésben szereplő *computational* szó nem

a hálózattudománnyal is, amely újszerű kutatási kérdésekkel gazdagítja a humán kutatásokat.¹⁶ Jelentős terület a kulturális örökség digitális archiválása és bemutatása, a digitális muzeológia is; akárcsak a *born-digital* tartalmak (például a ma már „muzeális” weboldalak az i. sz. II. évezred utolsó évtizedéből) hosszú távú megőrzése és elemzése.¹⁷ Ezen felül a digitális bölcsészetben 2004 óta az önreflexió is jelentős szerepet kap: reflexió a digitális alkalmazások társadalmi aspektusaira, reflexió a DH-kutatások módszertanára, számítógépes forráskritika (*source code criticism*) és a terület önmeghatározásáról szóló folyamatos diskurzus.¹⁸

A munkafolyamat – erre projektünk is példát szolgáltat – három fázisra osztható. Ez a hármasság minden tudományterületen megfigyelhető, de a digitális bölcsészetben látványosabban elkülönülnek egymástól. Először az adatokat kell létrehozni: szövegeket, képeket, térképeket vagy zenéket digitalizálunk, összegyűjtjük a *born-digital* tartalmakat, és ezekből relációs adatbázisokat, hálózatokat vagy annotált korpuszokat építünk. Második lépésben elkészítjük, elkészítettjük vagy „a polcra leemeljük” a mások által korábban már elkészített (*off-the-shelf*) eszközöket: keresőprogramokat, adatbázis-kezelőket, alaktani és mondattani elemzőket, adatbányászati (*big data*) szoftvereket... Harmadik lépésben pedig megválaszoljuk a kutatási kérdéseinket. A valóságban ez a három munkafázis összefonódik, hiszen az adatok gyűjtését és a felhasználandó eszközöket a kutatási kérdésünk határozza meg, míg a feltehető kutatási kérdéseket behatárolhatják a rendelkezésünkre álló erőforrások, adatok és eszközök.

A digitális bölcsészet haszna mellett 1+3 lépésben érvelek. Mindenekelőtt, önmagában érték lehet az adatok szisztematikus összegyűjtése és az ezek feldolgozásához szükséges eszközök megteremtése: a DH új korpuszokkal, adat-

feltétlenül utal szilíciumalapú számítógépre, abba belefér az elméleti számítástudomány is. Habár már a kezdetektől, az 1950-es évektől fontos motiváció volt a gyakorlati alkalmazhatóság, a terület fő hangsúlya az ezredforduló környékén tolódott el a természetes nyelvfeldolgozás (*natural language processing*, NLP), azaz a nyelvtechnológia (*language technology*) irányába.

¹⁶ Egy bibliatudományi példa Czachesz 2016.

¹⁷ Például egy késő huszadik századi magyar zsidó levelezési lista kutatási lehetőségeiről lásd Biró 2019 és 2020.

¹⁸ Vö. Schreibman et al. 2016.

bázisokkal, hálózatba rendezett információval gazdagította és gazdagítja a bölcsészeti- és társadalomtudományi kutatásokat. Azáltal, hogy arra kényszerítjük magunkat, hogy az adatainkat valamilyen formalizált struktúrában helyezzük el, nem csupán észrevehetjük a hiányzó adatpontokat, de új perspektívába is kerül kutatásunk tárgya. Az adatstruktúra formalizáltsága a vizsgált jelenség mélyebb fogalmi megértését, például fogalmi distinkciók megfogalmazását teheti lehetővé, illetve teszi szükségessé. Erre szolgált példát jelen tanulmányom is.

Túl az adatokon, a kutatási kérdések kapcsán a DH hasznosságát három esetrebontom. Először is, vannak azok a kérdések, amelyek korábban is megválaszolhatók voltak, de a DH hatékonyabbá teszi a kutató munkáját, időt és energiát takarít meg neki. Roberto Busa egy cédulakatalógussal is megalkothatott volna Aquinói Tamás korpuszából egy konkordanciát, és szógyakorisági szótárak már a két világháború között is megjelentek – fáradalmas munka eredményeképp.¹⁹ Ezzel szemben napjainkban két Google-kereséssel pár másodperc alatt eldönthető, hogy két kifejezés vagy írásmód közül melyik a gyakoribb a weben. Másodszor, bizonyos kutatási kérdések korábban elméletben vagy praktikusán megválaszolhatatlanok maradtak. A teljes 19. századi magyar irodalom konkordanciájának az összeállítására senki nem vállalkozott. A kortárs irodalom konkordanciája pedig elméletben is megalkothatatlan lett volna, mert új művek gyorsabban jelennek meg, mint ahogy ezekhez elkészítették és kiadták volna a konkordanciákat. Harmadszor, egyes kutatási kérdések korábban fel se merültek. Például mondattanilag elemzett korpuszokra a nyelvtechnológia miatt merült fel az igény, és ezek létrehozásában a nyelvtechnológia segített. Hálózatkutató elemzések, valamint az irodalomtudományban a távoli olvasás (*distant reading*)²⁰ és a *stilometria*²¹ csupán néhány példa a digitális bölcsészet által bevezetett újszerű kutatási kérdésekre, módszerekre, megközelítésekre.

¹⁹ Például Nemes 1933.

