

Orvosi korpuszok és történetiség

Kuna Ágnes

ELTE Alkalmazott Nyelvészeti és Fonetikai Tanszék

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

Összefoglaló

A tanulmány középpontjában az orvosi nyelvi korpuszok témaköre áll. Az írás bemutatja a modern kori orvosi szaknyelvhez kötődő adatbázisok néhány tendenciáját, majd ismerteti a *Corpus of Early English Medical Writing* angol nyelvű projekt főbb jellegzetességeit, ami át is vezet a folyamatban lévő magyar nyelvű orvosi korpusz építésének a lépéseire. A tanulmány fő célja, hogy ismertesse a *Magyar orvosi nyelvi korpusz 16–17. századi alkorpuszá-*nak (MOKO) munkálatait. Így kitér az anyaggyűjtésnek, az adatbázis koncepciójának, az átírás elveinek és a nyelvtechnológiai feldolgozásnak a kérdésére is.

Kulcsszavak: történeti pragmatika, orvosi történeti korpusz, MOKO, átírás, normalizálás, kézirat, recept, nyelvtechnológiai feldolgozás

1. Bevezetés¹

A magyar orvosi nyelv történetére vonatkozóan nem áll rendelkezésre egy szisztematikusan összeállított, elektronikus korpusz, pedig ez számos tekintetben segítené a korai szövegek megértését és feldolgozását, ahogy azt nemzetközi példák mutatják (Taavitsainen–Pahta–Mäkinen 2005; Taavitsainen et al. 2010; Taavitsainen–Hiltunen 2019),² és a magyar nyelvű történeti korpuszok viszonylatában is látható (Simon–Sass 2012; Dömötör et al. 2017; Novák et al. 2018; Gugán et al. 2023).³ A történeti korpuszok azon túl, hogy jól meghatározott szempontok szerint a lehető legrepresentatívabb képet mutatják az adott korra, témára, diskurzusműfajra⁴ vagy egyéb jelenségre összpontosítva, egyfajta filológiai háttérrel is biztosítanak mind a forrásokra, mind az átírásra vonatkozóan, és így nagyban segítik az egyes korok

* „To understand the characteristics of present-day scientific writing, we need to know about the earlier stages of language use.” (a szerző fordítása)

¹ Jelen tanulmány nagyban támaszkodik az alábbi, már megjelent munkákra: Kuna (2011; 2016a; 2024a: 8. fejezet) és Kuna–Kocsis–Ludányi (2017).

² Erre egy kiemelkedő példa a *Corpus of Early English Medical Writing 1375–1800* projekt (<https://tinyurl.com/jeny-2025-KunaA-01>; letöltés: 2024. 01. 26.).

³ *Ómagyar Korpusz* (<http://omagyarkorpusz.nytud.hu>; letöltés: 2024. 01. 26.); *Történeti Magánéleti Korpusz* (<http://tmk.nytud.hu>; letöltés: 2024. 01. 26.); *Magyar Történeti Szövegtár* (<https://tinyurl.com/jeny-2025-KunaA-02>; letöltés: 2024. 01. 26.).

⁴ A *szövegtípus*, *diskurzustípus* és a *diskurzusműfaj* terminusokat szinonim értelemben használom (Kuna–Simon 2017).



nyelvi jellemzőinek, kommunikatív gyakorlatainak a vizsgálatát. Ezekkel összefüggésben már több mint tíz éve megfogalmazódott a *Magyar orvosi nyelvi korpusz* (MOKO) 16–17. századi alkorpuszáinak a gondolata (Kuna 2011).

Ehhez kapcsolódóan jelen tanulmány célja, hogy bemutassa egy folyamatban lévő orvosi történeti korpusz készítésének elveit, lépéseit, valamint ennek a munkának a szűkebb és tágabb tudományos kontextusát. Ezzel összefüggésben elsőként rámutatok az orvosi korpuszok néhány jellegzetességére jelen és történeti viszonyban (2.), majd ismertetem a MOKO alapkoncepcióját (3.). Ezután térek rá a már elkészült, illetve folyamatban lévő munkálatokra, külön kiemelve az átírást (4.) és a nyelvtechnológiai feldolgozásra vonatkozó lépéseket (5.). Összegzésként azt mutatom be, hogy milyen hozadékkal bírhat a MOKO 16–17. századi alkorpusza, amelynek munkálatai jelenleg is folyamatban vannak.⁵

2. Korpuszok és orvosi nyelv

Az elmúlt két évtizedben egyre nagyobb számmal keletkeztek orvosi nyelvre vonatkozó korpuszok mind jelenkori,⁶ mind történeti viszonylatban. Magyar nyelven is egyre több egészségügyi kontextusra vonatkozó szövegtár áll rendelkezésre, új utakat nyitva a modern kori szaknyelvi kutatásoknak. A Szegedi Tudományegyetemen például több orvosi korpusz is épült: A *BioScope* orvosi és biológiai szövegeket dolgoz fel (Vincze et al. 2008); a *Szegedi Medical Korpusz* (SZEMEK) orvostudományi szaklapok tanulmányainak nyelvi elemzését teszi lehetővé (Keresztes–Borda–Vincze 2019). Az Eötvös Loránd Tudományegyetemen készül a *Pannónia Korpusz* részeként egy orvosi alkorpusz is, amely párhuzamos szövegei révén a szakfordítást és a szaknyelvoktatást szolgálja (Robin et al. 2016). Folyamatban van továbbá egy tudományos együttműködésen alapuló, közel hárommillió szövegszós *Covid Corpus* is, amely az internetes hírportálok koronavírushoz kapcsolódó terminológiai elemzését teszi lehetővé (Varga et al. 2023).⁷

Történeti viszonylatban is több orvosi korpusz jött létre a 21. században. Ebben kiemelt szerepet kap az angol nyelvre vonatkozó orvosi adatbázisok aránya, mint

⁵ Az eddigi munkálatokról lásd bővebben Kuna (2011; 2016a) és Kuna–Kocsis–Ludányi (2017).

⁶ Csak néhány példát kiemelve: *Medizinkorpusz* (<https://tinyurl.com/jeny-2025-KunaA-03>; letöltés: 2023. 12. 12.); *Medical Corpus: English corpus from the web* (<https://tinyurl.com/jeny-2025-KunaA-04>; letöltés: 2024. 01. 10.); *Oslo Medical Corpus* (<https://tinyurl.com/jeny-2025-KunaA-05>; letöltés: 2024. 01. 15.); *UFAL Medical Corpus v. 1.0* (<https://tinyurl.com/jeny-2025-KunaA-06>; letöltés: 2024. 01. 21.); *An Open Access Corpus of Scientific, Technical, and Medical Content* (<https://tinyurl.com/jeny-2025-KunaA-07>; letöltés: 2023. 11. 10.).

