

# A DIGITÁLIS MIKES-SZÓTÁR

Kiss Margit

PhD, tudományos munkatárs,  
MTA Irodalomtudományi Intézet  
kiss.margit@iti.mta.hu

*„Nem szükségeszerü, hogy kibékíthetetlen ellentét legyen a racionalista és az empirikus nézőpont között; ma már sokan tudják: adaton nem a nyelvész noteszába beírt privát bejegyzést kell érteni, hanem azt a nyersanyagot, amelyen állnak vagy buknak az elméletek.”*

(Dömötör, 2011, 49)

Tanulmányomban Mikes Kelemen életművének folyamatban lévő szótári feldolgozásáról adok helyzetképet.\* Egy olyan munkálat részleteit mutatom be, amelyben a modern számítógépes eljárások egyesülnek a hagyományos filológiai eszközökkel, s amelyben hasznosítani igyekszünk a már elkészült vagy készülöben lévő külföldi elektronikus szótárak tanulságait.

Már az 1900-as évek elején szótárszerűen feldolgozták Mikes egyes szavait: 1906-ban Szily Kálmán néhány tucat szócikket közöl a *Törökországi levelek* szóállományából. Egy Mikes-szótár elkészítését, az összes művek kritikai kiadásának befejezését követően, Hopp Lajos is tervezte, de realitássá ez az ő életében már nem válhatott. Mikes halálának 250. évfordulójához közeledve azonban az MTA Irodalomtudományi Intézetének XVIII. századi osztályán OTKA-pályázat keretében 2010 áprilisában elkezdődött a munka, amely-

nek célja Mikes Kelemen teljes szókincsének feldolgozása elektronikus írói szótár formájában. A kutatás vezetője Tüskés Gábor; Horváth Katalin és Szathmári István szakmai tanácsadóként segíti a munkát, a lexikológiai feladatokat e cikk szerzője végzi.

A magyar szótárirodalomban több írói szótár áll rendelkezésünkre: többek között a Balassi-, a Zrínyi-, a Petőfi-, és a Juhász Gyula-szótár. A kötetek terjedelmi kötöttségek és a kezelhetőség szempontja miatt – a Petőfi-szótár kivételével – nem tartalmazzák a feldolgozott korpusz teljes anyagát; a jelentéseket példázó idézetek és a további előfordulások száma egyaránt erősen korlátozott. Emellett szükségessé vált számos további, terjedelmet csökkentő megoldás és bonyolult utalórendszer kialakítása. A most készülő Mikes-szótárban az elektronikus feldolgozás lehetőségeinek köszönhetően nem korlátozzuk a jelentéseket illusztráló idézetek számát, s lehetőség nyílik a szavak, szókapcsolatok összes előfordulásának visszakérésére.

Ezzel nemcsak mennyiségileg, hanem minőségileg is új, ma még csak részben ismert lehetőségek nyílnak meg mind a történeti nyelvészeti, mind a stílus-, irodalom- és eszmetörténeti kutatások előtt. Ez egyedülálló alkalmat ad nemcsak a nyelvhasználat és a stílus sajátosságainak, valamint e sajátosságok változásának megfigyelésére, hanem elősegí-

\* Készült a K 81.337 sz. OTKA-pályázat keretében.

ti az író világképét érintő összetettebb kérdések megválaszolását is. Különösen fontos e kutatás nyelv- és szótörténeti jelentősége, hiszen az 1772 előtti időszak történeti nyelvészeti, szótörténeti-frazeológiai feldolgozottsága nagyon hiányos. A munka hiánypótló mind az anyag mennyisége, mind filológiai igényessége szempontjából, amely nemcsak a Mikes-kutatás számára hozhat új eredményeket, hanem a középmagyar kor nyelvész kutatóinak is hiteles és bő forrásanyagot nyújt. A *Mikes-szótár* a már elkészült *Zrínyi-szótárral* együtt két egymást követő évszázad irodalmi nyelvéről adhat átfogó képet két kiemelkedő író munkásságán keresztül.

