

Számítógép és írói szótár – különös tekintettel a készülő József Attila-szótárra*

1. Az elmúlt ötven év magyar írói szótárai. Az írói szótár műfaji előzményei egészen az ókorra nyúlnak vissza (az alábbiakban – a nyelvtechnológiai nézőpont okán – nem teszünk különbséget az egy művet, a műcsoportot, illetve az életművet feldolgozó írói szótárak között, noha ennek a filológiai jelentősége alapvető). A két fő műfaji forrás a konkordancia és a glosszárium volt (vö. GÁLDI 1965a: 369–370; BENKŐ 1979: 13–19, 64–68; KARPOVA 2011: 1–7). A konkordanciák célja a tájékozódás és a keresés volt a szövegben (pl. a Bibliában), a glosszáriumok az olvasókat segítették speciális szövegeknek (pl. antik szerzők műveinek) az olvasásakor. E két funkció grammatikai, szemantikai, stilisztikai, illetőleg nyelvtörténeti, általános nyelvészeti, valamint tárgyi-kulturális szempontokkal kiegészülve vezetett a korszerű írói szótár létrejöttéhez (vö. MATTAUSH 1990: 1552–1553). Noha az írói szótárak száma nem csekély, és napról napra nő, a lexikográfiai szakirodalom meglehetősen kevés figyelmet szentel a kérdésnek (általánosabb áttekintés pl. WACHA 1961; BENKŐ 1979 [bő szótárjegyzékkel és szakirodalommal]; KARPOVA 1989; MATTAUSH 1990 [bő szótárjegyzékkel és szakirodalommal]; KARPOVA 2011; valamint az MTA Nyelvtudományi Intézet 1964-es munkaértekezlete az írói szótárakról: GÁLDI 1965a; BENKŐ 1965; WACHA 1965; BACHÁT 1965; KISS J. 1965; MARTINKÓ 1965; SZATHMÁRI 1965; GÁLDI 1965b).

Miként az 1964-es munkaértekezlet ténye is mutatja, az írói szótárak ügye Magyarországon az 1950-es, 1960-as években lendült fel. Bár korábban is készültek íróiszótár-féleségek (a legnagyobb becsű: TOLNAI 1917; részletes bibliográfiát l. WACHA 1961: 190–191, illetve kibővítve: BENKŐ 1979: 23), a fellendülő íróiszótár-készítés mintájául a német és még inkább az orosz hagyomány szolgált, különösen a Puskin-szótár (VINOGRADOV 1956–1961), illetve az akkor már készülő Goethe-szótár (GoetheW., kapcsolódó irodalom: <http://www.uni-tuebingen.de/gwb/biblio.html>; SCHLAPS–SCHARES 2008), de egyebek mellett az Ibsen-szótár – jobbára negatív példaként szemlélt – hatása is kimutatható (l. WACHA 1962). Az így kijelölt út igen tartalmas, gazdag, szemantikailag és stilisztikailag részletező, ugyanakkor lassan, nehézkesen készülő szótárakat eredményezett, a kezdetekkor különösen magas színvonalon. A mai napig számos magyar írói szótár ennek a hagyománynak a vonalát követi kisebb-nagyobb eltérésekkel, egyszerűsítésekkel (JuhászSz. [kapcsolódó irodalom: i. m. 18; BENKŐ 1979: 245–6], PetőfiSz. [kapcsolódó irodalom: GÁLDI 1958, 1965a, 1965b; J. SOLTÉSZ 1965], ToldiSz., BánkBSz., BalassiSz., ZrínyiSz., RadnótiSz.). A most készülő Mikes Kelemen-szótár is minden bizonnyal ezen munkák sorába fog illeszkedni (KISS M. – TÜSKÉS 2011; KISS M. 2012).

* A tanulmány a Bolyai János kutatási ösztöndíj támogatásával készült.