⁷ Nemzetközi viszonylatban a beszélt nyelvi korpuszok is egyre nagyobb szerepet kapnak egészségügyi kontextusban. Kiemelkedő Hollandiában a *Nivel Primary Care Database* (<https://tinyurl.com/jeny-2025-KunaA-08>; letöltés: 2023. 10. 26.), amelyet 1970-es évektől gyűjtenek szisztematikusan a kutatók, és mára megközelítőleg 18 000 hang- és videófelvételt tartalmaz a kutatói adatbázis. Más projekteknél készült hanganyagok átiratai elérhetőek publikusan, mint például a *Physician-patient transcripts with 4C coding analysis from the Contextualizing Care research program* (<https://tinyurl.com/jeny-2025-KunaA-09>; letöltés: 2023. 10. 27.). Több nyelven is létezik kutatói adatbázis szimulált helyzetekre vonatkozóan, ilyen online anyagokat tartalmaz például a *DocCom* oktatási felület (<https://tinyurl.com/jeny-2025-KunaA-10>; letöltés: 2023. 10. 28.). Magyar viszonylatban egyelőre kisebb, egyéni adatbázisokkal dolgoznak a kutatók.

például a Malaga korpusz (*The Malaga Corpus of Late Middle English Scientific Prose*; <http://hunter.uma.es/>), vagy a kifejezetten recepteket magában foglaló *Corpus of Early English Recipes* (Alonso-Almeida-Ortega-Barrera-Quintana-Toledo 2012).

Az angol nyelvű korpuszok közül egy nagyobb projektet, a *Corpus of Early English Medical Writing* (CEEM)⁸ adatbázist emelem ki, abból az okból kifolyólag is, hogy ez különösen nagy jelentőségű, és számos történeti pragmatikai munka alapját is képezi. A CEEM projektje 1995 és 2019 között zajlott az angol nyelvű orvosi irodalomnak három nagy szakaszát felölelve, amelyet összegzően az 1. táblázat mutat be.

Alkorpusz neve	Szavak száma (db)	Szövegek száma (db)	Időszak	A korpusz elkészítésének időszaka
<i>Middle English Medical Texts</i> (MEMT)	495 322	86	1375–1500	1995–2005
<i>Early Modern English Medical Texts</i> (EMEMT)	2 millió	450	1500–1700	2000–2010
<i>Late Modern English Medical Texts</i> (LMEMT)	több mint 2 millió	628	1700–1800	2010–2018

1. táblázat

A Corpus of Early English Medical Writing (CEEM) alkorpuszainak adatai⁹

A projekt nagy előnye, hogy nem kizárólag a korpusz készült el, hanem a nyelv-történeti kutatásokat is meghatározó elemzések sora valósult meg magas tudományos színvonalon, más történeti kutatásokra is nagy hatást gyakorolva (lásd összegzően Taavitsainen–Pahta–Mäkinen 2005; Taavitsainen et al. 2010; Taavitsainen–Hiltunen 2019).

Ez a nagyszabású projekt rávilágít arra, hogy angol nyelvre vonatkozóan jóval korábbról vannak fennmaradt orvosi szövegek, mint magyarul. Másrészt azt is láthatóvá teszi, hogy a későbbi korokban nagyobb szövegszó-mennyiségű korpuszok létrehozása lehetséges, és a kor előrehaladtával egyre többféle és specifikusabb szövegtípus jelenik meg, összefüggésben az orvosi tudás bővülésével és specializálódásával (Taavitsainen–Hiltunen 2019). Ugyanakkor a receptek mindegyikben kiemelt diskurzusműfajnak számít. Harmadrészt megmutatkozik az is, hogy egy-egy (al)korpusz létrehozása hosszú időt vesz igénybe (8–10 év). Emellett még egy gyakorlati, ugyanakkor fontos tényező, hogy mindegyik korpusz aktív és támogatott kutatócsoport munkájának eredményeként valósulhatott meg.¹⁰ A CEEM projektjének felsorolt hozadékait a készülő *Magyar orvosi nyelvi korpusz* 16–17. századi alkorpuszának bemutatásakor is figyelembe veszem.

⁸ <https://tinyurl.com/jeny-2025-KunaA-11>; letöltés: 2024. 01. 26.)

⁹ Bővebb információk a kutatócsoportról, a szövegekről az alábbi linkeken és kötetekben található: <https://tinyurl.com/jeny-2025-KunaA-12>; <https://tinyurl.com/jeny-2025-KunaA-13>; <https://tinyurl.com/jeny-2025-KunaA-14>; Taavitsainen–Pahta–Mäkinen (2005); Taavitsainen et al. (2010); Taavitsainen–Hiltunen (2019).

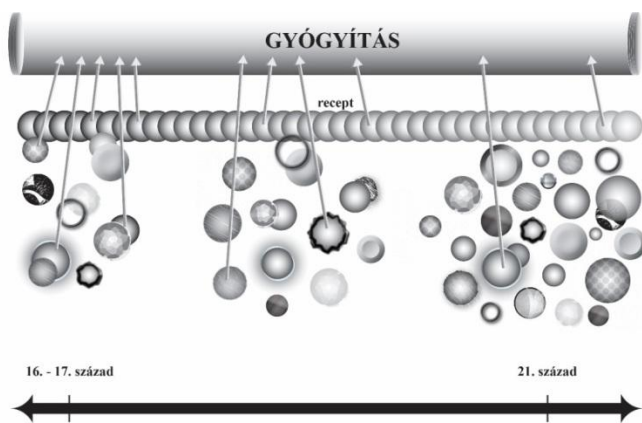
¹⁰ A projektben több mint 10 fő vett részt aktívan 23 éven keresztül, a megvalósításhoz a Finnish Academy Centre of Excellence program nyújtott anyagi támogatást (Taavitsainen–Hiltunen 2019: IX).

3. A Magyar orvosi korpusz (MOKO) alapkonceptiója¹¹

A medicina történetében különböző jellegzetes diskurzusműfajok alakultak ki az évszázadok folyamán a társadalmi és kommunikatív igények mentén, amelyek hozzáférést biztosítanak az orvosi nyelv fejlődésének vizsgálatához (Kuna 2024a). Az orvosi nyelv lényegében egy alakulástörténeti folyamat, múlt, jelen és jövő egysége (Csontos 2013). Ennek megfelelően a történeti és a jelenkori orvosi szaknyelv, illetve a gyógyításhoz kapcsolódó kommunikáció egyazon dinamikus folyamat részeként értelmezhető (Kuna 2011; 2019).

Ezekkel a gondolatokkal összefüggésben a MOKO projektjét egy nagyszabású alapgondolat vezérli: a magyar orvosi szaknyelv megismerését, feldolgozását nagyban segítené, ha az egyes korok fennmaradt, valamint a jelenkor keletkező szakszövegeit strukturált módon (al)korpuszokba rendeznénk. Ez lehetővé tenné az orvosi szaknyelv egész alakulástörténetének szövegalapú leírását, ami természetesen csak lépésenként (és hosszú távon) megvalósítható feladat mind a történeti, mind az egyre specifikusabb tudás szempontjait is szem előtt tartva. Ebben az alfejezetben kizárólag a 16–17. századi magyar szövegekre összpontosítok.