²⁰ Például Péter 2016.

²¹ Például Kiss 2019.

A MUNKAFOLYAMATUNK A DIGITÁLIS BÖLCSÉSZET PERSPEKTÍVÁJÁBÓL

A projekt során a régebbiről ismert és újonnan felfedezett verseket digitális kép formájában szereztük be vagy dokumentáltuk azokban a közgyűjteményekben, ahol a bennünket érdeklő verseket tartalmazó kötetek fellelhetők. Az egyes versek feldolgozása a következő lépésekből állt: (a) betűhív átírás, (b) normalizált (héberbetűs vokalizált) szöveg előállítás, (c) értelmező nyersfordítás, (d) műfordítás, (e) észrevételek, kommentárok fűzése a költeményhez. A folyamat a valóságban természetesen nem egyirányú, nem lineáris volt, hiszen a szöveg elemzése visszahatott annak értelmezésére, egy új értelmezés pedig esetleg az eredetileg pontozatlan héber szöveg újravokalizálását is maga után vonhatta. Sőt az is előfordulhatott, hogy egy később feldolgozott vers készítette a kollégát egy korábban már feldolgozott vers újragondolására.

Szinte evidens a 21. század első harmadában, hogy ezt a folyamatot digitális képek alapján, szövegszerkesztőt felhasználva hajtjuk végre.²² A digitális bölcsészet azonban módszertani önreflexióra szólít fel. Például, vajon milyen előnyei és milyen veszélyei vannak annak, hogy jellemzően digitális képekből dolgozunk, nem pedig az eredeti nyomtatvány fölé görnyedünk? Extrém esetben a felvétel minősége, felbontása, megvilágítása, hibái – de akár a képfájl megsérülése is a többszörös továbbítás, tárolás során – befolyásolhatja a projekt eredményeit: a publikált szöveget és a kutatási kérdésekre adott válaszokat. Egy másik „filozofikus” kérdés: vajon a 17. századi nyomtatott héber betűk mennyire feleltethetők meg a „betűhív” átírás Unicode karaktereinek? Az átírás abból az előfeltevésből

²² A digitális képek átírása során, a projekt elején próbálkoztunk OCR (*Optical Character Recognition*, automatikus karakterfelismerés) szoftverekkel. A tapasztalataink rosszak voltak, a kipróbált programok túl sok hibát ejtettek az átírás során, és nem tűnt hatékonyak ez a módszer. Időközben sokat fejlődött a pre-modern héber (nyomtatott és kéziratos) karakterfelismerés, a digitális judaisztika meghatározó szakértői az eScriptorium (<https://msia.escriptorium.fr/>) használatát javasolják 2023 nyarán. De nem biztos, hogy a szoftver megtanulása, a képek beolvasása, majd a még mindig nem tökéletes átírás javítása hatékonyabb megoldás lenne, mint a szövegek kézzel történő begépelése (és ezáltal azok beható megismerése). Továbbá, 2023 nyarán is még csak a héber betűk felismerésére képes a technológia, a magánhangzójelek automatikus átírása még messze van (Daniel Stökl Ben Ezra, személyes kommunikáció).

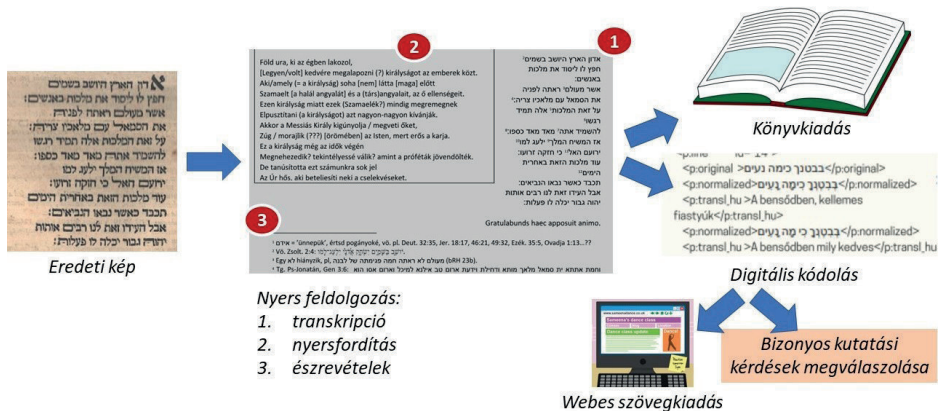
indul ki, hogy a héber ábécé 22 absztrakt „grafémát” tartalmaz, legyen szó a 17. vagy a 21. századról, és eltekinthetünk az „allográfoktól”: szempontunkból irreleváns a betűméret, a karakterek közötti távolság, a betű pontos alakja és így tovább.²³ Aki ezekre esetleg a jövőben kíváncsi lesz, az tekintse meg az átíráshoz mellékelt képet! A karakterek közötti távolság egy dichotóm kérdésre redukálható: van vagy nincs ott szóköz? Eltérő döntéseket hoztunk volna azonban a feldolgozás módszerét illetően, ha a projektünkhöz egy nyomdatörténész is csatlakozott volna.

