

A MAGYAR ORVOSI NYELVI KORPUSZ 16–17. SZÁZADI ALKORPUSZA: TERVEZET, ÁTÍRÁS, ANNOTÁLÁS¹

KUNA ÁGNES, KOCSIS ZSUZSANNA, LUDÁNYI ZSÓFIA
(MTA Támogatott Kutatócsoportok Irodája, Pázmány Péter Katolikus Egyetem, Károli
Gáspár Református Egyetem – Eötvös Loránd Tudományegyetem –
MTA Nyelvtudományi Intézet)

„A múlt írásai hozzásegítenek ahhoz, hogy jobban megértsük
a jelenkor tudományos nyelvének jellemzőit.”
(Taavitsainen–Pahta 2004: XVI)

„[...] használd a múltat a jelen magyarázatához,
és használd a jelent a múlt eléréséhez.”
(Romaine 1988: 1452)

1. Bevezető gondolatok – a magyar orvosi nyelv fejlődése mint folyamat

Az egészség megőrzése, valamint a gyógyítás az emberi kultúra szerves részét képezi a kezdetektől fogva. Az egyes korokban és kultúrkörökben a betegségekről különböző elképzelések éltek, és ez a fogalmi és nyelvi létrehozásban is nyomon követhető. Mindezt alapvetően meghatározza a testről való tudásunk, a hit- és hiedelemvilág, az orvoslás tudományos nézetei, az intézményesítés foka, tehát a vizsgált időszak és a gyógyítás társas-kulturális tényezői. Az egyes korok gyógyítói gyakorlata a használt eszközökön túl elsősorban írásos dokumentumok révén őrződött és őrződik meg. A medicina történetében különböző jellegzetes szövegtípusok alakultak ki az évszázadok folyamán a társadalmi és kommunikatív igények mentén, amelyek hozzáférést biztosítanak az orvosi nyelv fejlődésének vizsgálatához.

Tanulmányunkban az orvosi nyelvet alakulástörténetként kezeljük (vö. Csontos 2013). Ennek megfelelően a történeti és a jelenkori orvosi szaknyelvet, illetve a gyógyításhoz kapcsolódó kommunikációt egyazon dinamikus folyamat részének tekintjük (vö. Kuna 2011, 2012, 2015). Továbbá abból indulunk ki, hogy az orvosi nyelv története megragadható a gyógyításhoz kapcsolódó szövegek és szövegtípusok történeteként (vö. Steger 1984; Barz–Fix–Schröder–Schuppner 2000; Maitz–Molnár 2001), és a szaknyelv szisztematikus vizsgálatát nagyban elősegíti, ha rendelkezésre áll egy módszeresen felépített korpusz. Ezzel összefüggésben célunk, hogy ismertessük a készülőben lévő Magyar orvosi nyelvi korpusz (MOKO) tervezetét, és azon belül is elsősorban a magyar nyelvű írásos orvosi munkák kezdeti szakaszára vonatkozó munkafolyamatokat. Ennek megfelelően ismertetjük a MOKO alapkoncepcióját és a korpusz várható hozadékait;

¹ A tanulmány az MTA TKI és a PPKE ITK posztdoktori kutatóprogram támogatásával készült.

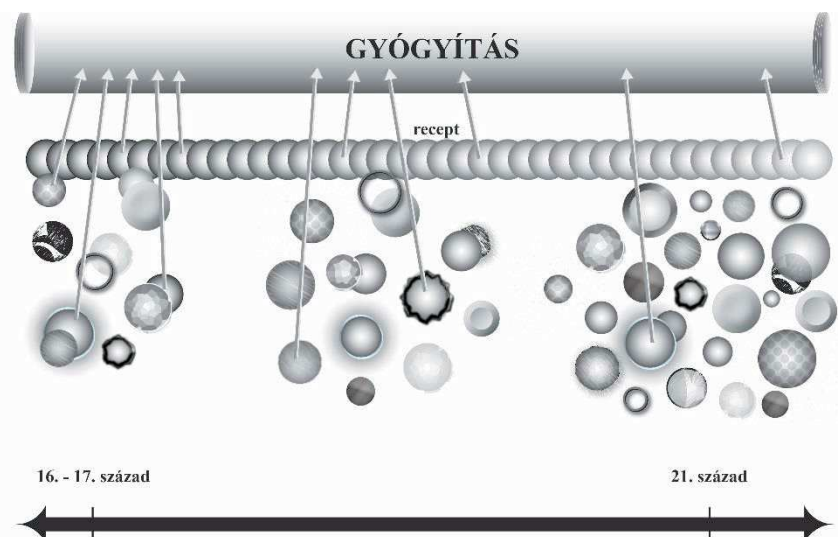
majd részletesen bemutatjuk a 16–17. századi alkorpusz munkafázisait, kitérve: a korpusz építésének főbb alapelveire és átfogó koncepciójára; a szövegátírás és a normalizálás szempontjaira; továbbá a nyelvtechnológiai feldolgozás menetére és terveire.

2. A MOKO alapkonceptiója

A MOKO projektje a magyar nyelvű medicina egészét felöleli.² Az egyes munkafázisok azonban az orvoslás történeti alakulásához, azok korszakaihoz alkalmazkodnak. Így első lépésben a magyar nyelvű orvoslás kezdeti szakaszát, a 16–17. századi dokumentumokat dolgozza fel. A korpusz rendezőelvét az egyes korokban fellelhető reprezentatív szövegtípusok adják. Ezzel az orvosi nyelv történetét az adott korszakban jelen lévő szövegtípusok történetének tekinthetjük, ami lehetővé teszi a szaknyelv fejlődésének dinamikus vizsgálatát, a szaknyelvet alakító társadalmi és kulturális tényezők bevonását. A szövegtípus alkalmas elméleti keretet ad a szaknyelvi korpusz kialakításának, továbbá a szaknyelv szociokulturális komplexitásának vizsgálatának is (Steger 1984; Görlach 2004; Kohonen 2004; Taavitsainen 2001; Kuna 2011). A szövegtípus elméleti keretként való alkalmazásakor a kommunikatív-kognitív megközelítésből indulunk ki, amely szövegtípusok nem absztrakt, önmaguktól és önmagukért létező nyelvészeti kategóriáknak tekint, hanem a társas-kommunikatív igényeket kielégítő megismerési és (egyéni és társas) használati sémáknak (Taavitsainen 2001; Tolcsvai Nagy 2001, 2006; Kocsány 2002, 2006; Kuna 2016a, 2017). Az egyes korok szövegtípus-repertoárjának, illetve a szövegtípusok egymáshoz való kapcsolatának, hálózatának a feltérképezése tehát az orvosi nyelv dinamikus alakulásának, alakulástörténetének a megismerését is lehetővé teszi.