Az orvoslás és ennek nyelvi feldolgozása, az orvosi nyelv alakulástörténete a kor előrehaladtával egyre komplexebb képet mutat a szaktudás növekedésével, az egészségügyi ellátás szociokulturális változásával összhangban. Egyre nagyobb számban jelennek meg új diskurzusműfajok, és a szövegek száma is nagymértékben gyarapszik. Többek között ezt ábrázolja sematikusán az 1. ábra.



1. ábra

A gyógyítás diskurzustartományja és a recept alakulástörténete

¹¹ A korpuszépítés szükségszerűen csapatmunka. A MOKO-n több hallgató és kutató dolgozott, dolgozik. Ezúton köszönöm a 2015–2018 között működő Orvosi Nyelvi Diákműhely (KRE) MA-s tagjainak, Kornis-Balotay Boglárkának, Mogyorósi Kamillának, Skordai Annának, valamint az akkor még doktorandusz Kocsis Zsuzsannának és Kocsis Rékának az aktív és inspiráló munkáját. A projekt részben az MTA posztdoktori egyéni kutatói támogatásával készült, amelyet a szerző a 2016–2018 közötti időszakra nyert el. A további munkálatokban Ludányi Zsófia is részt vállalt (Kuna-Kocsis-Ludányi 2017).

Azt is láthatjuk továbbá, hogy egyes diskurzustípusok hosszú évszázadokon keresztül kísérik a medicina történetét azon oknál fogva, hogy fontos szerepet töltenek be a gyógyítás gyakorlatában (társadalmi relevancia, Heinemann 2000). Ilyen diskurzustípusok a receptek például, hiszen a gyógyítás korai szakaszától a mai napig jelen lévő szövegtípusról van szó. Ugyanakkor az is megfigyelhető, hogy az orvosi recept szerepe nagy eltérést mutat a történeti időkhöz képest (Kuna 2009). Amíg a 16–17. században a receptek az orvosi (és egyéb tudás) átadásának, menedzselésének fontos részét képezték, addig ma egy szűkkörű, intézményesen előírt és jellemzően virtuálisan megvalósuló kommunikáció egy elemeként jelennek meg (Kuna 2019: 28–31).

Az 1. ábra a recept szövegtípusának alakulástörténetén túl arra is felhívja a figyelmet, hogy az egyes korokban fellelhető reprezentatív szövegtípusok felfejtése, továbbá a szöveghagyomány bevonása fontos alapját képezi az orvosi korpusz kialakításának. Lényegében az orvosi nyelv története az adott korszakban jelen lévő diskurzustípusok egyfajta történeteként értelmeződhet (Steger 1984; Barz et al. 2000; Maitz–Molnár 2001), ami lehetővé teszi a szaknyelv fejlődésének dinamikus vizsgálatát, a szaknyelvet alakító társadalmi és kulturális tényezők bevonását is, ahogy arra a CEEM is rámutat. A diskurzustípusok tehát alkalmas keretet adhatnak a szaknyelvi korpusz kialakításához, a szaknyelv szociokulturális komplexitásának megragadásához (Steger 1984; Taavitsainen 2001; Görlach 2004; Kohnen 2004; Kuna 2011). Hiszen ezek nem absztrakt, önmaguktól és önmagukért létező nyelvészeti kategóriák, hanem a társas-kommunikatív igényeket kielégítő megismerési és (egyéni és társas) használati sémák (Taavitsainen 2001; Kocsány 2002; 2006; Tolcsvai Nagy 2006; Kuna 2016b; 2017; 2024a). Így alkalmasak a kommunikatív mintázatok bemutatására.

Az egyes korok diskurzustípus-repertoárjának, illetve ezek egymáshoz való kapcsolatának és hálózatának a feltérképezése tehát az orvosi nyelv dinamikus alakulásának, alakulástörténetének a megismerését is lehetővé teszi.

3.1. A MOKO 16–17. századi alkorpusz

A történeti orvosi szaknyelv feldolgozásának kezdeti lépését a 16–17. századi orvosi munkáknak a korpuszba rendezése jelenti, összhangban a magyar nyelvű írásos orvosi nyelv történetével (Kuna 2024a). A gyógyításban már igen korai időszaktól kezdve megfigyelhető a mindennapi és az egyre tudományosabb elvárások megjelenése, és ezzel az orvosi szövegek felhasználói szintek szerinti rétegződése, ami a mai napig fennáll, és hatással van a korpuszok kialakítására is (Kuna–Ludányi 2019; Kuna 2024b). A vizsgált korban is megfigyelhető a tudományos-mindennapi skálája, illetve rétegzettsége, mégis fontos megjegyezni, hogy ebben az időszakban a magyarnyelvűség szorosan összefonódott a mindennapi igényekkel és az ismeretterjesztéssel. Ezek a szövegek tehát azt a célt szolgálták, hogy a gyógyításra vonatkozó tudás minél szélesebb réteghez jusson el. Az anyanyelvűség jelentősebb szempont volt annál is, mint hogy az egyes munkák a kéziratos vagy nyomtatott formában jelentek-e meg, illetve terjedtek-e (Szlatky 1980: 131).

Az ismertetett szempontok együttesen határozzák meg, hogy milyen szövegek kerüljenek a MOKO 16–17. századi alkorpuszába annak érdekében, hogy a gyógyításhoz kötődő nyelvhasználatról árnyaltabb képet kaphassunk. Ezzel összhangban

elsőként a 2. táblázatban szereplő orvosi munkák feldolgozását tervezzük (részletebben I. Kuna 2024a).¹²

	Cím	A keletkezés ideje	A felhasznált utóközlés ideje	Rövidítés
KÉZIRATOK	<i>Ars Medica</i>	1577 k.	1943	AM
	<i>Medicinae Variae</i>	1598	1989	MV
	<i>Egy néhány rendbeli orvosságok</i>	1614–1635 k.	1989	ENRO
	<i>Orvoskönyv lovak orvoslása</i>	1619 e.	1989	OLO
	<i>Testi orvosságok könyve</i>	1619 k.	1989	TOK
	<i>Kis patika</i>	1628	1989	KP
	<i>Gyógyszerek I.</i>	17. sz. eleje	1989	GY1
	<i>Gyógyszerek II.</i>	17. sz. eleje	1989	GY2
	<i>Házi patika</i>	1663 k.	1989	HP
	<i>Orvosságos könyveczke</i>	1664 k.	1989	OK
	<i>Próbálós bizonyos orvosságok</i>	1666; 1693	1989	PBO
	<i>Orvosságos könyv Apafi Anna számára</i>	1677	1989	OKAASZ
	<i>Orvosságos könyv (Bihari Gergely)</i> ¹³	1677	2019	OKBG
	<i>Medicusi és borbélyi mesterség</i>	1668–1703	1989	MBM
	<i>Pettényi Borbély Márton orvosló könyve</i>	1683–1701	1983	PBMOK
	<i>Orvosságos könyv (Gellén Gergely)</i>	1680	2017	OKGG
	<i>Próbált orvosságok</i>	1684	1989	PO
	<i>Liber Medicinarum</i>	1693 k.	1983	LM
	<i>Mindenféle orvosságoknak rendszedése</i>	17. sz. 2. fele	1989	MOR
	<i>Phlebotomia</i>	17. sz. 2. fele	1983	Ph.
<i>Orvoskönyv némely füveknek hasznáról</i>	17. sz.	1989	ONFH	
NYOMTAT- VÁNYOK	<i>Herbárium</i>	1578	1979	Herb.
	<i>Hasznos és fölötte szükséges könyv</i>	1588	/1983	HFSZK
	<i>Orbis sensualium pictus</i>	1658	1959	OSP
	<i>De conservanda</i>	1694	1983	DC
	<i>Pax Corporis</i>	1695	/1984	PC