Hagyományosan ezután következik a feldolgozott forrásanyag, esetünkben a köszöntőversek (kommentált) kiadása, valamint ezek elemzése tudományos tanulmányokban. Utóbbiból egy csokrot tart az olvasó éppen a kezében, egyéb tanulmányok máshol jelentek meg (lásd e kötet bevezetőjében), és remélhetőleg további tanulmányok is meg fognak jelenni a jövőben. A költemények teljes kiadása szintén terveink közt szerepel.

A 21. században ez a munkafolyamat kiegészül két további fázissal (lásd az 1. ábrát). Az első fázisban a szöveget digitálisan átkódoljuk számítógép által is olvasható (*machine-readable*) formába annak érdekében, hogy a második fázisban könnyebben lehessen azzal hasznos dolgokat végezni. Ha a verseket egy weboldalon megjelenítjük, akkor a felhasználók interaktívan férhetnek hozzá a feltöltött adatokhoz, könnyebben kereshetnek a szövegben, és olyan funkciókkal is élhetnek, amelyek egy hagyományos nyomtatott (vagy pdf vagy epub formátumban, e-könyvként kiadott) kötetben nem állnak a rendelkezésére. Ha a költeményeket egy számítógépes korpuszként tudjuk kezelni, akkor a kortárs digitális irodalomtudomány eszközeit (távoli olvasás, *stilometria* stb.) is alkalmazhatjuk. Újszerű kutatási kérdésekre újszerű eszközökkel adhatunk válaszokat,

²³ A fonológiából ismert *fonéma* és *allofón* fogalmakra utalok. A magyarban a /t/ fonéma hehezetes [t^h] és hehezet nélküli [t] allofónjai közti különbséget észre se vesszük, egyes nyelvekben ezek viszont éppúgy eltérő fonémák, mint magyarban a /t/ és a /d/. Hasonlóképpen, egy <o> betű „sarkossága” latinbetűs írás esetén érdektelen, ellentétben a héber ו és ן közti különbséggel. A betű magassága számít a latin <o> és <O> esetén, a héber ו ,ׁ és ׀ esetén, de nem számít a latin <t> vagy a héber ם esetén.

amint azt Götz Andrea is bemutatja jelen kötetben.²⁴ Végül, ha a költeményeket egy standard kódolás segítségével egy nagyobb korpusz részévé tesszük, akkor mások nagyobb volumenű kutatásaihoz is hozzájárulunk. A projektünk során feldolgozott költeményeknek ott a helyük mind a régi magyar irodalom, mind a középkori héber nyelven született irodalom digitális korpuszában.



1. ábra: A közzétöverserk feldolgozásának a folyamata.

A digitális feldolgozás nem fogja felváltani, hanem kiegészíti a hagyományos szövegkiadási és publikálási formákat. A papíralapú könyvek legnagyobb előnye és egyben legnagyobb hátránya is az, hogy ezek kézzel fogható fizikai objektumok. Az alexandriai könyvtár leégése óta mindannyian tudjuk, hogy ha tűzvész, árvíz, háború vagy gondatlan állománykezelés miatt a példányok megsemmisülnek, elvesznek, ismeretlen helyre kerülnek, akkor a könyvekben felhalmozott tudás is elvész. Ezzel szemben a digitálisan őrzött adatok másolása, terjesztése, tárolása (tetszőleges számú biztonsági másolatban), bár nem ingyenes,²⁵ de sokkal olcsóbb, mint a kézzel vagy nyomdagéppel sokszorozott, klimatizált könyvtári

²⁴ Götz 2023.

²⁵ Az adatokat tároló hardverek, (saját vagy „felhőbeli”) szerverek, valamint a hozzáférést biztosító eszközök bekerülési és működtetési költségei nem elhanyagolhatóak.

helyiségekben őrzött, évszázadonként restaurálásra szoruló, papíralapú kiadványoké.

Ugyanakkor a megőrzött fizikai tárgy évszázadok vagy évezredek múlva is fizikai tárgy marad – mint azt épp az általunk feldolgozott kötetek illusztrálják. A digitálisan őrzött adatokat viszonylagos rendszerességgel új adathordozóra szükséges átmásolni, mert a jelenlegi adathordozóink élettartalma legfeljebb néhány évtized. Több évezredes papirusz- és pálmalevélfeliratokat olvasunk, de a 20. század elejéről származó gramofonfelvételek vagy a 20. század második feléből származó mágnesszalagok kiolvasása már nehézségeket okoz. A két-három évtizeddel ezelőtti fájlformátumokat már nem biztos, hogy meg tudja nyitni a gépünk. A 2020-as években kifejlesztett webportálokat vajon a 2030-as, 2040-es évek böngészői képesek lesznek kezelni? Vagy egy másik kérdés: örököseim remélhetőleg a könyvtáramat elszállítják valahová, de vajon tudni fogják, hogy a nevemen lévő *domaint* mikor kell meghosszabbítani?

Összefoglalva, úgy vélem, hogy jelenleg a digitális adatok hosszútávú fenntarthatósága bizonytalanabbnak tűnik, mint a fizikai tárgyaké. Hasonlóan ahhoz, ahogy dokumentumok is leginkább könyvtárakban, levéltárakban maradtak fenn kutatható formában, úgy a digitális teret tekintve is jobban bízom a központi – állami vagy ahhoz hasonló – projektekben (adatbázisokban, repozitóriumokban stb.), amelyek stabil szakmai háttérrel rendelkeznek, mint a magánszemélyekhez vagy üzleti vállalkozásokhoz köthető webportálokból.