A szótár alapjául az 1966–1988 között megjelent kritikai kiadás szolgál, amely hat kötetben mintegy 6000 oldalnyi nyomtatott szöveget tartalmaz. Nagy értéke a leendő szótárnak – különösen, ha figyelembe vesszük az eddigi magyar nyelvű írói szótárakat –, hogy a teljes életművet a kritikai kiadás alapján dolgozzuk fel. Ezzel a filológiai, nyelvészeti, esztétikai és más jellegű kutatásoknak is megbízható elektronikus szövegbázisát teremthjük meg.

A munka főbb állomásai a következők: a szövegek elektronikus rögzítése, ellenőrzése; a címszólista elkészítése; előszerkesztés; a jelentésszerkezet kialakítása; a címszóval alkotott szókapcsolatok bemutatása.

A szótárkészítésben korábban alkalmazott hosszadalmas és a kézi másolás miatt pontatlan cédulázást mára felváltotta számos gyors és hatékony számítógépes szövegfeldolgozási eszköz. A számítógép a munkálatok több fontos és időigényes részében ad megbízható segítséget, amilyen például a szövegek elektronikus rögzítése. Annak érdekében, hogy a nyomtatott szövegek a számítógépen is rendelkezésünkre álljanak és szerkeszthetők le-

gyenek, munkacsoportunk a kötetek szkennelését követően az *Abby FineReader* szövegfelismertető szoftvert alkalmazta, majd korrektúrázta a felismertetett szövegeket.

A szövegfelismertető szoftver beállításait sokáig teszteltük, mire elértük az optimális eredményt, az oldalankénti négy-öt hibát. A régi szöveg sajátos központozása, helyesírási egyenetlensége, a névelővel egybeírt szóalakok, az egybe- és különírás variánsai, a kis- és nagybetűk maitól eltérő használata mind nehezítették a szövegfelismertetést. A korrektúra során a 6000 oldalnyi szöveget összeolvastuk a felismertetett szöveggel, s javítottuk a hibákat. A nyomtatásban is olvashatatlannak, illetőleg nyomdahibás szövegrészleteket a kéziratokból és mikrofilmről javítottuk.

2011 eleje óta rendelkezünk Mikes műveinek betűhív átiratával elektronikus formában, amely a MEK honlapján mindenki számára elérhető (URL1). A szövegek elektronikus rögzítését a MEK-kel kötött együttműködési megállapodás keretében végeztük.

A szótári munkában nélkülözhetetlen, hogy a rendelkezésre álló szövegcsoportnak ne csak egy-egy adott szavára tudjunk keresni, hanem az ugyanolyan funkcióban álló, a szócikkíráshoz szükséges tartalmi elemeket is ki tudjuk gyűjteni, és csoportosíthatjuk ezeket. E célból a szövegbe különféle jelölőelemeket, ún. *tag*-eket, címkeket illesztettünk be. A következő egységeket címkéztük: versbetétek, címek, idegen nyelvű szövegrészletek, margináliák, rövidítések, idézetek, fordításrészletek. Példán szemléltetve:

„Az első példa ajudás két fiáról vagyon; ennek apátriárkának négy fia volt. a kik közül. akét első igen gonosz, és rosz erkölcsü volt. ihon mit mond az írás az elsőtől, Heröl. **{foreign}***fuit quoque Her primogenitus juda. nequam in Conspectu Domini., et ab eo occisus*

*est, {/foreign} {translated lang=„latin”} Herjudásnak az első fia, igen gonosz volt Isten előtt. meg is ölte ötet, {/translated}* amásodikárol, onánrol. mindgyárt azután azt mondgya az irás, hogy az Isten meg verte ötet halálal, mert utálatos vétket tselekedet., {foreign} *Id circo percussit eum Dominus, eo quod rem detestabilem faceret, {/foreign}* ez avétek, noha oly utálatos Isten előtt., és oly nagy büntetéssel büntetetik meg, mindazon által. igen közönséges az ifiak közöt. amelyért gyakorta szál reájok az Isten haragja. látható, vagy láthatatlan képen.”