2. táblázat
A MOKO tervezett kéziratjai

A kéziratok és nyomtatott munkák közül hat kézirat kétszintű átírása már megvalósult:

- Váradi Vásárhelyi István: *Kis patika* (KP; 1628; Duod. Hung. 66.; 64 f., 26 üres p.)
- *Mindenféle orvosságok rendszedése* (MOR; 17. sz. 2. fele; Oct. Hung. 87)
- *Orvosságos könyveczke* (OK; 1664k.; Oct. Hung. 87)
- Szentgyörgyi János: *Testi orvosságok könyve* (TOK; 1619 k.; Oct. Hung. 500.)
- Becskereki Váradi Szabó György: *Medicusi és borbélyi mesterség* (MBM; 1668–1703; Nr. 5295.; Székelyudvarhelyi Könyvtár)
- *Házi Patika* (HP; 1663 k.; Oct. Hung. 485; 52 levél)

¹² A lista nem teljes, és bővíthető levelezésekben található szövegekkel is. Fontos döntés továbbá, hogy a töredékes munkák milyen mértékben képezzék részét a korpusznak.

¹³ A rövidítésbe ebben az esetben a szerző nevét is bevontam, hogy egyértelmű legyen, melyik kézitról van szó.

A *Medicusi és borbélyi mesterség* kéziratot kívül az itt felsorolt kéziratok mind-egyikét digitalizáltuk az Országos Széchényi Könyvtár hozzájárulásával.¹⁴ Az MBM esetében a Hoffmann (1989)-féle átíratot vettük alapul. Az átírás és a feldolgozás többkörös folyamatban, meghatározott szabályzat szerint zajlott a Károli Gáspár Református Egyetemen működő *Orvosi Nyelvi Diákműhely* keretein belül. Ezeknek a főbb lépéseit ismertetem a következő alfejezetekben.¹⁵

3.2. Az átírás elvei

A 16–17. századi alkorpusz kialakításakor figyelembe vettük a meglévő magyar nyelvű történeti korpuszok, valamint a nemzetközi orvosi nyelvi szövegtárak kialakítási és átírási elveit is,¹⁶ majd saját szabályzatot alakítottunk ki a többkörös átírási folyamatok szervezésére és rendszerezésére.¹⁷ Elsődleges feladat az eredeti szövegek betűhű, valamint normalizált átíratainak az elkészítése, ami alapfeltétele a nyelvtechnológiai feldolgozásnak is. Ezen munkafázisokat követően válik majd lehetségessé a korpusz honlapjának és keresőjének létrehozása. A tervek szerint az egyes kéziratokat PDF-formátumban is elérhetővé szeretnénk tenni (lásd 2. ábra) annak érdekében, hogy a kutatók teljességében is hozzáférjenek az egyes korai munkákhoz.¹⁸ Így kutathatóvá válik például a helyesírás, az írásjelhasználat, a szövegbeli elrendezés és számos más jelenség is. A továbbiakban az átírási és a nyelvtechnológiai feldolgozás lépéseit egy kéziraton, a *Mindenféle orvosságoknak rendszedése* (MOR) alapján ismerhetjük meg.



2. ábra
Az átírás fázisai és formátumai

¹⁴ Ezúton is köszönöm Liphay Endre segítőkészségét, és Kocsis Zsuzsanna, valamint Kocsis Réka digitalizálásban nyújtott segítségét.

¹⁵ Az itt bemutatott munkálatokat Kocsis Zsuzsannával és Ludányi Zsófiával közösen publikáltuk (Kuna–Kocsis–Ludányi 2017), az eredmények itteni publikálása szerzőtársaim hozzájárulásával történik.

¹⁶ Lásd a korábban szóba kerülő korpuszokat. Taavitsainen–Pahta (2004); Taavitsainen–Pahta–Mäkinen (2005); Taavitsainen et al. (2010); Voigts–Kurtz (2001); német nyelvű kutatás Riecke (2004); Eckkrammer (2016); magyar nyelven Simon–Sass (2012); Dömötör et al. (2017); Novák et al. (2018).

¹⁷ A közel 10 oldalas szabályzatban minden eddig felmerült elvet és problémát rögzítettünk az átírásra vonatkozóan. Csak néhányat kiemelve, ilyen kérdések az idegen nyelvű vagy tájnyelvi elemek, a törölt vagy elliptikus részek, az iniciátorok vagy azok hiányának a jelölése, a tömbösödött receptek számozása, az érthetetlen vagy olvashatatlan szövegrészek, számnevek, tulajdonnevek jelölése.

¹⁸ Ez kizárólag az engedélyek beszerzését követően lehetséges. Az OSZK-val a szakmai együttműködés megkezdődött, mivel azonban a projekt nem tart publikálási fázisban, ez a kérdés is valamelyest háttérbe szorult.

Az adatbázis építése során az egyik elsődleges feladat az eredeti szövegek digitális átíratának elkészítése. A feldolgozásra kijelölt szöveg a *Mindenféle 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 (Korompay 2003; Kniezsa 1952a; 1952b). Az írásjelek közül a pont és a vessző található meg, előbbi mondat-, utóbbi tagmondathatároló funkcióban, többnyire következetesen.

A receptek a kéziratban könnyen elkülöníthetők, mivel bekezdéssel vannak egymástól elválasztva, s részben számozottak, ahogy ezt a 3. ábrán is láthatjuk.



3. 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 sorszámozástól, 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, ahogy azt az (1)-es példa is szemlélteti.

(1) 1.
(Ala megh gjúl lonak Csomobā)
(Alá meggyűl lónak csomóba'n)

Borotvald el rola szőrít azon helen,
 Borotváld el róla szőrét azon helyen,

az fog hajjmat o hajal törd ősze,
 a fokhajymát óhájjal törd össze,

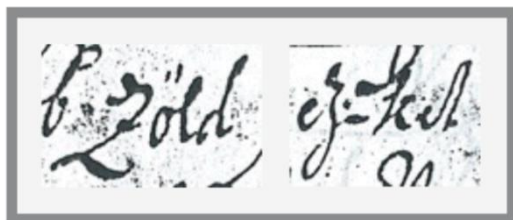
es melegen kend Véle,
 és melegen kend véle,

ha lehet reaja kős benne
 ha lehet, reája kősD benne

hogj hamaréb meg fagadgjon.
 hogy hamarabb megfakadjon.