MEGJELENÍTENDŐ INFORMÁCIÓ A DIGITALIZÁCIÓ SORÁN

Az eddigiekből is kiderült, hogy a szövegek feldolgozása során számos kérdést kellett – reflektált formában vagy ösztönösen – eldönteni. Így például a karakterek közti távolságot, mint információt, nem mértük le, nem tüntettük fel sehol, azt „kidobásra ítéltük”. Amennyiben valaki új olvasatot javasol, és kétségbe vonja a döntésünket valamelyik szóközről (vagy annak hiányáról), kénytelen lesz visz-

szamenni az eredeti nyomtatványhoz (mint fizikai tárgyhoz) vagy az arról készült digitális képhez. Nyomtatott szövegekről lévén szó, döntésünk védhetőbb, mint héberbetűs kéziratok esetén. Egy új olvasat valószínűleg nem arra fog hivatkozni, hogy a két betű közti hézag akár szóköznek is értelmezhető, hanem arra, hogy a nyomdász valószínűleg kifejejtette a szóközt. Fontos hangsúlyozni, hogy a projektünkben feldolgozott költemények egyetlen kiadásban, kis példányszámban jelentek meg, így a valószínűleg nem elhanyagolható számú nyomdahibát nem javították egy második kiadásban.

A döntéseket néha elvi alapon hoztuk meg. Például azt, hogy a normalizált szöveg a tüberiási héber standard nyelvtanát tükröző, vokalizált szöveg legyen, hiszen ez az a formája a héber nyelvnek, amelyet a költemények szerzői és címzettjei maguk is referenciának tekintettek. A betűhív átírás ezzel szemben megőrzi az esetleges nyomdahibákat, és azt is, hogy a szöveg néha pontozott, néha pontozatlan, sőt alkalmanként latinbetűs átírásban jelenik meg a nyomtatványokban.²⁶ A betűhív átírás tükrözi a nyelvi-nyomdászati produktumot, míg a normalizált szöveg az irodalmi teljesítményt. Különböző kutatók másra és másra lesznek kíváncsiak, ezért mindkettőnek van létjogosultsága.

Más esetekben, mint bármely kutatás esetén, a döntések egy része reflektálatlan maradt, és későbbi korok kutatói minden bizonnyal szóvá fogják tenni. Éppen ezek a módszertani kritikák hajtják előre a tudományt. A véletlen és a realitások alapján meghozott döntések mellett az ökonómiai szempontok is jelentősen befolyásolnak egy ilyen projektet. Például nagyon sok szempontból lehetne annotálni egy korpuszt: a szavak morfológiai sajátosságait feltüntetve, a bibliai utalásokat feltüntetve, a nyomdahibákat feltüntetve stb. Azonban mindaddig felesleges ebbe emberi erőforrást befektetni, amíg nincs olyan kutatási kérdés, amely ezt az investíciót indokolttá teszi. Egy potenciális jövőbeli kutató talán nem pontosan úgy annotálná a korpuszt, mint ahogy azt mi tennénk, jóeséllyel a befektetett munkánkat úgyis felülvizsgálná.²⁷

²⁶ Például Becskehazi István verse Székelyhidi Istvánnak (1651, RMNY III 2356).

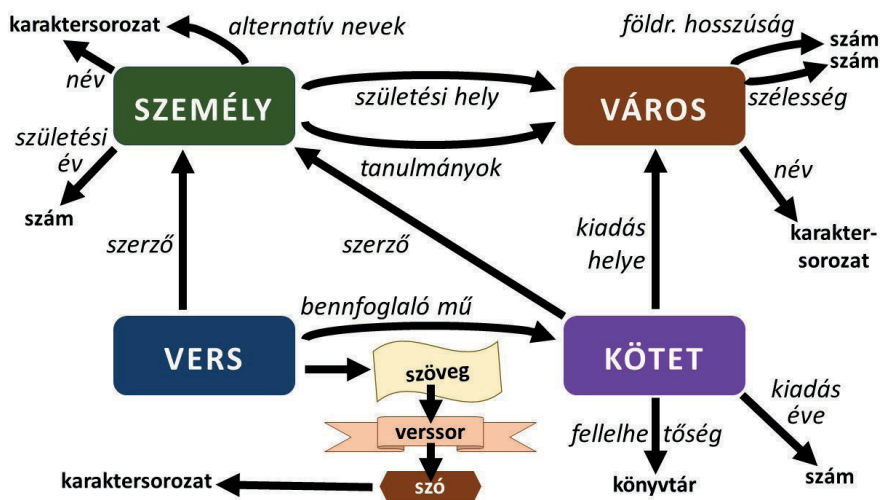
²⁷ Lásd még Götz 2023 a jelen kötetben.

Most tekintsük át a ténylegesen rendelkezésünkre álló adatokat. Az informatika az adatok feldolgozásának a tudománya. Milyen adattípusok jelennek meg a projektben? A nyomtatványokról készült képfájlok, valamint szövegek, számok, földrajzi helyek, személyek...