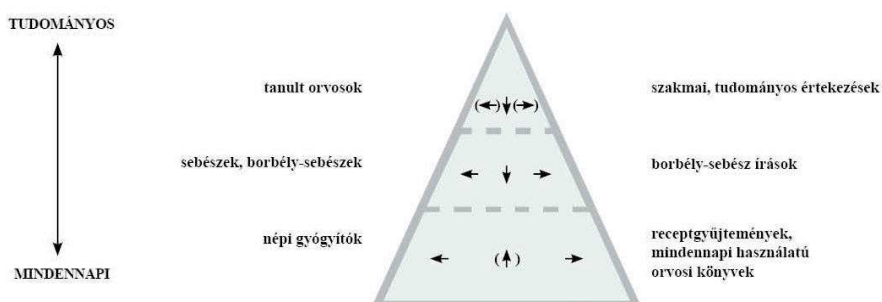
Ebben a folyamatban megfigyelhető, hogy bizonyos szövegtípusok a társas-kommunikatív igények mentén alakulva végigkísérik az orvosi diskurzustartományon belül a gyógyításhoz kapcsolódó kommunikációt (pl. recept), mások eltűnnek, és számos új szövegtípus keletkezik a kommunikatív relevanciának (Heinemann 2000) köszönhetően (pl. beutaló, betegtájékoztató, internetes blogok stb.). Az alakulástörténetben a jelen felé haladva megfigyelhető, hogy az intézményesítés és a tudományosság központi szerepe azt is eredményezi, hogy a gyógyításhoz kapcsolódó kommunikáció egyre komplexebbé válik, ami a szövegtípusok számának megszorodásában is megmutatkozik. Ezt szemlélteti sematikus az 1. ábra.

² A hosszú távú projekthez 2016. július 13-án a PPKE ITK-n tudomány- és intézményközi kutatócsoport alakult Kuna Ágnes vezetésével.



1. ábra: A gyógyításhoz kötődő szövegtípusok időbeli alakulása (Kuna 2015: 33)

Fontos szempont továbbá, hogy a gyógyításban már igen korai időszakról kezdve megfigyelhető a mindennapi és az egyre tudományosabb elvárások megjelenése, és ezzel az orvosi szövegek felhasználói szintek szerinti rétegződése, ami a magyar orvosi nyelv kezdetétől fogva a mai napig megmutatkozik a szövegtípusok és a nyelvhasználók szintjén egyaránt (vö. Kuna 2011, 2013, 2016b). A 16–17. századi orvosi diskurzustartományra vetítve ezt ábrázolja szematikusan a 2. ábra. A MOKO készítésekor az említett komplexitást és rétegzettséget is szükséges figyelembe venni ahhoz, hogy a gyógyítás-hoz kötődő nyelvhasználatról árnyaltabb képet kaphassunk.



2. ábra: A gyógyítók és az orvosi szöveghagyomány rétegzettsége és a kommunikáció iránya a 16–17. században (vö. Kuna 2011: 56)

3. A MOKO I. szakasza – 16–17. század

A MOKO létrehozása hosszú távú célként tűzi ki a magyar nyelvű orvosi irodalom egészének rendszerezett és szakaszos feldolgozását. A korpusz kialakítása jelenleg kezdeti fázisban van, a 16–17. századi receptgyűjtemények, herbáriumok és orvosi könyvek feldolgozása folyik. A magyar nyelvű írásos orvosi irodalom kezdeti szakasza a 16–17. századra tehető, ekkortól maradtak ugyanis fenn nagyobb terjedelmű írásos emlékeink. „[...] A kéziratos és nyomtatott könyvek együttesen jelentik a magyar nyelvű orvosi irodalom kezdeti időszakát, s a nyomtatott, illetve kéziratos formából adódó különbségek kisebb jelentőségűek az azonos témából és előadásmódból, a magyar nyelvből következő egyezéseknél” (Szlatty 1983: 384). Ennek megfelelően mind a kéziratos, mind a nyomtatásban megjelent műveket feldolgozzuk.

Számos korai írás részben vagy egészben hozzáférhető utánnomásban vagy valamilyen feldolgozásban (vö. pl. Varjas 1943; Szabó T. 1979; Szlatty 1983; Hoffmann 1989; Szabályár 1984; Szabó T.–Bíró 2000; S. Sárdi 2002). Több kézirat azonban teljesen kiadatlan. A korpusz létrehozásakor fontos szempont, hogy egységes elvek alapján váljon kutathatóvá a magyar orvosi nyelv írásbeliséggel rendelkező, kezdeti szakasza. A szövegtár épít a már megjelent, kiadott munkákra, így az alább felsorolt kéziratos és nyomtatott művek is szerepelnek benne. (A lista nem teljes, bővíthető, változtatható.)

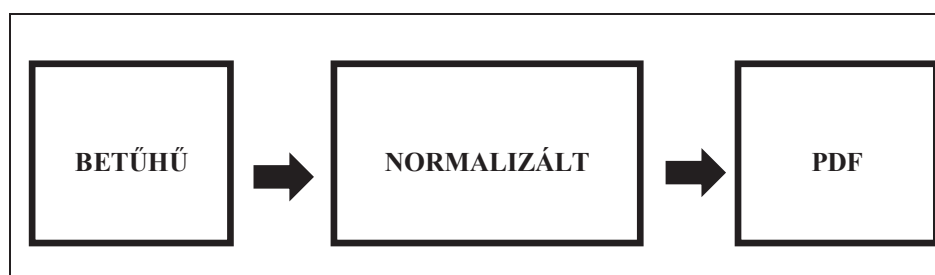
Kéziratok:

Ars Medica 1577 k./1943
 Medicinae Varias 1598/1989
 Egy néhány rendbeli orvosságok 1614–1635 k./1989
 Orvoskönyv lovak orvoslása 1619 e./1989
 Testi orvosságok könyve 1619 k./1989
 Kis patika 1628/1989
 Gyógyszerek I. [17. sz. eleje]/1989
 Gyógyszerek II. [17. sz. eleje]/1989
 Orbis sensualium pictus 1658/1959
 Házi patika 1663 k./1989
 Orvosságos könyveczke 1664 k./1989
 Próbálós bizonyos orvosságok 1666; 1693/1989
 Orvosságos könyv Apafi Anna számára 1677/1989
 Medicusi és borbélyi mesterség 1668–1703/1989
 Pettényi Borbély Márton orvosló könyve 1683–1701/1983
 Próbált orvosságok 1684/1989
 Liber Medicinarum 1693 k./1983
 Orvosságoknak rendszedése [17. sz. 2. fele]/1989
 Phlebotomia [17. sz. 2. fele]/1983
 Orvoskönyv némely füveknek hasznáról [17. sz.]/1989

Nyomatott művek:

Herbárium 1578/1979
 Hasznos és fölötte szükséges könyv 1588/1983
 De conservanda 1694/1983
 Pax Corporis 1695/1984

A 16–17. századi alkorpusz kialakításakor figyelembe vesszük a meglévő magyar nyelvű történeti korpuszok, valamint a nemzetközi orvosi nyelvi szövegtárak kialakítási és átírási elveit is.³ Így elsődleges feladat az eredeti szövegek betűhű, valamint normalizált átíratainak az elkészítése. Ez alapfeltétele a nyelvtechnológiai feldolgozásnak, amelyet egy kéziraton készült próbaelemzésen keresztül ismertetünk. Ezen munkafázisokat követően hozzuk létre a korpusz honlapját és keresőjét, amelyen az egyes kéziratokat PDF formátumban is elérhetővé tesszük annak érdekében, hogy a kutatók teljességében is hozzáférjenek az egyes korai munkákhoz. Így kutathatóvá válik például a helyesírás, az írásjelhasználat, a szövegbeli elrendezés stb.



3. ábra: A kéziratok feldolgozásának lépései

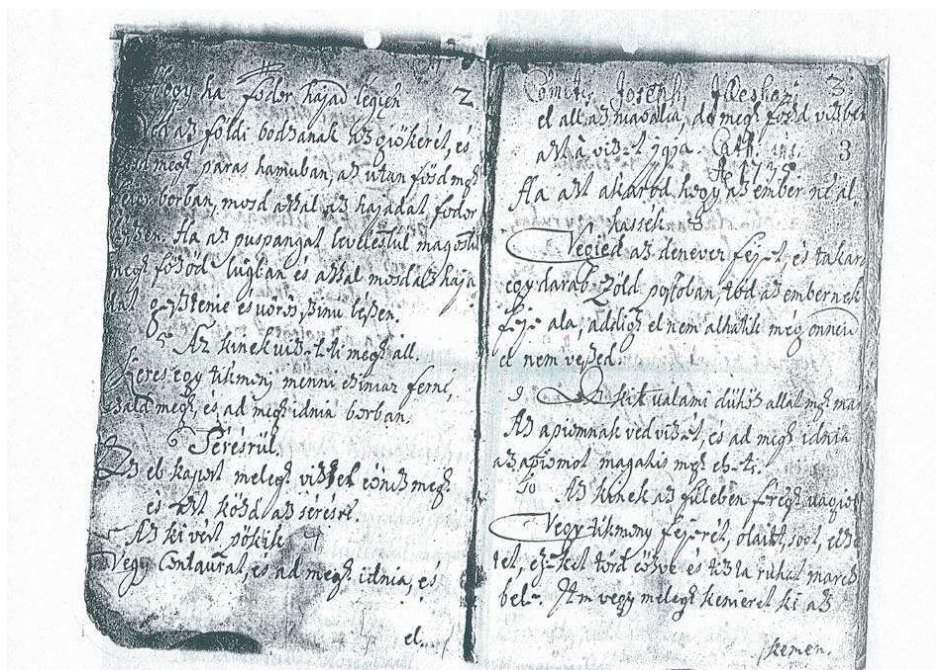
A továbbiakban rátérünk az átírási és a nyelvtechnológiai feldolgozás lépéseire egy kézirat, a Mindenféle orvosságoknak rendszedése kézirat alapján (OSZK Oct. Hung. 87.).

3.1. Az átírás elvei

Az adatbázis építése során az egyik elsődleges feladat az eredeti szövegek digitális átíratainak elkészítése. A feldolgozásra kijelölt szöveg az Orvosságoknak rendszedése címet viselő 17. századi receptgyűjtemény. A kézirat összesen 232 receptet tartalmaz, melyeket kurzív írással jegyeztek le magyar nyelven, helyenként latin szövegrészekkel. A receptek helyesírása kevert típusú, mind betűkapcsolatos (*cz, gh, gi, gy, sz*), mind mellékjelező megoldásokat (*é, ő, ű*) alkalmaz a hangminőségek megkülönböztetésére. Az írásjelek közül a pont és vessző található meg, az előbbi mondat, az utóbbi tagmondat-határoló funkcióban, többnyire következetesen.

A receptek a kéziratban könnyen elkülöníthetők, mivel bekezdéssel vannak egymástól elválasztva, s részben számozottak.

³ Lásd Hivatkozott elektronikus történeti korpuszoknál; Taavitsainen–Pahta 2004; Taavitsainen–Pahta–Mäkinen 2005; Taavitsainen–Pahta et. al. 2010; Voigts–Kurtz 2001; német nyelvű kutatás Riecke 2004; Eckkammer 2005.



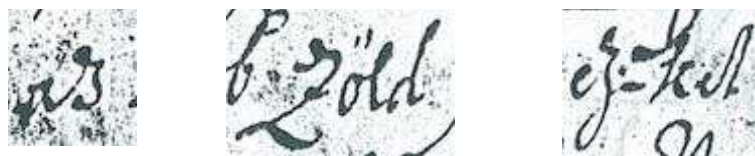
4. ábra: Részlet az eredeti kéziratból

A receptek eredeti számozása azonban nem minden esetben következetes. A kézirat egyes részeiben minden recept külön sorszámot kap, míg más helyeken a számozás nem a receptekhez, hanem a betegségekhez van rendelve, vagyis ha egy kór gyógyítására több recept is van, azok nem kapnak külön-külön számozást. Az átírat készítése során eltértünk az eredeti sorszámazástól, és egy új, a teljes szövegre kiterjedő egységes megoldást vezettünk be, aminek a célja a könnyebb átláthatóság és kezelhetőség volt. A digitális átíratok újraszámozása során minden recept külön sorszámot kapott, függetlenül attól, hogy ugyanazon betegség gyógyítására szolgált vagy sem.