A korrektúrázott és címkékkel ellátott szövegből adatbázist készítettünk. Ez a rendszer már önmagában is többszemponú vizsgálatra alkalmas. Kereshetünk egy szó összes előfordulására, de kijelölhetjük azokat a műveket, amelyekben keresni szeretnénk. Akár több szó kapcsolatát is megjelölhetjük az „és”, „vagy”, „kizáró” stb. logikai formákkal. Azzal, hogy Mikes összes műveiből elektronikus adatbázist készítettünk, eddig kiaknázatlan vizsgálati lehetőségek nyílnak meg, s új eredmények várhatók a szövegkutatás terén.

Az adatbázis mellett rendelkezésünkre állnak különféle gyakorisági listák is. Ezeket az egységes címkézésnek köszönhetően tudtuk létrehozni. A különböző típusú szóstatistikákból megtudhatjuk például, hogy az ún. tartalmatlan szavakon (például: *a, az, és, hogy* stb.) túlmenően a tartalmas szavak közül Mikes legtöbbször a *szent* szót használta, rögtön ezt követi a gyakorisági sorban az *Isten*, de a sor elején állnak még a *nagy, Kristus* szavak is. A gyakorisági értékeket a művek szerint is megkaphatjuk. Elkészült a konkordancia-lista, ami a szótárkészítés alapja. Ebben a szövegtörzs a ragozott szóalakokra bontva található, s a ragozott alakokat rövid szövegkörnyezet illusztrálja. Így például:

*bánhatik*–3 (ragozott alak, előfordulások száma) / és hogy ezekel is úgy *bánhatik* az Isten, (TL 71) / hogy minden jóvainkal úgy *bánhatik* valamint néki tettzik, (KJÉ 634) / azzal szabad, és azzal szabadosan *bánhatik*, (IJE 110) (A példamondatokat zárójelben a műcím rövidítése és az oldalszám követi.)

*bánhatnak* – 1 / midön az emberekel gonoszul nem *bánhatnak*, arra kérék akristust. (KJÉ 634)

A digitálisan rögzített szövegekből további előzetes számadatok is kiolvashatók. Így például a szövegszavak (Mikes összes művének összes szava) száma: kb. 1,5 millió. Ehhez összehasonlításként érdemes megemlíteni, hogy *A magyar nyelv nagyszótárának* számítógépes szövegtörzs kb. 25 millió szövegszó, a készülő József Attila-szótár anyaga kb. 340 ezer szövegszó. Ha a teljes konkordancia-listát kinyomtatnánk, az kb. 50 ezer oldalnyi terjedelmet tenne ki. A különböző típusú ragozott alakok (szóalakok) száma: 162 ezer. Az 50 feletti előfordulással rendelkező szóalakok száma: 1352. Eddigi becsléseink alapján összesen kb. 10–15 ezer címszó lesz majd a szótárban. Ha ezt a mennyiséget összevetjük a szövegszavak számával, láthatjuk, hogy az egy címszóhoz tartozó előfordulások száma viszonylag nagy. Érdemes azonban megjegyezni, hogy jelentős különbségek vannak a szavak gyakorisági értékei között. Tudjuk például, hogy összesen 258 *be* igekötős címszó lesz a szótárban.