A számok egyszerű adattípusok. Számok lehetnek például évszámok, oldal-számok és földrajzi koordináták is. Szöveges adatformában jelennek meg elsősorban a költemények: különböző nyelveken, különböző írásrendszerekkel, de fontos, hogy verssorokba (és sok esetben versszakokba) szerkesztett formában. Rövidebb szöveges adattípust, karaktersorozatot használnak a személynevek, városnevek és a kötetek címei. Csavar viszont az, hogy a személyneveket, különösen pedig a városneveket mind magyarul, mind angolul a felhasználóink rendelkezésére szeretnénk bocsátani, de egyúttal releváns lehet a korabeli latin és német/holland elnevezés is, és ezek nemritkán többféle alternatív írásmóddal is léteztek. Elvileg itt is felmerülhetnek átírási kérdések: például egységesítsük, vagy őrizzük meg a korabeli különbséget a <s> és a <f> karakterek között?

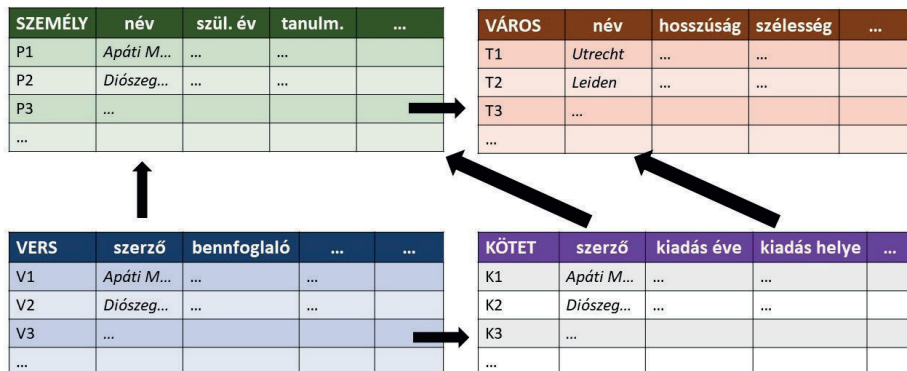
A kiadás helye, illetve szerzőink és címzettjeink életrajzában megjelenő fontos városok, országok, térségek, egyéb helyszínek (a származás és a születés helye, a tanulmányok helyszínei, későbbi állomáshelyek stb.) szintén önálló adattípusnak tekinthetők. Ezekhez név és földrajzi koordináták tartoznak. A nevek megjelenhetnek a mai felhasználók részére magyarul és angolul, illetve különböző latin, német, holland korabeli írásmódokkal. A földrajzi koordináták lehetővé teszik a helyszínek megjelenítését térképeken, ami nem csupán egy látványos funkció a weboldalon, hanem a geolokáció elemzése a kortárs digitális bölcsészetben is egyre jelentősebb szerepet tölt be. Felvetődhet a földrajzi helyek ontológiába történő szervezése, például a városok hozzárendelése egy tartományhoz, régióhoz, országhoz. Tekintve azonban azt, hogy a kutatásunk során érintett városok száma nem túl nagy, ezek többsége jól ismert, viszont 17. századi történelmük fordulatos, ezért nem láttuk értelmét egy ilyen fejlesztésnek (vö. a korábban említett ökonómiai szempontokkal a döntések meghozatala során).

A projektünk fogalmi elemzése végeredményben a 2. ábrán látható adatmodellt eredményezte, és erre épül a projekt eredményeit bemutató portál is (<https://cgh.hu/>).



2. ábra: A Héber köszöntőversek-projekt adatmodellje.

Következő lépésben létrehozunk egy relációs adatbázist négy táblázattal (3. ábra). Az egyikben a személy, a másikban a városok, a harmadikban a költemények, és a negyedikben a kötetek találhatóak. Mindegyik személyhez megadjuk a nevét, az alternatív neveit, a születési évét, azt a várost, amelyben tanult, és így tovább. A város, amelyben tanult, egyben átvezet a városokat tartalmazó táblázathoz, ahol megtalálhatók például a város földrajzi koordinátái. A költeményeket tartalmazó táblázatban szintén feltüntetjük a szerzőt, valamint azt a kötetet is, amelyben ez a költemény megtalálható. A költemény címet úgy deríthetjük ki, ha előbb megnézzük a költemény bennfoglaló kötetét, majd ezt a kötetet kikeressük a köteteket tartalmazó táblázatból, és megnézzük annak a szerzőjét.



3. ábra: Relációs adatbázis a személyek, a kötetmények, a városok és a kötetek adataival.

Ezt követően megfogalmazhatunk kereséseket (a MySQL adatbáziskezelő nyelvet idéző pszeudokódban):

- (1) **SELECT** volume **FROM** volumes **WHERE**
volume.place_of_publication.name == 'Leiden'
- (2) **SELECT** person.name, person.alternative_names **FROM**
persons **WHERE** person.studied_in.lattitude < 50.5

Az (1) utasítás azt kéri az adatbázistól, hogy gyűjtse nekünk össze azokat a köteteket a köteteket tartalmazó táblázatból, amelyekre igaz az, hogy a kötet publikációs helyének a neve (itt már áttérünk a városokat tartalmazó táblázatra) egyenlő azzal a karaktersorozattal, hogy „Leiden”. A (2) utasítás pedig azokat a személyeket gyűjti össze a személyeket tartalmazó táblázatból, akikre igaz az, hogy olyan városban tanultak (itt ismét áttérünk a városokat tartalmazó táblázatra), amelynek földrajzi szélessége kisebb, mint $50,5^\circ$, és ezen személyek kapcsán kérjük mind a nevüket, mind a nevük lehetséges alternatív változatait.