Az eredeti kézirat alapján kétféle digitális átíratot készítettünk, egy betűhű és egy normalizált változatot. Minkét esetben figyelembe vettük a már kialakult szöveghelyi gyakorlatot, s követtük a már meglévő mintákat ott, ahol a nyelvelmékek közti különbség lehetővé tette. Így például az idegen nyelvi kifejezések (!lat!Probatum), valamint a hiányzó mód- és esetjelölések normalizálásában (Végy egy dióhéjat, töltsD meg vargaszurokkal) a Középmagyar történeti magánéleti korpusz megoldásait vettük alapul (Dömötör 2014).

A betűhű átírat létrehozása során a nyelvészeti forráskiadás gyakorlatához igazodva (Benkő 1972) készítettünk el egy graféma- és ékezhű átíratot, amely bizonyos mértékben paleográfaiailag is hű. A szöveghűség elvét szem előtt tartva a grafémák egységesítését redukáltuk a mellékjel nélküli betűvariánsokra. Abban az esetben, ha egy (mellékjel nélküli) betűnek két vagy annál több jele volt a szövegben, s bizonyosak lehettünk, hogy ezeknek nincs hangértéket megkülönböztető funkciója, a különböző variánsok helyett csak egyet, a ma használt grafémát vezettük be. Ezt a megoldást alkalmaz-

maztuk a szövegben három különböző betűvel jelölt *z* vagy a két variánsban előforduló *s* esetében is.



5. ábra: A *z* graféma variánsai a szövegben



6. ábra: Az *s* graféma variánsai a szövegben

A szöveghűség elvét a nagy- és kisbetű használat esetében is érvényesítettük, valamint megőriztük a szöveg eredeti tagolását és írásjeleit.

A nyelvemlék fizikai állapotának romlása következtében a betűk helyenként elhalványodtak vagy egyáltalán nem láthatók. A kopást, illetve más módon történt sérüléseket a szövegben minden esetben jelöltük, és jegyzettel láttuk el a nyelvészeti forráskiadás elveinek megfelelően (Benkő 1972). A sérült részeket nem pótoltuk vissza a szövegbe, ugyanakkor a jegyzetben feltüntettük, hogy milyen szövegelem hiányozhat, amennyiben az kikövetkeztethető volt.

A betűhű átirat mellett elkészítettük a szöveg normalizált változatát is. A normalizálás során igyekeztünk egy olyan szövegváltozatot készíteni, ami egyszerre segíti a modern kor olvasóját a szöveg befogadásában és megértésben, ugyanakkor – amennyire lehetséges – megőrzi annak eredeti jellegét. A nyelvjárási sajátosságok és archaizmusok esetében több különböző megoldást alkalmaztunk. A könnyebb kereshetőség és az egyértelműség érdekében a hangtani jelenségek többségét nem őriztük meg, hanem normalizáltuk.⁴ Kivételt jelentettek ez alól azok esetek, mikor a nyelvjárási vagy archaikus alak a köznyelvi használatban is elfogadott, például a labiális-illabiális váltakozás egyes szavak esetében (*fel* ~ *föl*, *seper* ~ *söpör*). A toldalékolás eltéréseit minden esetben normalizáltuk, az eredeti szövegtől való eltérést jelezve (*vízbe* > *vízbe'n*). A tájszavakat és az archaizmusokat megőriztük, jelentésüket szószedetben adtuk meg.

⁴ A korábbi történelmi korpuszok (pl. *Biblia Mediaevalis Hungarica*, Középmagyar történelmi magánéleti korpusz) tanulsága, hogy a bennük való keresést jelentős mértékben megkönnyíti, ha a szinkrón köznyelvi változatot is csatolunk a betűhű átiratok mellé. A normalizált átiratok mellékelésének további előnye, hogy a nyelvtörténelmi háttérrel rendelkező szakember, sőt akár még a laikus felhasználók számára is érthető, kereshető lesz a korpusz.

A szöveg nem magyar nyelvű elemeit (pl. a latin kifejezéseket) jelölve benne hagytuk az átiratban, megértésükhöz szintén szószedetet állítottunk össze. A magyar nyelvi elemeket a latin nyelveiktől a todalékolás alapján különítettük el a Középmagyar történeti magánéleti korpusz megoldását követve (Dömötör 2014). Egy jövevényszó akkor minősült magyarnak az adott szöveggörnyezetben, ha magyar nyelvi todalékot kapott, így például az *ápiumot*, *mastixot*, *rusikát*, *tormentillát* kifejezések magyar nyelvűnek tekinthetők, míg a *Citrium*, *fenum graecum*, *mastix* lexémák nem.

A szöveg műfaji sajátosságai további, az olvasót segítő egyedi megoldásokat igényeltek az átirat készítése során. A táj-, idegen nyelvi és archaikus szavakat tartalmazó szószedetek mellett szükségesnek tűnt további, orvosi szakkifejezéseket tartalmazó szótár megalkotása is. A gyűjteménybe a betegség- és növénynevek, mértékegységek, eszközök, eljárások megnevezései kerültek be, például *glét*, *orbánc*, *madragfü*, *sérés*, *spiki-nárd olajt*, *vérállás*.

A betűhű átirathoz hasonlóan a sérült szövegrészek a normalizált változatban is fel vannak tüntetve, megjegyzéssel és a lehetséges olvasattal ellátva. A jegyzetekbe ezen felül a bizonytalan olvasatú helyek is jelezve vannak, szintén az összes lehetséges olvasattal ellátva.

3.2. Nyelvtechnológiai feldolgozás

A nyelvtechnológiának kulcsfontosságú szerepe van abban, hogy a múlt kulturális örökségét széleskörűen elérhetővé és feldolgozhatóvá tegye. A magyar nyelvű orvosi irodalom rendszerezett feldolgozásához olyan adatbázisra van tehát szükség, amelyben az elemek különféle nyelvészeti, történeti, paleográfiai stb. információkkal vannak ellátva (vö. Simon–Sass 2012).