A következő s egyben jelenleg is folyó munkafázis a címszólista-készítés és előszerkesztés. A címszólista készítésének fő szempontja az egy címszóhoz sorolandó szavak rendszerszerű csoportosítása. A teljesség elvéhez híven nem válogatunk a címszavak között, s minden mikesi szó önálló címszó lesz mai formájában. A címszólista elkészítéséhez a

konkordancialistát használjuk fel. A konkordancialista a ragozott szóalakokból és a hozzájuk tartozó összes szóelőfordulásból áll. Minden szóelőforduláshoz egy mondatnyi szöveggörnyezet tartozik. A címszavasítás menete kötött: az azonos ragozási formájú, toldalékolású szóelőfordulások alkotják a szóalakot. A szóalakokból meghatározzuk az alak- vagy írásváltozati címszót, amelyből több is lehet, majd ezekből megalkotjuk a szótári címszót. Példával illusztrálva:

**háború** – szótári címszó

**haboru** – alak- vagy írásváltozati címszó

**haborutul** – szóalak

Gabriel menyből. Üdvözle Istenöl., Öriz *haborutul*. Évától vett jajtul (KG 555) – [szóelőfordulás a rövid szöveggörnyezettel, zárójelben a forrásmegjelöléssel (KG 555 = Keresztényi Gondolatok 555. o.)]

Az eljárás viszonylag egyszerű, ám az említett kb. 50 ezer oldalnyi anyag összes példamondatát át kell nézni, s a megfelelő helyre tenni. Az erre a munkafolyamatra kidolgozott ún. lemmatizáló szoftverek (pl. *WordCruncher*) nagy hibahatárokkal dolgoznak, s adaptálásuk számos nehézségbe ütközött volna, ezért elvetettük ezek alkalmazását. Külön nehézséget jelentenek a maitól eltérő formák, például a különírt igekötők (*meg halván*), a különírt toldalékok (*leg csudálatosab*), az összetett igei szerkezetek (*meg vala mondva*), a névelővel egybeírt névszók (*ahit*), valamint a régies névszói egybe-, illetve különírási módozatok (*dus gazdag*) stb. Azért van szükség erre a munkára, mert egy-egy szónak nagyon sok alak-, illetve írásváltozata van, amelyeket a korpuszban nem, vagy csak hosszas kereséssel találunk meg. A módszer előnye, hogy a címszavasítás mellett egyben előszerkesztés is folyik, mivel az egy címszóhoz tartozó összes alak- és írásváltozat, különböző toldalékolású

forma az ezeket tartalmazó példamondatokkal együtt már most egy szócikkben található. A különböző státuszú szavak (címszó, alakváltozatok, szóalakok stb.) típusonkénti címkézése lehetővé teszi, hogy a teljes szöveggörnyezetről e funkciók alapján ki tudjuk gyűjteni és rendszerezni tudjuk a szavakat.

A címszavak mai alakjának meghatározásában segítségünkre vannak az értelmező szótárak, a tájszótárak, a helyesírási szótárak, az idegen szavak szótárai, valamint különféle vallástörténeti és szaklexikonok. Ha az adott címszó más szótárban (beleértve a kb. 25 millió szövegszavas nagyszótári számítógépes szöveggörnyeztet) nem fordul elő, azt külön megjelöljük. Ezek a szavak alkotják az igazi nyelvtörténeti, lexikográfiai kincseket: Mikés ún. nem szótárazott szavai, mint például a *dívánvezér*, *dölnékezik*, *drágafü*. Más nyelvtörténeti újdonságok is várhatók: a *gimnázium* szónak például az etimológiai szótárakban 1787-es az első előfordulása, Mikestől viszont már 1754-ből van rá adatunk.