A figyelmes olvasóban ezen a ponton számos kérdés merülhet fel: például mi történik azokkal a peregrinusokkal, akik több városban is tanultak? Ezeknek a

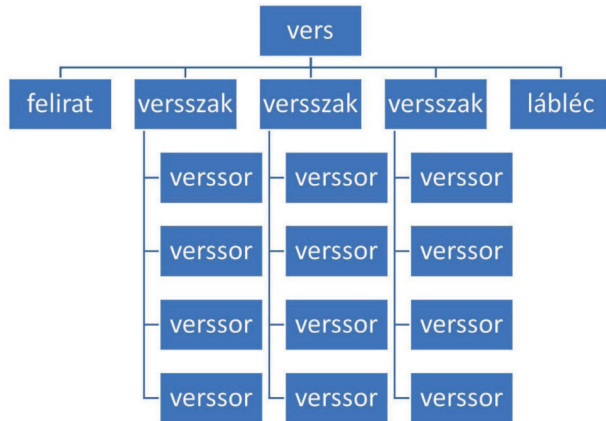
kérdéseknek a megoldása az implementáció részleteitől függ. A projektünkben egyelőre nem okozott gondot, de más esetekben gondot okozhat a bizonytalan és a hiányos információ: hogyan jelöljük azt, ha például a kötet megjelentetésének az évszáma nem áll rendelkezésünkre, vagy ha a szerző születésére csak egy intervallumot tudunk megadni? Mindezen helyzetek kezelésére vonatkozóan az adatbázisok informatikai elméletének számos megoldási javaslata van.

Annak érdekében, hogy a tőlünk független és jövőbeli kutatásokkal össze lehessen kapcsolni a miénket, fontos az adataink összekapcsolása szabványos referenciapontokkal. A személyek esetén például az egyelőre még csak fejlesztés alatt álló *Nemzeti Névtér*²⁸ fog várhatóan ilyen funkciót betölteni a következő évtizedekben, míg a kötetek esetén az RMK-beli azonosító tűnik hasonló standard hivatkozásnak.

PEREGRXML: A SZÖVEG MINT HIERARCHIKUS STRUKTÚRA

A köszöntőversek struktúrája kötött: a betűk szavakba, a szavak sorokba, a sorok sok esetben versszakokba szerveződnek, továbbá a vershez legtöbbször tartozik még egy felirat és egy lábléc is (4. ábra). Sok esetben a verssorok két félsorra bomlanak. A prózai szövegekkel ellentétben, a szöveg tagmondatokra (mondat egységekre) és mondatokra való tagolása kisebb szerepet játszik.

²⁸ <https://magyarnemzetinevter.hu/>.



4. ábra: Egy költemény hierarchikus struktúrája.

A hierarchikus struktúrákat napjainkban az XML (eXtensible Markup Language, „kiterjeszhető jelölőnyelv”) szabvány segítségével lehet leghatékonyabban kódolni, például:

```

<poem book="RMK III 2452" author="p09" addressee="p17">
  <stanza id="1">
    <line id="11">Első versszak első sora</line>
    <line id="12">Második sora</line>
    <line id="13">Harmadik sora</line>
  <line id="14">Negyedik sora</line>
  </stanza>
  <stanza id="2">
    <line id="21">Második versszak első sora</line>
    <line id="22">Második sora</line>
    <line id="23">Harmadik sora</line>
    <line id="24">Negyedik sora</line>
  </stanza>
</poem>

```

Ebben a PeregrXML-nek elnevezett rendszerben a költeményt a <poem> és a </poem> tagek veszik körül. Azon belül a két versszakot a <stanza> és a </stanza> tagek jelölik, ezen belül találhatóak a verssorok, a <line> és a </line> tagekkel körbevéve. Az egyes szinteken további információkat is kódolhatunk. Így adhatjuk át a számítógépnek azt a messze nem evidens struktúrát, amelyet mi, az európai kultúrában nevelkedett emberként, a költeményre ránézve automatikusan észreveszünk. Az XML szabvány használatának sok előnye van: rugalmasan bővíthető, miközben bizonyos szigorú szabványok a programok számára könnyen kezelhetővé teszik.