Tekintve, hogy mind a történeti, mind a jelenkori orvosi szaknyelv eltér a mai magyar sztenderd nyelvváltozattól, amelyre nyelvtechnológiai eszközeink többsége készült, az orvosi szövegek feldolgozása más módszereket kíván. A 16–17. századi orvosi receptek alkorpuszának elkészítéséhez mintául szolgál számunkra a Magyar generatív történeti szintaxis projektum ómagyar korpusza (Simon–Sass 2012), illetve a Középmagyar történeti magánéleti korpusz (Dömötör 2014).

Jelen alfejezetben egy recepten, a Mindenféle orvosságoknak rendszedése kéziratban mutatjuk be a korpuszépítés egyes lépéseit. A kézirat PDF formátumban állt rendelkezésünkre. Első lépésként egyszerű, formázást nem tartalmazó szöveggé (txt) alakítottuk, és manuális korrekciót végeztünk.

A majdani keresőfelület kialakításához a Magyar nemzeti szövegtár (Oravecz–Váradi–Sass 2014) és a Magyar történeti szövegtár (Sass 2017) nemrégiben elkészített új felületét vesszük mintául. Ezen korpuszok keresőfelülete a Sketch Engine korpuszkezelő eszköz szabadon elérhető változatán, a NoSketch Engine-en alapul. Ahhoz, hogy a készülő korpuszban az egyes nyelvi jelenségek kereshetők legyenek (például ha latin szavakat, orvosi szakkifejezéseket vagy régi mértékegységneveket szeretnénk keresni), „a releváns információkat elektronikusan előhívható és interpretálható módon kell tárolni” (Simon–Sass 2012: 12). A gyakorlatban ez azt jelenti: a szövegfájlokat olyan formátumúra kell alakítani, hogy a NoSketch Engine képes legyen azt feldolgozni. Ez a formátum egy vertikális formájú (word-per-line) szövegfájl, amely soronként egy szövegszót vagy központosási jelet (összefoglaló néven token) tartalmaz. Minden egyes

szövegszó mellett az alábbi adatoknak kell szerepelniük: betűhű alak, normalizált alak, szótó, morfológiai elemzés (vö. Simon–Sass–Novák 2013). Mindemellett szerepelhetnek egyéb mezők is (latin szó-e, betegségnev, tájszó, értelmezés stb.). A különféle mezők tabulátorral vannak elválasztva, vagyis a nyelvtechnológiában szokásos ún. TSV (tab-separated values) fájlokat kell létrehozni. A mezők (oszlopok) száma tetszőleges, akárhány attribútumot felvehetünk: minden egyes oszlop egy lehetséges keresési szempont.

A TSV fájloknak ezenkívül bizonyos metaadatokat is tartalmazniuk kell. Így például az első oszlop egy fontos metaadat, az ún. lókuszelölő, amely azt adja meg, hogy az adott szövegszó pontosan hol található meg a dokumentumban. Mivel jelen esetben csak egy receptgyűjtemény feldolgozása történt meg, így csupán egyetlen lókuszelölő szám szerepel a táblázatban, az a sorszám, amellyel a digitális átiratok újrászámozása során az egyes recepteket elláttuk. Ugyanakkor a pontos beazonosíthatóság érdekében a későbbiekben szükséges lehet többféle lókuszelölő használata, ez az egyes receptek felépítésétől függ; a lókuszelölők tehát szövegenként változhatnak.

A későbbi kutatások szempontjából fontosnak tartjuk, hogy azt az információt is megőrizzük, az adott szövegszó a receptnek az iniciátor vagy instrukciós részében helyezkedik-e el. Ezt a metaadatot a táblázat második oszlopában található LOCUS attribútummal jelöljük, melynek értéke {inic} vagy {inst} lehet a szó elhelyezkedésétől függően.

A TYPE attribútum értéke jelöli az esetleges szemantikai, lexikai jegyeket, amelyeket az elkészített listák alapján kézzel címkéztünk. A táblázat harmadik oszlopa azt mutatja, miként jelöltük például a betegségneveket (a TYPE attribútum a {disease} értéket kapja). Láthatjuk, hogy az egyes betegségnevek nem csupán egyetlen szóból állhatnak, hanem gyakran szónál nagyobb egység fejezi ki egy-egy betegség nevét (pl. *amely embernek a fogát kivonják, és a vér meg nem állna*).

Az esetleges idegennyelvűséget az opcionális LANG attribútum jelzi, amelynek értéke az adott nyelvet jelöli (a táblázatban a latint).

Az egyes normalizált alakokhoz tartozó szótó (lemma) előállításához és a morfológiai elemzés elkészítéséhez a hunmorph programot használtuk.

A táblázat minden egyes oszlopa egy-egy keresési szempontot jelent. Terveink között szerepel, hogy lehetővé tegyük az egyes szövegtípusok közti keresést (akár külön alkorpuszokban). Természetesen az egyes kéziratokon belül is lehetséges különféle kereséseket végezni: a betűhű alak szerint, illetve a normalizált alak szerint (teljes szóra, szókezdetre, szóvégre is). Kereshetünk latin szavakat, betegségneveket, értelmezéseket.