Az eredeti kézirat alapján kétféle digitális átíratot készítettünk, egy betűhű és egy normalizált változatot. Minkét esetben figyelembe vettük a már kialakult szövegközlési gyakorlatot, s követtük a már meglévő mintákat ott, ahol a nyelvemlékek 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; Dömötör et al. 2017).¹⁹

A betűhű átírat létrehozása során a nyelvészeti forráskiadás gyakorlatához igazodva (Benkő 1972) készítettünk el egy graféma- és ékezetű átíratot, amely bizonyos mértékben paleográfaiailag is hű (Bak 2000). A szövegűség 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 alkalmaztuk a szövegben három különböző betűvel jelölt s vagy a két variánsban előforduló z esetében is (lásd 4. és 5. ábra).



4. ábra
 A z graféma variánsai a szövegben

¹⁹ Ezúton is köszönetet szeretnék mondani Dömötör Adrienne-nek, aki rendelkezésünkre bocsátotta a *Középmagyar történeti magánéleti korpusz* átírási elveit, és konstruktívan segítette átírási munkánkat.



5. ábra
Az s graféma variánsai a szövegben

A szöveghűség elvét a nagy- és kisbetűk használata 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 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.²⁰ 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 szövszedetben adtuk meg.

(2) 227.

(Szem fiais Haljog ellenis embernek)
(Szemfájás hályog ellen is embernek)

Tjukmonjat kemennjen meg sütnj,
Tyúkmonyát* keményen megsütni,

azt kette vagnj,
azt kettévágni,

az sargaiat kj vetnj,
a sárgáját kivetni,

²⁰ A korábbi történeti korpuszok (pl. *Biblia Mediaevalis Hungarica*, *Történeti magánéleti korpusz*) tanulsága, hogy a bennük való keresést jelentős mértékben megkönnyíti, ha a szinkrón köznyelvi változatot is csatoljuk a betűhű átiratok mellé. A normalizált átiratok mellékelésének további előnye, hogy a nyelvtörténeti háttérrel nem rendelkező szakember, sőt akár még a laikus felhasználók számára is érthető lesz a szöveg, kereshető a korpusz.

helette egetet timsot beleie tennj,
 helyette égetett timsót beléje tenni,
 es be czjnalnj szösz köze,
 és becsinálni szösz közé,
 par hamuban tennj,
 pár hamuba'n tenni,
 egj tikmon sűtig ot tartanj,
 egy tyúkmony sűltig ott tartani,
 azutan tukmonj hajat levetnj
 azután tyúkmonyháját* levetni
 s meg tiztitanj, aprora szagatnj,
 s megtisztítani, apróra szagatni,
 Rosa Vízben hannj,
 Rózsavízben hagyni,
 azt az Rosa Vízet embernek szemeben czepegetnj
 azt a rózsavízet emberek szemébe'n csepegetetni
 Isten utan hasznal.
 Isten után használ.
 !!lat!Pro!
 !!lat!Pro!

A szöveg nem magyar nyelvű elemeit (pl. a latin kifejezéseket) jelölve benne hagytuk az átiratban, megértésükhöz szintén szöszedetet állítottunk össze. A magyar nyelvi elemeket a latin nyelviekől a toldalékolás alapján különítettük el a *Középmagyar történeti magánéleti korpusz* megoldását követve (Dömötör 2014; Dömötör et al. 2017). Egy jövevényszó akkor minősült magyarnak az adott szöveggörnyezetben, ha magyar nyelvi toldalékot kapott, így például az *ápiumot*, *mastixot*, *rusikát*, *tormentillát* kifejezések magyar nyelvűnek tekinthetők, míg a *Citrium*, *fenum graecum*, *mastix* lexémák nem.

- (3) **29.**
(!!lat!Contra Pestem ovium!!lat!)
(!!lat!Contra Pestem ovium!!lat!)

Az Diofanak berkejet /vegj/
 A diófának berkéjét /vedd/
 so kózot agj juhoknak.
 só között adD juhoknak.

A szöveg műfaji sajátosságai további, az olvasót segítő egyedi megoldásokat igényeltek az átirat készítése során. Az idegen nyelvi, tájnyelvi és archaikus szavakat tartalmazó szöszedetek mellett szükségesnek tűnt további, orvosi szakkifejezéseket tartalmazó szógyűjtemény megalkotása is, amelybe a betegség- és növénynevek, mértékegységek, eszközök, eljárások megnevezései kerültek be, például *glét*, *orbánc*, *madragfű*, *sérés*, *spikinárd olaj*, *vérállás*.

A betűhű átíráshoz hasonlóan a sérült szövegrészek a normalizált változatban is fel vannak tüntetve, megjegyzéssel és a lehetséges olvasattal ellátva. A jegyzetekben ezen felül a bizonytalan olvasatú helyek is jelezve vannak, szintén az összes lehetséges olvasattal ellátva. Az átírás ezen módja, illetve jelölése az alapja a nyelvtechnológiai feldolgozásnak, amelyre a következő alfejezetben térek ki.

3.3. Nyelvtechnológiai feldolgozás

A nyelvtechnológiának kulcsszerepe van abban, hogy a múlt kulturális örökségét széleskörűen elérhetővé és feldolgozhatóvá tegyék. A magyar nyelvű orvosi irodalom rendszerezett feldolgozásához olyan adatbázisra van tehát szükség, amelyben az elemek különféle nyelvészeti, történeti, paleográfiai információkkal vannak ellátva (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* projekt ómagyar korpusza (Simon-Sass 2012), illetve a *Középmagyar történeti magánéleti korpusz* (Dömötör 2014; Dömötör et al. 2017; Novák et al. 2018). Jelen alfejezetben az átíráshoz hasonlóan a *Mindenféle orvosságoknak rendszedése* egy receptjének feldolgozásán keresztül mutatjuk be a korpuszpépítés egyes lépéseit.²¹

Az új keresőfelület kialakításához a *Magyar nemzeti szövegtár* (Oravecz-Váradis-Sass 2014) és a *Magyar történeti szövegtár* (Sass 2017) újabb felületét, illetve a *Középmagyar emlékirat- és drámakorpuszt* (Gugán et al. 2023) vettük mintául. Ezen korpuszok keresőfelülete a Sketch Engine korpuszkezelő eszköz szabadon elérhető változatán, a NoSketch Engine-en (Rychlý 2007; Kilgarriff et al. 2014) alapul. Ahhoz, hogy a készülő korpuszban az egyes nyelvi jelenségek kereshetők legyenek (például ha latin szavakat, orvosi szakkifejezéseket vagy régi mértékegységeveket 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 központozási jelet (összefoglaló néven token) tartalmaz. Minden egyes szövegszó mellett az alábbi adatoknak kell szerepelniük: betűhű alak, normalizált alak, szótó, morfológiai elemzés (Simon-Sass-Novák 2013). Mindemellett szerepelhetnek egyéb mezők is (latin szó, betegségnév, 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. lókusztjelölő, amely azt adja meg, hogy az adott szövegszó pontosan hol található meg a dokumentumban. Mivel jelen esetben csak egy receptgyűjtemény feldolgozása történt meg, csupán egyetlen lókusztjelölő szám szerepel a táblázatban, az a sorszám, amellyel a digitális átíratok újra-

²¹ Az átírt anyag nyelvtechnológiai feldolgozása anyagi támogatás híján sajnos a mai napig nem valósult meg. Az itt ismertetett példa lényegében egy mutatónyelv.

számozása során az egyes recepteket elláttuk. Ugyanakkor a pontos azonosíthatóság érdekében a későbbiekben szükséges lehet többféle lókuszelőző használata, ez az egyes receptek felépítésétől függ; a lókuszelőzők tehát szövegenként változhatnak.