A bővítésnek két iránya van. Egyrészt, szeretnénk visszaadni a szöveg feldolgozási folyamatának a fázisait: a betűhív átíratot, a normalizált szöveget és a fordításokat is. Másrészt, lehetőséget kell teremtenünk az eltérő olvasatok, a bizonytalanság, a szerző vagy a nyomdász által elkövetett hibák jelzésére is. Egy lehetséges megvalósítást mutat be az 5. ábra. A projekt korai fázisában Diószegi Kalmár Pálnak Báthori Mihályhoz címzett költeményében (1653, Leiden, RMK III. 1840) egy kifejezés több lehetséges olvasata között bizonytalankodtunk (14. sor), megengedve azt, hogy egy szóköz kimaradt a nyomtatásnál. Ezért az XML-kód két normalizált (vokalizált) szöveget, és ennek megfelelően két fordítást is tartalmazott.

```
<?xml version="1.0" encoding="UTF-8"?>
<p:poem xmlns:p="peregr.kre.hu">
  <p:location book_id="Leiden5_RMK_III_1840" date="1653" />
  <p:photo src="img5.jpg" />
  <p:photo src="img7.jpg" />
  <p:author id="dkp" name="Diószegi Kalmár Pál" />
  <p:addressee id="bm" name="Báthori Mihály" />
  ...
  <p:line id="13">
    <p:original>שׁמַע דְּבַר יְיָ אֲזַנְךָ אֶת אִמְרֵי />
    <p:normalized>שמע דבר יי' א'ז'ן א'ת א'מ'ר'י />
    <p:transl_en>Turn your ear, listen to my words</p:transl_en>
  </p:line>
  <p:line id="14">
    <p:original>בְּבִטְנְךָ כִּי טַח כִּימֵה נְעִים />
    <p:normalized>ב'ב'י'ט'נ'ך' כ'י' ט'ח כ'י'מ'ה נ'ע'י'ם />
    <p:transl_en>In your belly, the pleasant Pleiades</p:transl_en>
    <p:normalized>ב'ב'י'ט'נ'ך' כ'י' ט'ח כ'י'מ'ה נ'ע'י'ם />
    <p:transl_en>In your belly, for how pleasant it is</p:transl_en>
  </p:line>
  <p:line id="15">
    <p:original>לְרֹשְׁתְּךָ יוֹם שְׁלֹשׁוֹת />
    <p:normalized>ל'ר'ש'ת'ך' יו'ם ש'ל'ו'ש'ו'ת />
    <p:transl_en>To return to you the day of retribution.</p:transl_en>
  </p:line>
  ...
</p:poem>
```

5. ábra: Részlet egy PeregrXML-ben kódolt versből.

Időközben megjelent és a digitális bölcsészek körében rendkívül elterjedt a TEI-szabvány (*Text Encoding Initiative*),²⁹ amely lehetővé teszi azt, hogy az immár feldolgozott, kevesebb bizonytalanságot tartalmazó szövegeinket egy mások által is ismert formában bocsássuk a hazai és nemzetközi kutatók, a következő generációk rendelkezésére.

ÖSSZEFOGLALÁS

Tanulmányom célja az volt, hogy áttekintsem a digitális bölcsészet hozzájárulását a 17. századi héber köszöntőverseket feldolgozó projektünkhöz. Önmagában a nyomtatványokról készült, jó minőségű digitális képek elkészítése és a versek átgépelése is tekinthető már digitális bölcsészetnek. De a 21. századi digitális bölcsészet több szempontból is meghaladja a 20. századi bölcsészettudományi informatikát. Egyrészt, tudatos önreflexióra kényszerít bennünket, így tanulmányomban is felvettem néhány „filozofikus” kérdést a projekt során – tudatosan vagy ösztönösen – meghozott döntésekkel kapcsolatban. Másrészt, a korábbiaknál komplexebb információhalmaz feldolgozását teszi lehetővé. Valóban, a projektünkben is találkozunk a képi információ (a nyomtatványokról készült felvételek) a szöveges, számszerű és földrajzi információval. Ezekből kereshető relációs adatbázisokat hozunk létre, és hierarchikusan strukturált szövegeket bocsátunk a jövőbeli kutatók rendelkezésére. A projekt digitális bölcsészeti szempontú elemzése vezetett el ahhoz az adatmodellhez is, amely a projekt eredményeit bemutató honlap (<https://cgh.hu/>) szerkezetét megadja.

Hosszú út áll még előttünk ahhoz, hogy a középkori keresztény héber nyelven magyar szerzők által írt költeményeket egyrészt a régi magyar irodalom, másrészt a kora újkorai héber irodalom szerves részeként kezeljék a művelődéstörténet,

²⁹ <https://tei-c.org/>. A TEI szabvány szerint kódolt szövegeken végzett filológiai munkát a *Versioning Machine* (<http://v-machine.org/>) teszi különösen praktikussá. A projektünkben egyelőre nem merült fel ilyen jellegű igény, de a jövőben, új típusú kutatási kérdések kapcsán, hasznos lehet kipróbálni ezt a szoftvert.

kultúratörténet, irodalomtörténet, nyelvtörténet kutatói. „Szerves rész” alatt itt a digitális bölcsész arra gondol, hogy egy digitális korpuszba lettek integrálva: ha az adott korpusz szabványai szerint kódoljuk az általunk vizsgált költeményeket, akkor biztosítva lesz azokhoz a hosszú távú hozzáférés, a kereshetőség. Innentől kezdve az is rá fog találni a 17. századi magyar peregrinusok héber nyelvű költeményeire, aki nem kimondottan ezeket keresi. Mostantól a jövő kutatógenerációi az éppen kurrens módszertanokat fogják tudni alkalmazni arra a tágabb korpuszra, amelyeknek ezek a köszöntőversek is részei.