Lókuszt	Lókusztípusa	Típus	Nyelv	Betűtű alak	Normalizált alak	Lemma	Morfológiai elemzés
15	LOCUS{finc}	TYPE{disease}		Az mely	A mely	amely	NOUN
15	LOCUS{finc}	TYPE{disease}		emberek	emberek	ember	NOUN<CAS<DAT>>
15	LOCUS{finc}	TYPE{disease}		az	a	a	ART
15	LOCUS{finc}	TYPE{disease}		fogat	fogát	fog	NOUN<POSS><CAS<ACC>>
15	LOCUS{finc}	TYPE{disease}		ki unyvszak	ki vonják	ki von	VERB<PLUR><DEF>
15	LOCUS{finc}	TYPE{disease}		.	.	.	PUNCT
15	LOCUS{finc}	TYPE{disease}		és	és	és	CONJ
15	LOCUS{finc}	TYPE{disease}		az	a	a	ART
15	LOCUS{finc}	TYPE{disease}		vére	vére	vér	NOUN<POSS>
15	LOCUS{finc}	TYPE{disease}		megh	meg	meg	PREV
15	LOCUS{finc}	TYPE{disease}		nem	nem	nem	ADV
15	LOCUS{finc}	TYPE{disease}		allana	álana	áll	VERB<COND>
15	LOCUS{finc}	TYPE{disease}		.	.	.	PUNCT
15	LOCUS{instr}			Narancs alma	Narancsalma	narancsalma	NOUN
15	LOCUS{instr}			héjat	héját	héj	NOUN<POSS><CAS<ACC>>
15	LOCUS{instr}			.	.	.	PUNCT
15	LOCUS{instr}			tímsot	tímsót	tímsó	NOUN<CAS<ACC>>
15	LOCUS{instr}			.	.	.	PUNCT
15	LOCUS{instr}		LANG{latin}	bonus	bonus	bonus	bonu NOUN ATTRIB ADJ
15	LOCUS{instr}		LANG{latin}	armenust	armenust	armenus	NOUN<CAS<ACC>>
15	LOCUS{instr}			.	.	.	PUNCT
15	LOCUS{instr}			ezeket	ezeket	ez	NOUN<PLUR><CAS<ACC>>
15	LOCUS{instr}			forrald	forrald	forral	VERB<SUBJUNC-IMP><PERS<?>><DEF>
15	LOCUS{instr}			eöszve	össze	össze	PREV
15	LOCUS{instr}			.	.	.	PUNCT

1. táblázat: Részlet a feldolgozott, elemzett szövegből

4. Összegzés

A Magyar orvosi nyelvi korpusz tehát egy hosszú távú projektként célként tűzi ki a magyar nyelvű orvosi irodalom egészének rendszerezett feldolgozását. A munka számos hozadékkal bír, amelyek közül itt néhányat ki is emelünk:

- Megfoghatóvá és széles körben elérhetővé, kutathatóvá teszi a magyar nyelvű, gyógyításhoz kötődő szövegeket és nyelvhasználatot az egyes korokban és egészében. Az elektronikus szövegtár előnye, hogy módosítható, bővíthető, és szükség esetén javítható.
- Lehetőséget ad a szövegtípusok összekapcsolódásának, hálózatának a leírására. Ez önmagában az orvosi nyelv szövegtípus szerinti állományának a feltáráshoz vezet, amely a szaknyelv összefüggő működésének teljesebb megismeréséhez is hozzájárul, továbbá a szövegtípus-fogalom árnyalásához is vezet.
- Lehetővé teszi a magyar nyelvű orvoslás korszakolásának újragondolását.
- Rámutat az egyes korok társas-kulturális változásaira az orvoslás terén (tudományos fejlődés; az írásbeliségnek és a szóbeliségnek a szerepe; a képi ábrázolás; a gyógyítók rétegzettsége; érzelmi bevontság; idegen nyelvi hatás stb.).
- A magyar orvosi nyelv változásai folyamatként, azok folytonosságában vizsgálhatók.
- Lehetővé válik egyes korok jellegzetességeinek, irányzatainak a vizsgálata.
- Rendszerezett nyelvészeti, orvostörténeti és kulturális kutatásokat tesz lehetővé, biztosítva az egyes szakterületek együttműködését.
- Néhány kutatható jelenség például: a szókincs alakulása (szótár létrehozása); a tudományos nyelv alakulása; a tudományos stílus változása; a szaknyelvi helyesírás alakulása; egyes nyelvi funkciók leírása (pl. az utasítás, a meggyőző kommunikáció alakulása az egyes korokban stb.).
- Megnyitja a nyelven belüli és nyelvek közötti összehasonlító elemzések útját, ezzel leírhatóvá válnak az egyes nyelvek és korok irányzatai, mintázatai, valamint az ún. „univerzális” vagy „páneurópai” jellegzetességek.
- A korpusz révén a magyar orvosi nyelv kutatása jobban beleilleszthetővé válik a külföldön folyó vizsgálatokba, ami a nemzetközi együttműködést is előíranyozza.

Források

- Ars Medica 1577 k./1943 = Várad Lencsés György: Egész orvosságról való könyv azaz Ars Medica (1577); kiadta Varjas Béla 1943. *XVI. századi magyar orvosi könyv*, Kolozsvár, Sárkány Nyomda. (Magyar Elektronikus Könyvtár, <http://mek.oszk.hu/01100/01159>, letöltve 2016. 09. 21.)
- De conservanda 1694/1983 = Felvinczi György: *De conservanda bona valetudine Liber Scholae Salernitanae. Az Anglia (!) országban levő Salernitana Scholanak Jó Egészségről való Meg-tartásnak módgyáról írott Könyve*; szemelvények Szlatky Mária (szerk.) 1983. „Minden doktorságot csak ebből késértek.” *Szemelvények orvosi kézikönyvekből*, Budapest, Magvető, 325–51.