A későbbi kutatások szempontjából fontosnak tartjuk megőrizni azt az információt is, hogy 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 (lásd 3. táblázatban a *latint*).

Lókus	Lókus Lókus típusa	Típus	Nyelv	Betűhű alak	Normalizált alak	Lemma	Morfológiai elemzés
15	inic	disease		Az mely	Amely	amely	NOUN
15	inic	disease		embernek	embernek	ember	NOUN<CAS<DAT>>
15	inic	disease		az	a	a	ART
15	inic	disease		fogat	fogát	fog	NOUN<POSS><CAS<ACC>>
15	inic	disease		ki uonyszak	kivonják	kivon	VERB<PLUR><DEF>
15	inic	disease		,	,	,	PUNCT
15	inic	disease		és	és	és	CONJ
15	inic	disease		az	a	a	ART
15	inic	disease		vére	vére	vér	NOUN<POSS>
15	inic	disease		megh	meg	meg	PREV
15	inic	disease		nem	nem	nem	ADV
15	inic	disease		allana	állna	áll	VERB<COND>
15	inic	disease		.	.	.	PUNCT
15	instr			Narancs alma	Narancsalma	narancsalma	NOUN
15	instr			heját	héját	héj	NOUN<POSS><CAS<ACC>>
15	instr			,	,	,	PUNCT
15	instr			timsot	timsót	timsó	NOUN<CAS<ACC>>
15	instr			,	,	,	PUNCT
15	instr		latin		bonus	bonus	bonu NOUN[ATTRIB] ADJ
15	instr		latin	armenust	armenust	armenus	NOUN<CAS<ACC>>
15	instr			,	,	,	PUNCT
15	instr			ezeket	ezeket	ez	NOUN<PLUR><CAS<ACC>>
15	instr			forrald	forrald	forral	VERB<SUBJUNC-IMP> <PERS<2>><DEF>
15	instr			eőszve	össze	össze	PREV
15	instr			.	.	.	PUNCT

3. táblázat

A nyelvtéchnológiai feldolgozás szempontjai

Az egyes normalizált alakokhoz tartozó szótő (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 diskurzusműfajok közti keresést (akár külön alkorpuszokban). Természetesen az is cél, hogy az egyes kéziratokon belül lehetséges legyen keresések elvégzése: a betűhű alak szerint, illetve a normalizált alak szerint (teljes szóra, szókezdetre, szóvégre). Kereshetünk latin szavakat, betegségneveket, értelmezéseket is.

4. Összegzés

A *MOKO 16–17. századi alkorpusz*ának készítése folyamatban van. Idáig első-sorban a középmagyar receptirodalom kéziratos munkáival dolgoztunk, és jelenleg összesen hat receptgyűjtemény (kb. 5000–6000 recept) kétszintű átírása valósult meg. A feldolgozás során saját szabályzatban rögzítettük a betűhű és a normalizált átírás szabályait és a nyelvtchnológiai feldolgozáshoz szükséges jelrendszert. Ehhez figyelembe vettük a korábbi magyar nyelvű történeti, valamint a nemzetközi orvosi nyelvi korpuszok kialakítási és átírási elveit, ugyanakkor a diskurzustípushoz és a gyógyításhoz kötődő szaknyelvhez alakítottuk őket. A feldolgozás során a kéziratok alapján készítettük el a betűhű átiratot, amelyet tagmondatokra bontva, normalizálva is átirítottunk. Ezenkívül a betűhű átiratoknak egy egybefüggő, kritikai átírását is elkészítettük, valamint több, értelmezést segítő listát is összeállítottunk (archaizmusok, betegségek, mértékegységek, növénynevek, egyéb szakszavak, rövidítések), amelyek egy későbbi történeti orvosi szótár alapjául is szolgálhatnak. Az átírást többkörös, többszemélyes szigorú ellenőrzési munkálatok után dolgoztuk fel. A további munkálatok megvalósításához a szakmai háttér rendelkezésre áll, ugyanakkor ez egy időigényes, anyagi támogatás hiányában nagyon nehezen megvalósítható folyamat.

Összességében a *Magyar orvosi nyelvi korpusz* mint koncepció egy hosszú távú projekt, amely célként tűzi ki a magyar nyelvű orvosi irodalom rendszerezett feldolgozását. A munka számos hozadékkal bírhat, melyek közül az alábbiak már a munka jelen fázisában nyilvánvalóak:

- 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 diskurzusműfajok összekapcsolódásának, hálózatának a leírására. Ez önmagában az orvosi nyelv diskurzusműfaj szerinti állományának a feltárásához vezet, amely a szaknyelv összefüggő működésének teljesebb megismeréséhez is hozzájárul, továbbá a diskurzusműfaj-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 a korpuszban: 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).
- Megnyitja a nyelven belüli és nyelvek közötti összehasonlító elemzések útját, ezzel leírhatóvá válnak az egyes nyelvek és korok irányzatai, mintázatai, valamint különböző hatókörű jellegzetességek.
- A korpusz révén a magyar orvosi nyelv kutatása jobban beilleszthetővé válik a külföldön folyó vizsgálatokba, ami a nemzetközi együttműködésnek is előmozdítója.

Összességében megállapítható, hogy a történeti szövegek feldolgozása, korpuszba rendezése lehetővé teszi, hogy a történeti szövegeket nyelvhasználati eseményként és egyidőben produktumként értelmezzük. A korpusz kialakítása segíti továbbá a tudományközi együttműködések is (orvosi, orvostörténeti, néprajzi területekkel például). További fontos szempontként jelent meg, hogy a szövegtipológia hangsúlyosan jelenjen meg az orvosi történeti korpusz kialakításában, rávilágítva arra, hogy a beszédesemények, a diskurzusműfajok meghatározzák a nyelvhasználati jellegzetességeket a kommunikációban, továbbá arra is, hogy a pragmatikai szemlélet következetes alkalmazása a történeti szövegek esetében is megvalósítható, amelyekhez a korpuszok új módszertani lehetőségeket is kínálnak. Mindez pedig hozzájárul ahhoz is, hogy újabb és a korábbiánál átfogóbb ismeretekkel gazdagodhassunk a magyar orvosi szaknyelv kezdeti időszakáról.