FELHASZNÁLT IRODALOM

- Biró 2018 Biró Tamás, „Early Modern Christian Hebrew: A hitherto overlooked variety of Medieval Hebrew”, konferenciaelőadás: *XIth Congress of the European Association for Jewish Studies* (EAJS 2018), Kraków, July 19, 2018.
- Biró 2019 Biró Tamás, „»...milyen jó dolog magyar zsidónak lenni, de ez mar egy másik vita témaja...« Lojalításhierarchiák egy késő 20. századi digitális korpuszban”, konferenciaelőadás: *Zsidó lojalításhierarchiák dinamikája a modernkori Közép-Európában*, SZTE BTK, Szeged, 2019. november 7.
- Biró 2020 Biró Tamás, „»...milyen jó dolog magyar zsidónak lenni, de ez mar egy másik vita témaja...« Kutatási lehetőségek egy késő 20. századi digitális korpuszban”, konferenciaelőadás: *KDFI A magyar internet története*, BME, Budapest, 2020. február 14.
- Czachesz 2016 Czachesz István, „Network Analysis of Biblical Texts”, *Journal of Cognitive Historiography* 3/1–2 (2016) 43–67.
- Fantoli 2013 Margherita Fantoli, „Steven E. Jones, Roberto Busa, S. J., *And The Emergence Of Humanities Computing: The Priest And The Punched Cards*, Routledge, London – New York 2016, 196 Pp., ISBN:

- 9781315643618”, *Mediterranea: International Journal on the Transfer of Knowledge* 8 (2023) 673–681.
- Götz 2023 Götz Andrea, „What can Hebrew poems tell us about 17th-century Hungarian peregrination? Using metadata and frequency analyses for sociopragmatic research” in Koltai K. – Zsengel-lér J. (szerk.), *שיר פשוט ומחלוק, Héber üdvözlőversek és a peregrináció* (Studia Hebraica Hungarica 9), Budapest: MHT – L’Harmattan, 2023, 153–175.
- Jurafsky–Martin 2008 Daniel Jurafsky – James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (2. kiadás), Upper Saddle River, NJ: Pearson – Prentice Hall, 2008.
- Kiss 2019 Kiss Margit, „Stilometriai elemzés lehetőségei magyar történeti szövegtörzseken”, *Digitális Bölcsészlet* 2 (2019) T:15–T:33.
- Koltai 2021 Koltai Kornélia: „Bibliai és posztbiblikus nyelvi rétegek dialógusa (egymásra gyakorolt hatása) a magyar peregrinusok 17. századi héber nyelvű üdvözlőverseiben”, in Farmati A. – Gábor Cs. (szerk.), *A dialógus formái a magyar régiségben*, Kolozsvár: Egyetemi Műhely Kiadó, 2021, 455–476. (<http://real-eod.mtak.hu/11518/2/FarmatiGaborEgyFuzcc-5IMP.pdf>)
- Kopel 2014 Moshe Kopel, “The Responsa Project: Some Promising Future Directions”, in N. Dershowitz – E. Nissan (eds.), *Language, Culture, Computation: Computing for the Humanities, Law, and Narratives. Essays Dedicated to Yaacov Choueka on the Occasion of His 75 Birthday*, Part II. Berlin – Heidelberg: Springer, 2014.
- Ladon 2022 Joshua S. Ladon, *The Mediated Torah: Source Sheets and Jewish Knowledge Construction in the Digital Age*, Doktori disszertáció, Jewish Theological Seminary of America, 2022.

- Manning–Schütze 2000 Christopher D. Manning – Hinrich Schütze, *Foundations of Statistical Natural Language Processing* (2. jav. kiadás), Cambridge, MA: MIT Press, 2000.
- Nemes 1933 Nemes Zoltán, „A magyar parlamenti nyelv leggyakoribb szavai”, *Acta litterarum ac scientiarum Regiae Universitatis Francisco-Josephinae: sectio philologico-historica* 5 (1933) 17–47.
- Péter 2016 Péter Róbert, „A big data kihívás és lehetőség a bölcsészettudományokban: digitális szövegek és metaadatok távoli olvasása”, *Magyar Tudomány* 177/11 (2016) 1323–1330.
- Postma–Van Sluis 1995 Ferenc Postma – Jacob van Sluis (szerk.), *Auditorium Academiae Franekerensis. – Bibliographie der Reden, Disputationen und Gelegenheitsdruckwerke der Universität und des Athenäum in Franeker 1585–1843* (FA: Minsken en Boeken 23), Leeuwarden/ Ljouwert: Frysk Akademy, 1995.
- Schreibman et al. 2004 Susan Schreibman, Raymond George Siemens és John Unsworth, *A Companion to Digital Humanities*, Malden, MA – Oxford, UK – Victoria, Aus: Blackwell, 2004.
- Schreibman et al. 2016 Susan Schreibman, Raymond George Siemens és John Unsworth, *A New Companion to Digital Humanities*, Malden, MA – Oxford, UK – West Sussex, Aus: Blackwell, 2016.
- Zsengellér 2023 Zsengellér József „Miért éppen a 17. század?” in Koltai K. – Zsengellér J. (szerk.), *שיר פשוט ומחלוק, Héber üdvözlőversek és a peregrináció* (Studia Hebraica Hungarica 9), Budapest: MHT – L’Harmattan, 2023, 31–48.