A szókészletről készülő különféle mutatók lehetővé teszik az összevetést a korábbi és későbbi korok szövegeiből készült hasonló típusú összeállításokkal. A mutatók révén többek között új adatokhoz juthatunk az összehasonlított szerzők egyéni nyelvének szókészletéről és a magyar szókincs változásáról. Már a munkának ebben a fázisában megtudhatjuk, hogy az életműben az összes szóhoz mérten milyen az egyes szavak gyakorisági (akár művenkénti) eloszlása, és Mikés nyelve mennyire tér el a köznyelvitől. Példán szemléltetve: a *dohány*, *dohányos*, *dohányoz*, *dohányozás*, *dohányozhat*, *dohányozni*, *dohányzik* szócsalád egyedül a *Törökországi levelek*-ben fordul elő. Ugyanakkor a *dorbézolás*, *dorgálás* és ezek egyéb alakjai csak valóságos tárgyú művekben szerepelnek. Megvizs-

gálhatjuk, hogy konkrét szavak hogyan jelennek meg az összetételekben és a szóképzésben: például *császár*, *császármé*, *császárság*, *császárválasztás*. Ezek alapján megfigyelhetjük az ugyanabba a szócsaládba tartozó alakok művenkénti eloszlását, például a *döglött* a *Törökországi levelek*-ben fordul elő, míg a *dögbalál* a vallásos művekben. Kigyűjthetjük a címszavak alak- és írásváltozatait, mint például a *Galilea* tizenhat alakváltozatát: *Galilaea*, *galilaea*, *Galilaéa*, *galilaéa*, *Galilea*, *galilea*, *Galiléa*, *galiléa*, *galliaéa*, *Gallilaea*, *gallilaea*, *Galliláea*, *Gallilea*, *gallilea*, *Galliléa*, *galliléa*. A szókinszstatistikából kiderül, hogy a már-már tartalmatlannak tartott *dolog*, *cselekszik*, *cselekedni* családja népes, míg a *csinál* arányaiban jóval szerényebb az előbbiekhöz képest. Ugyanakkor a *dicsőség*, *dicsőséges* kifejezeten nagy számban fordul elő a mikesi szókinszben. Elemezhető a szinonimák használata: például a *csecsemő*, *csecsszopó* közel azonos eloszlású a mikesi életműben. A készülő szótár alapul szolgálhat eszme- és művelődéstörténeti kutatásoknak. Megvizsgálhatók a szavak többek között az alábbi területekről: étkezés, például *csokoládé*, *csokoládécsinálás*, *dinnye*, *dió*, *cseresznye*; egyházi terminológia, például *dalmatika*, *Dominus Vobiscum*; történelmi szakkifejezések, például *díván*; helységnevéhasználat, például *dantzka*. Értékes adatokat kaphatunk az idegen szavak használatáról, például *deskripció*, *dialektika*, *diárium*, *diskurzus*, *doktrína*.

A szóösszetételei határok jelölése révén nemcsak az összetett szavakra, hanem az elő- és utótagok szerint is kereshetünk a címszavak között. Lehetőség nyílik arra, hogy mélyrehatóan megvizsgáljuk az összetételeket és szerkezetűpusaikat. Helyesírás-történeti szempontból elemezhető az alak- és írásváltozatok eloszlása, a mondattagolás alakulása.

Már ebből a vázlatos áttekintésből is kirajzolódik Mikes tipikus szókapcsolatainak a köre. Ezek módszeres vizsgálatával közelebb juthatunk nyelvezetének megismeréséhez, például *gondolván + lenni*, jelző + *gondviseléssel*, *gyengeségekre/gyengeségére való nézve*.

A további munka során először a címszavakba rendezett anyagot fogjuk elektronikus formában közzétenni. A tervezett megoldás egy többszortatú felületen, a címszó, az alakváltozatok, a szóalakok, a szóelőfordulások és a bővebb szövegkörnyezet bemutatásával, számadatokkal teszi könnyebbé a további feldolgozást. A szóanyag az imént felsorolt kategóriák szerint lesz csoportosítható, s a különböző keresési lehetőségekkel önmagában is egyedülálló forrásbázist kínál a további kutatásoknak. A munka eredményeként a 1,5 millió szavas mikesi életmű részleteiben és egészében is könnyebben kezelhetővé válik. Példa:

**Címszókeresés:** *dézsma*\*

**Találatok:** *dézsma*, *dézsma**bor*, *dézsma**l*

**Alakváltozat(ok):** *dézsma* *bor*

**Szóalak(ok):** *dézsma* *bor**l*

**Szóelőfordulás(ok):** hagymárol, és a *dézsma* *bor**l* kívánna irni a feleséginek (TL 40)

Az eddigi eredmények alapján létrehozunk egy kísérleti weboldalt, ahová folyamatosan fogjuk feltölteni az elkészült anyagokat. Mivel az internetfelhasználók többnyire nemcsak passzív befogadói a weboldalakon lévő információknak, hanem észrevételeikkel maguk is bővíthetik az olvasottakat, a Mikes-szótár internetes változatában mi is ezt szeretnénk alkalmazni: a szócikkeknel lehetőség lesz arra, hogy a felhasználók szakmai megjegyzéseket tehesenek. Például egy adott szónál utalhatnak a hozzá kapcsolódó szótörténeti forrásmunkákra, folyóiratcikkekre, de más tudományterületről származó szakiro-

dalmi hivatkozásokat vagy egyéb szakmai megjegyzéseket is feltüntethetnek. Az ilyen típusú megjegyzések nem férnek bele egy zárt szócikkstruktúrába, de a megfelelő helyen közölve alaposabb vizsgálatra hívhatják fel a figyelmet, s bővíthetik ismereteinket. Mikes szöveganyagából kiindulva ezáltal egy átfogó, több tudományterületet érintő, a szűkebb értelemben vett szótáron messze túlmutató, sokoldalú forrásbázis jöhet létre.

A szótár a fentiekén túl kiegészül a jelentések bemutatásával és a frazeológiával. A címszavakra és alakváltozataikra egyaránt kereshetünk, s megismerhetjük a jelentésszerkezetet az összes példamondattal együtt, ahonnan a teljes korpusz is elérhető. Végül megtalálhatjuk a szócikkvégi utalásokat és a frazeológia blokkját. Példa:

**Címszókeresés:** *galamb*\*

**Találatok:** *galamb*, *galambárus*, *galambtojás*

**Címszó, alakváltozat(ok):** *galamb|árus fn*  
2 *galamb arus* 2

**Szóalakok:** *galamb arusok* 2

## IRODALOM

Dömötör Adrienne (2011): Nyelvtörténet, nyelvváltozat, adatbázis. In: Hegedűs Orsolya – Psenáková Ildikó (szerk.): *Tudomány az oktatásért – oktatás a*

**Jelentések, példamondatok:** ,az a személy, aki *galambokat* árul, azokkal kereskedik': / apénz váltok asztalit, és a *galamb arusok* székeit, fel fordítá. (É 195) / fel fordítá a pénz váltok asztalit. és *agalamb arusok* székeit. (KJÉ 721)

Mikes életműve ma már nem, vagy csak nehezen érthető nyelvi magyarázatok nélkül, s nyelvének értékei igen gyakran rejtve maradnak az olvasó előtt. A szótár hozzájárulhat a XVIII. század első felét felölelő időszak nyelvezetének szótörténeti-etimológiai, alak- és mondatörténeti, dialektológiai, stilisztikai és irodalmi nyelvi vizsgálatához, mivel az elektronikus feldolgozásnak köszönhetően teljes egészében elének állítja Mikes életművét és nyelvi hagyatékát. A szótár az életmű több szempontú feldolgozásához adhat hatékony segítséget, s egyben segédeszköze lehet a magasabb szintű középiskolai és egyetemi oktatásnak.

**Kulcsszavak:** *Mikes, szótár, korpusz, szó, elektronikus, nyelv, internet*

*tudományért.* I. Univerzita Konstantína Filozofa v Nitre, Fakulta stredoeurópskych štúdií, Nitra  
URL: <http://mek.oszk.hu/09000/09000/index.phtml>