Hivatkozások

- Alonso-Almeida, Francisco – Ivalla Ortega-Barrera – Elena Quintana-Toledo 2012. Corpus of Early English Recipes: Design and Implementation. In Nila Vázquez (szerk.) *Creation and Use of Historical English Corpora in Spain*. Newcastle upon Tyne: Cambridge Scholars Publishing. 37–49.
- Bak Borbála 2000. A XVI–XVIII. századi magyar nyelvű források kiadásának kérdései. *Fons (Forráskutatás és Történeti Segédtudományok)* 7/1:91–137.
- Barz, Irmhild – Ulla Fix – Marianne Schröder – Georg Schuppener (szerk.) 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*. Budapest: Magyar Nyelvtudományi Társaság.
- Csontos Nóra 2013. A pragmatika mint szemléletmód érvényesítésének lehetősége a történetiségben. *Magyar Nyelv* 109/2:149–157.

- 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 Magyarágtudományi Társaság. 11–21.
- Dömötör Adrienne – Gugán Katalin – Novák Attila – Varga Mónika 2017. Kiútkeresés a morfológiai labirintusból: Korpuszépítés ó- és középmagyar kori magánéleti szövegekből. *Nyelvtudományi Közlemények* 113:85–110. doi:[10.15776/NyK.2017.113.3](https://doi.org/10.15776/NyK.2017.113.3)
- Eckkrammer, Eva Martha 2016. *Medizin für den Laien. Vom Pesttraktat zum digitalen Ratgebertext*. Berlin: Frank & Timme.
- Görlach, Manfred 2004. *Text Types and the History of English*. (Trends in Linguistics. Studies and Monographs 139) Berlin; New York: De Gruyter Mouton. doi:[10.1515/9783110197167](https://doi.org/10.1515/9783110197167)
- Gugán Katalin – Bácsi Enikő – Dömötör Adrienne – Horváth László – Mohay Zsuzsanna – Varga Mónika – Sass Bálint 2023. *Középmagyar emlékirat- és drámakorpusz*. <https://ked.nytud.hu/>
- Heinemann, Wolfgang 2000. Textsorte - Textmuster - Texttyp. In Klaus Brinker – Gerd Antos – Wolfgang Heinemann – Sven F. Sager (szerk.) *Text- und Gesprächslinguistik 1. Halbband*. Berlin; New York: De Gruyter Mouton. 507–523. doi:[10.1515/9783110194067-050](https://doi.org/10.1515/9783110194067-050)
- Hoffmann Gizella (szerk.) 1989. *Medicusai és borbélyi mesterség. Régi magyar ember- és állatorvosló könyvek Radvánszky Béla gyűjtéséből*. Szeged: MTA Irodalomtudományi Intézet.
- Keresztes Csilla – Borda Bernadett – Vincze Veronika 2019. Mi derül ki a SZEMEK-ből? Egy magyar nyelvű orvos korpusz morfológiai sajátosságai. *Porta Lingua* 2019:97–109.
- Kilgarriff, Adam – Vít Baisa – Jan Bušta – Miloš Jakubíček – Vojtěch Kovář – Jan Michelfeit – Pavel Rychlý – Vít Suchomel 2014. The Sketch Engine. *Lexicography* 1/1:7–36. doi:[10.1007/s40607-014-0009-9](https://doi.org/10.1007/s40607-014-0009-9)
- Kniezsa István 1952a. *A magyar helyesírás története*. Budapest: Magyar Tankönyvkiadó Vállalat.
- Kniezsa István 1952b. *Helyesírásunk története a könyvnyomtatás koráig*. Budapest: Akadémiai Kiadó.
- Kocsány Piroska 2002. *Szöveg, szövegtípus, jelentés: a mondás mint szövegtípus*. (Nyelvtudományi Értekezések 151) Budapest: Akadémiai Kiadó.
- Kocsány Piroska 2006. A szövegtipológia eredményei és/vagy eredménytelenségei. In Tolcsvai Nagy Gábor (szerk.) *Szöveg és típus. Szövegtipológiai tanulmányok*. (Segédkönyvek a nyelvészet tanulmányozásához 63) Budapest: Tinta Könyvkiadó. 17–26.
- Kohnen, Thomas 2004. Methodological problems in corpus-based historical pragmatics. In Karin Aijmer – Bengt Altenberg (szerk.) *Advances in Corpus Linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23)*. Amsterdam; New York: Rodopi. 237–247.

- Korompay Klára 2003. Helyesírás-történet. In Kiss Jenő – Pusztai Ferenc (szerk.) *Magyar nyelvtörténet*. Budapest: Osiris Kiadó. 101–102, 281–299, 579–593.
- Kuna Ágnes 2009. Az orvosi vény alakulása a kezdetektől napjainkig. In Balázs Géza – H. Varga Gyula (szerk.) *Semiotica Agriensis 6. konferenciakötet*. Eger: Líceum Kiadó. 401–410.
- Kuna Ágnes 2011. *A 16–17. századi magyar nyelvű orvosi recept szövegtipológiai és pragmatikai vizsgálata funkcionális kognitív keretben*. PhD disszertáció. Budapest: Eötvös Loránd Tudományegyetem Bölcsészettudományi Kar.
- Kuna Ágnes 2016a. A Magyar orvosi nyelv korpusza. *Magyar Orvosi Nyelv* 16/1:26–30.
- Kuna, Ágnes 2016b. Genre in a cognitive functional framework. Medical recipe as a genre in 16th and 17th century Hungarian. In Ninke Stukker – Wilbert Sporen – Gerard Steen (szerk.) *Genre in Language, Discourse and Cognition*. (Applications of Cognitive Linguistics 33) Berlin; Boston: De Gruyter Mouton. 193–224.
- Kuna Ágnes 2017. Történeti szövegtipológia. In Forgács Tamás – Németh Miklós – Sinkovics Balázs (szerk.) *A nyelvtörténeti kutatások újabb eredményei IX*. Szeged: Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék. 223–237.
- Kuna Ágnes 2019. *Nyelv, meggyőzés, gyógyítás. A meggyőzés nyelvi mintázatai a 16–17. századi orvosi receptben*. (Segédkönyvek a nyelvészet tanulmányozásához 210) Budapest: Tinta Könyvkiadó.
- Kuna Ágnes 2024a. *Fejezetek a receptirodalom történetéből. A 16–17. századi magyar orvosi recept hagyomány nyelvészeti megközelítése*. Budapest: Szépirodalmi Figyelő Alapítvány.
- Kuna Ágnes 2024b. *Az orvos-beteg kommunikáció nyelvészeti elemzése. Elmélet, gyakorlat, módszer*. Budapest: ELTE Eötvös Kiadó.
- Kuna Ágnes – Kocsis Zsuzsanna – Ludányi Zsófia 2017. A Magyar orvosi nyelv 16–17. századi alkorpusza. Tervezet, átírás, annotálás. In Forgács Tamás – Németh Miklós – Sinkovics Balázs (szerk.) *A nyelvtörténeti kutatások újabb eredményei IX*. Szeged: Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék. 239–253.
- Kuna Ágnes – Ludányi Zsófia 2019. Terminológiai elvek az orvosi szaknyelvben és a gyógyító kommunikációban. Problémák, tendenciák, ideológiák. In Fóris Ágota – Bölcskei Andrea (szerk.) *Terminológiasztratégiai kihívások a magyar nyelvterületen*. Budapest: L'Harmattan. 105–120.
- Kuna Ágnes – Simon Gábor 2017. Műfaj, szövegtípus, szövegfajta. Nézőpontok, kategóriák, modellek a szövegnyelvészeti kutatásban. *Magyar Nyelv* 113/3:257–275.
- Maitz Péter – Molnár Anna 2001. Nyelvtörténetírás és történeti szövegnyelvészet. A szöveg helye és a szövegkutatás feladatai a történeti nyelvészetben. In Csátár Péter – Maitz Péter – Tronka Krisztián (szerk.) *A nyelvtantól a szövegtanig. Tanulmányok Kocsány Piroska tiszteletére*. Debrecen: Kossuth Egyetemi Kiadó. 322–336.
- Novák, Attila – Katalin Gugán – Mónika Varga – Adrienne Dömötör 2018. Creation of an annotated corpus of Old and Middle Hungarian court records and private correspondence. *Language Resources and Evaluation* 52/1:1–28. doi:[10.1007/s10579-017-9393-8](https://doi.org/10.1007/s10579-017-9393-8)