- Egy néhány rendbeli orvosságok 1614–1635 k./1989 = Máriássy János: *Egy néhány rendbeli lóorvosságok és más orvosságok* (1614–1635 k.); Hoffmann 1989: 53–75.
- Gyógyszerek I. [17. sz. eleje]/1989 = *Gyógyszerek I.* (17. sz. eleje?); Hoffmann 1989: 435–45.
- Gyógyszerek II. [17. sz. eleje?]/1989 = *Gyógyszerek II.* (17. sz. eleje?); Hoffmann 1989: 447–58.
- Hasznos és fölötte szükséges könyv 1588/1983 = Frankovith Gergely: *Hasznos és fölötte szükséges könyv* (1588); Sztatky 1983: 96–161.
- Házi patika 1663 k./1989 = *Házi patika.* (1663 k.); Hoffmann 1989: 227–46.
- Herbárium 1578/1979 = Melius Juhász Péter: *Herbárium* (1578); Szabó T. 1979.
- Kis patika 1628/1989 = Váradi Vásárhelyi István: *Kis patika* (1628); Hoffmann 1989: 211–25.
- Liber Medicinarum 1693 k./1983 = Johann Ponyikay: *Liber Medicinarum* (1693); Sztatky 1983: 314–22.
- Medicinae Variæ 1598/1989 = *Medicinae Variæ. Melius Péter Herbáriumához kötött recipék* (1598); Hoffmann 1989: 11–52.
- Medicusai és borbélyi mesterség [1668–1703]/1989 = Becskereki Váradi Szabó György: *Medicusai és borbélyi mesterség* (1668–1703); Hoffmann 1989: 341–434.
- Orbis sensualium pictus 1658/1959 = Comenius, Johannes Amos: *Orbis Sensualium pictus. A látható világ* (1685); kiadta Dr. Geréb György 1959; Johannes Amos Comenius: *A látható világ.* Budapest, Magyar Helikon.
- Orvoskönyv lovak orvoslása 1619 e./1989 = Török János: *Orvoskönyv. Lovak orvoslása* (1619 e.); Hoffmann 1989: 77–171.
- Orvoskönyv némely füveknek hasznáról [17. sz.]/1989 = *Orvoskönyv némely füveknek hasznáról* (17. sz.); Hoffmann 1989: 473–96.
- Orvosságoknak rendszedése [17. sz. 2. fele]/1989 = *Orvosságoknak rendszedése.* (17. sz. 2. fele); Hoffmann 1989: 459–72.
- Orvosságos könyv Apafi Anna számára 1677/1989 = Újhelyi István: *Orvosságos könyv Apafi Anna számára* (1677); Hoffmann 1989: 295–334.
- Orvosságos könyveczke 1664 k./1989 = Orvosságos könyveczke. (1664 k.); Hoffmann 1989: 247–66.
- Pax Corporis 1695/1984 = Pápai Páriz Ferenc: *Pax Corporis* (1695/1764); kiadta Szablyár Ferenc 1984.
- Pettényi Borbély Márton orvosló könyve 1683–1701/1983 = Pettényi Borbély Márton orvosló könyve (1683–1701); szemelvények Sztatky Mária (szerk.) 1983. „Minden doktorságot csak ebből kérésérték.” *Szemelvények orvosi kézikönyvekből.* Budapest, Magvető, 165–225.
- Phlebotomia [17. sz. 2. fele]/1983 = *Phlebotomia* (17. sz. második fele); szemelvények Sztatky Mária (szerk.) 1983. „Minden doktorságot csak ebből kérésérték.” *Szemelvények orvosi kézikönyvekből.* Budapest, Magvető, 236–69.
- Próbálós bizonyos orvosságok 1666; 1693/1989 = Révay István: *Próbálós bizonyos orvosságok* (1666); Pekler György másolatában (1693); Hoffmann 1989: 267–94.
- Próbált orvosságok 1684/1989 = *Próbált orvosságok* (1684); Hoffmann 1989: 335–9.
- Testi orvosságok könyve 1619 k./1989 = Szentgyörgyi János: *Testi orvosságok könyve* (1619. k.); Hoffmann 1989: 173–201.

Hivatkozások

- Barz, Irmhild – Fix, Ulla – Schröder, Marianne – Schuppener, Georg 2000: *Sprachgeschichte als Textsortengeschichte: Festschrift zum 65. Geburtstag von Gotthard Lerchner*, Frankfurt am Main, Peter Lang.
- Benkő Loránd 1972: *Tudnivalók a Magyar Nyelv közleményeinek alakításmódjához* (A Magyar Nyelvtudományi Társaság Kiadványai 129.), Budapest.
- Csontos Nóra 2013: A pragmatika mint szemléletmód érvényesítésének lehetősége a történetiségben, *Magyar Nyelv* 109, 149–57.
- Dömötör Adrienne 2014: Az ó- és középmagyar kori magánéleti nyelvhasználat morfológiailag elemzett adatbázisa, in Fazakas Emese – Juhász Dezső – T. Szabó Csilla – Terbe Erika – Zsemlyei Borbála szerk.: *Tér, idő, társadalom és kultúra metszéspontjai a magyar nyelvben*, Budapest–Kolozsvár, ELTE Magyar Nyelvtörténeti, Szociolingvisztikai, Dialektológiai Tanszék–Nemzetközi Magyarstudományi Társaság, 11–21.
- Eckkrammer, Eva Martha 2005: *Medizin für den Laien vom Pesttraktat zum digitalen Ratgebertext. Augliederung, Pragmatik, Struktur-, Sprach- und Bildwandel fachexterner Textsorten unter Berücksichtigung des Medienwechsels. Habilitationsschrift. Salzburg.*
(<http://www.uni-salzburg.at/pls/portal/docs/1/556745.PDF>; a letöltés dátuma 2010. 09. 10)
- Görlach, Manfred 2004: *Text types and the history of English*, Berlin–New York, Mouton de Gruyter.
- Heinemann, Wolfgang 2000: Textsorte – Textmuster – Texttyp, in Brinker, Klaus – Antos, Gerd – Heinemann, Wolfgang – Sager, Sven Frederik Hrsg.: *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung* (HSK. 16.1.), Berlin–New York, Walter de Gruyter, 507–23.
- Hoffmann Gizella szerk. 1989: *Medicusi és borbélyi mesterség. Régi magyar ember- és állatorvosló könyvek. Radvánszky Béla gyűjtéséből*, Szeged, József Attila Tudományegyetem.
- Kocsány Piroska 2002: *Szöveg, szövegtípus, jelentés: A mondás mint szövegtípus = Nyelvtudományi Értekezések* 151, Budapest, Akadémiai.
- Kocsány Piroska 2006: A szövegtipológia eredményei és/vagy eredménytelenségei, in Tolcsvai Nagy Gábor szerk.: *Szöveg és szövegtípus. Szövegtipológiai tanulmányok*, Budapest, Tinta, 17–26.
- Kohnen, Thomas 2007: Text types and the methodology of diachronic speech act analysis, in Fitzmaurice, Susan M. – Taavitsainen, Irma eds.: *Methods in Historical Pragmatics* (Topics in English Linguistics 52.), Berlin, Mouton de Gruyter, 139–66.
- Kuna Ágnes 2011: *A 16–17. századi magyar nyelvű orvosi recept szövegtipológiai és pragmatikai vizsgálata funkcionális-kognitív keretben*, Budapest, ELTE (doktori disszertáció).
- Kuna Ágnes 2012: Az orvosi receptek alakulása a társadalmi és kulturális tényezők függvényében a 16. századtól napjainkig, in Balázs Géza – Veszelszki Ágnes szerk.: *Nyelv és kultúra. Kulturális nyelvészet*, Budapest, Magyar Szemiotikai Társaság, 156–62.