- Oravecz, Csaba – Tamás Váradi – Bálint Sass 2014. The Hungarian Gigaword Corpus. In Nicoletta Calzolari – Khalid Choukri – Thierry Declerck – Hrafn Loftsson – Bente Maegaard – Joseph Mariani – Asuncion Moreno – Jan Odijk – Stelios Piperidis (szerk.) *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik: European Language Resources Association (ELRA). 1719–1723.
- Riecke, Jörg 2004. *Die Frühgeschichte der mittelalterlichen medizinischen Fachsprache im Deutschen*. Berlin; New York: Walter de Gruyter. doi:[10.1515/9783110895223](https://doi.org/10.1515/9783110895223)
- Robin Edina – Dankó Szilvia – Götz Andrea – Nagy A. Lilla – Pataky Éva – Szegh Henriette – Zolczer Péter 2016. Fordítástudomány és korpuszkutatás: bemutatkozik a Pannonia Korpusz. *Fordítástudomány* 18/2:5–26.
- Rychlý, Pavel 2007. Manatee/Bonito – A modular corpus manager. In Petr Sojka – Aleš Horak (szerk.) *RASLAN 2007. Recent Advances in Slavonic Natural Language Processing*. Brno: Masaryk University. 65–70.
- Sass Bálint 2017. Keresés korpuszban. A kibővített Magyar Történeti Szövegtár új keresőfelülete. In Forgács Tamás – Németh Miklós – Sinkovics Balázs (szerk.) *A nyelvtörténeti kutatások újabb eredményei IX*. Szeged: Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék. 267–277.
- Simon Eszter – Sass Bálint 2012. Nyelvtechnológiai és kulturális örökség, avagy korpuszépítés ómagyar kódexekből. In Prószéky Gábor – Váradi Tamás (szerk.) *Általános Nyelvészeti Tanulmányok XXIV. Nyelvtechnológiai kutatások*. Budapest: Akadémiai Kiadó. 243–264.
- Simon Eszter – Sass Bálint – Novák Attila 2013. *Kódozási szabályzat*. Kézirat. Budapest.
- Steger, Hugo 1984. Sprachgeschichte als Geschichte der Textsorten/Texttypen und ihrer kommunikativen Bezugsbereiche. In Werner Besch – Oskar Reichmann – Stefan Sonderegger (szerk.) *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*. Berlin; New York: Walter de Gruyter. 186–204.
- Szlatky Mária 1980. Magyar nyelvű, kéziratos orvosló könyvek a 17. századból. *Magyar Könyvszemle* 96/2:131–148.
- Taavitsainen, Irma 2001. Changing conventions of writing: The dynamics of genres, text types, and text traditions. *European Journal of English Studies* 5/2:139–150. doi:[10.1076/ejes.5.2.139.7309](https://doi.org/10.1076/ejes.5.2.139.7309)
- Taavitsainen, Irma – Turo Hiltunen (szerk.) 2019. *Late Modern English Medical Texts. Writing medicine in the eighteenth century*. Amsterdam: John Benjamins Publishing Company. doi:[10.1075/z.221](https://doi.org/10.1075/z.221)
- Taavitsainen, Irma – Päivi Pahta (szerk.) 2004. *Medical and Scientific Writing in Late Medieval English*. Cambridge: Cambridge University Press.
- Taavitsainen, Irma – Päivi Pahta – Turo Hiltunen – Martti Mäkinen – Martilla Ville – Maura Ratia – Carla Suhr – Jukka Tyrkkö 2010. *Early Modern English Medical Texts*. CD-ROM. Amsterdam: John Benjamins Publishing Company.
- Taavitsainen, Irma – Päivi Pahta – Martti Mäkinen 2005. *Middle English Medical Texts*. CD-ROM. Amsterdam: John Benjamins Publishing Company.

- Tolcsvai Nagy Gábor 2006. A szövegtipológia megalapozása kognitív nyelvészeti keretben. In Tolcsvai Nagy Gábor (szerk.) *Szöveg és típus. Szövegtipológiai tanulmányok.* (Segédkönyvek a nyelvészet tanulmányozásához 63) Budapest: Tinta Könyvkiadó. 64–90.
- Varga Éva Katalin – Márton Emese – Indig Balázs – Sárközi-Lindner Zsófia – Palkó Gábor 2023. Erdélyi és anyaországi orvosi terminológia pandémia idején. *Alkalmazott Nyelvtudomány* 23/1:76–95. doi:[10.18460/ANY.K.2023.1.004](https://doi.org/10.18460/ANY.K.2023.1.004)
- Vincze, Veronika – György Szarvas – Richárd Farkas – György Móra – János Csirik 2008. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics* 9/S11:S9. doi:[10.1186/1471-2105-9-S11-S9](https://doi.org/10.1186/1471-2105-9-S11-S9)
- Voigts, Linda Ehsam – Patricia Deery Kurtz 2001. *Scientific and Medical Writings in Old and Middle English: An Electronic Reference.* CD-ROM. Ann Arbor: University of Michigan Press.

A szerzőről

Kuna Ágnes az Eötvös Loránd Tudományegyetem Alkalmazott Nyelvészeti és Fonetikai Tanszékének tanszékvezető habilitált egyetemi docense. Kutatásaiban az orvosi nyelv jelenkori és történeti vonatkozásait vizsgálja.

Elérhetősége: kuna.agnes@btk.elte.hu