- Kuna Ágnes 2013: *A magyar nyelvű orvosi diskurzustartomány, a magyar orvosi nyelv tendenciái. A Magyar Orvosi Nyelv Korpusza*. Tervezet. Kézirat, Budapest.
- Kuna Ágnes 2015: *A meggyőzés nyelvi mintázatai a 16–17. századi orvosi receptben. Kognitív nyelvészeti elemzés*, Budapest, ELTE (doktori disszertáció).
- Kuna, Ágnes 2016a: Genre in a cognitive functional framework. Medical recipe as a genre in 16th and 17th century Hungarian, in Stukker, Ninke – Spooren, Wilbert – Steen, Gerard eds.: *Genre in Language, Discourse and Cognition*, Berlin–New York, De Gruyter, 193–224.
- Kuna Ágnes 2016b: A Magyar Orvosi Nyelv Korpusza, *Magyar Orvosi Nyelv* 16/1: 26–30.
- Kuna Ágnes 2017: Történeti szövegtipológia, in Forgács Tamás – Németh Miklós – Sinkovics Balázs szerk.: *A nyelvtörténeti kutatások újabb eredményei IX*, Szeged, SZTE Magyar Nyelvészeti Tanszék, 223–37.
- Maitz Péter – Molnár Anna 2001: Nyelvtörténetírás és történeti szövegnyelvészet, in Csatár Péter – Maitz Péter – Tronka Krisztián szerk.: *A nyelvtantól a szövegtanig. Tanulmányok Kocsány Piroska tiszteletére*, Debrecen, Kossuth Egyetemi Kiadó, 322–35.
- Oravecz, Csaba – Várad, Tamás – Sass, Bálint 2014: The Hungarian Gigaword Corpus, in Calzolari, Nicoletta – Choukri, Khalid – Declerck, Thierry – Loftsson, Hrafn – Maegaard, Bente – Mariani, Joseph – Moreno, Asuncion – Odijk, Jan – Piperidis, Stelios eds.: *Proceedings of Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, 1719–23.
- Riecke, Jörg 2004: *Die Frühgeschichte der mittelalterlichen medizinischen Fachsprache im Deutschen I–II*, Berlin–New York, Walter de Gruyter.
- S. Sárdi Margit szerk. 2002: *Szentgörgyi János: Testi orvosságok könyve*, Máriabesenyő, Attraktor.
- Sass Bálint 2017: A kibővített Magyar Történeti Szövegtár új keresőfelülete, in Forgács Tamás – Németh Miklós – Sinkovics Balázs szerk.: *A nyelvtörténeti kutatások újabb eredményei IX*, Szeged, SZTE Magyar Nyelvészeti Tanszék, 267–77.
- Simon Eszter – Sass Bálint 2012: Nyelvtechnológia és kulturális örökség, avagy korpuszépítés ómagyar kódexekből, in Prószéky Gábor – Várad Tamás szerk.: *Általános Nyelvészeti Tanulmányok XXIV. Nyelvtechnológiai kutatások*, Budapest, Akadémiai, 243–64.
- Simon Eszter – Sass Bálint – Novák Attila 2013: *Kódolási szabályzat*. Kézirat, Budapest.
- Steger, Hugo 1984: Sprachgeschichte als Geschichte der Textsorten/Texttypen und ihrer kommunikativen Bezugsbereiche, in Besch, Werner – Reichmann, Oskar – Sonderegger, Stefan Hrsg.: *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*. Bd. I. Berlin–New York, Walter de Gruyter, 186–204.
- Szablyár Ferenc 1984: Pápai Páriz Ferenc: *Pax Corporis*. Budapest, Magvető.
- Szabó T. Attila 1979: Melius Péter és a kolozsvári Herbárium, in Melius Péter: *Herbárium. Az fáknak, füveknek nevekről és hasznairól. Bevezető tanulmánnyal és magyarázó jegyzetekkel sajtó alá rendezte Szabó T. Attila*, Bukarest, Kriterion.

- Szabó T. Attila – Bíró Zsolt 2000: *Ars Medica Electronica: Váradi Lencsés György (1530–1593)*. CD-ROM. BioTár Electronic, Gramma 3.1. & 3.2. Budapest–Kolozsvár–Szombathely–Veszprém, MTA–EME–BDF–VE.
- Szlatky Mária szerk. 1983: „*Minden doktorságot csak ebből késértek.*” *Szemelvények orvosi kézikönyvekből*, Budapest, Magvető.
- Taavitsainen, Irma 2001: Changing conventions of writing: the dynamics of genres, text types, and text traditions, *European Journal of English Studies* 5/2, 139–50.
- Taavitsainen, Irma–Pahta, Päivi eds. 2004: *Medical and scientific writing in late medieval English*, Cambridge, Cambridge University Press.
- Taavitsainen, Irma – Pahta, Päivi – Mäkinen, Martti eds. 2005: *Middle English Medical Texts* (MEMT CD-ROM), Amsterdam/Philadelphia, John Benjamins.
- Taavitsainen, Irma – Pahta, Päivi – Hiltunen, Turo – Mäkinen, Martti – Marttila, Ville – Ratia, Maura – Suhr, Carla – Tyrkkö, Jukka eds. 2010: *Early Modern English Medical Texts*, CD-ROM, Amsterdam, John Benjamins.
- Tolcsvai Nagy Gábor 2001: *A magyar nyelv szövegtana*, Budapest, Nemzeti Tankönyvkiadó.
- Tolcsvai Nagy Gábor szerk. 2006: *Szöveg és típus. Szövegtipológiai tanulmányok*, Budapest, Tinta.
- Varjas Béla 1943: *XVI. századi magyar orvosi könyv*, Kolozsvár, Sárkány Nyomda.
- Voigts, Linda Ehrsam – Kurtz, Patricia Deery eds. 2001: *Scientific and medical writings in Old and Middle English: An electronic reference*, CD-ROM, Ann Arbor, University of Michigan Press.

Hivatkozott elektronikus történeti korpuszok

Corpus of Early English Medical Writing

(CEEM) (<http://www.helsinki.fi/varieng/CoRD/corpora/CEEM/>)

Középmagyar történeti magánéleti korpusz (<http://tmk.nytud.hu/>)

Magyar generatív történeti szintaxis (<http://omagyarkorpusz.nytud.hu/>)

Magyar történeti szövegtár (<http://www.nytud.hu/hhc/>)

The Malaga Corpus of Late Middle English Scientific Prose (<http://hunter.uma.es/>)