Fontos azonban említést tenni egy másik, ettől jelentősen eltérő hagyományról is. Az íróiszótár-féleségek ezen csoportjának középpontjában nem a szemantikai-stilisztikai kidolgozás áll, hanem a formálisabb, grammatika-központúbb, informatikusabb megközelítés. Ennek a hagyománynak a centruma a debreceni egyetem. PAPP FERENC iskolateremtő személyiségének köszönhetően itt jött létre az 1960-as években a számítógépes nyelvészet, különösen a számítógépes lexikográfia legfontosabb központja.

Az első minta „A magyar nyelv szóvégmutato szótára” volt (VégSz.; kapcsolódó irodalom: PAPP 1964, 1965, 2000 [bő irodalommal: 37–39]), amely „A magyar nyelv értelmező szótára”-nak (ÉrtSz.) „bizonyos grammatikalizált tartalomjegyzéke” (VégSz. 8), számkódolt információkkal. Ezt a modellt követte – nem szótár-, hanem korpuszalapon – számos íróiszótár-féleség (JAKAB–KISS A. 1978, 1980, 1983, 1994, 1997; JAKAB–BÖLCSKEI 1988; JAKAB–KISS A. 2001; JAKAB 2002). Ezen művek többségének korpusza kódex vagy középmagyar nyelvemlék. Ez a debreceni nyelvtörténeti hagyományokra vezethető vissza (vö. pl. a rokon műfajú, de az informatika előtti korból származó WEISZ 1940 e.). Készült ugyanakkor hasonló jellegű szótár – szintén nem informatizált keretben – Debrecentől távolabb is (KÁROLY 1965).

Születtek még íróiszótár-féleségek az elmúlt néhány évtizedben, amelyek nem illeszkednek a fenti iskolák egyikébe sem: a) a JókaiSz. Jókai nehezebben érthető szavainak gyűjteménye, szómagyarázatokkal (tehát valójában sokkötetnyi regény jegyzetanyagának összefésülése), azaz a szerkesztők LEHR (1882, 1905) és TOLNAI (1917) eljárás módjának követői; b) BÜKY sajátos megközelítésű munkáiban (2002, 2004) Füst Milán első három verskötetének szintaktikai-stilisztikai feldolgozását adja betűrendes formában.

2. Nyelvtechnológia a lexikográfiában. PAPP FERENC fentebb idézett cikkeit épp csak egy-két évvel követte KELEMEN JÓZSEF két hosszabb tanulmánya (1966a, 1966b); ezekben a szerző a lyukkártya-technológia lexikográfiai felhasználására tett javaslatokat. Ezeknek – az akkor már elavulóban levő technológiához való erős kötöttségük miatt – értelemszerűen messze nem volt meghatározó iskolateremtő hatásuk. (A technológiai kötöttséget jól mutatják az alábbiak: „Az elektronikus számológépeknek [ti. számítógépeknek – M. A.] szövegből történő szótári célú szógyűjtésére szintén jelentős kísérletek történtek. [...] Úgy látszik, az elektronikus számológépek jelentősége inkább az anyag gyors rendezésében, mint a szótári adatgyűjtésben van. A szöveget ilyenkor is rendszerint lyukszalagra vagy lyukkártyára veszik fel, s arról gépi úton viszik át mágnesszalagra, a gép »memóriájába«, »emlékező mű«-vébe” [KELEMEN J. 1966a: 38–39]. Az igazsághoz hozzátartozik, hogy aztán így folytatja: „Az elektronikus számológépeknek a szövegből történő szógyűjtésben való szerepére később kívánok visszatérni, amikor az eljárás részleteit közelebbről megismerhetem” [i. m. 39].)

PAPP FERENC irányításával a debreceni számítógépes nyelvészeti iskola a fenti modellen kívül is alkotott maradandót. PAPP, miután elkészítette az egyik legkülönösebb és legizgalmasabb, sajnálatos módon azóta is egyedülálló: fonémaalapú

számítógépes feldolgozását Ady költői életművének (JÉKEL–PAPP 1974), egy emlékezetes cikkében a konkordancia mint íróiszótár-nyersanyag fontosságát bizonyította (PAPP 1975). E tanulmány bizonyos mértékig LADÓ (1962)-t követte, de jóval inkább meghaladta, és új, elméletileg is erősen megalapozott kontextusba helyezte a korábbi felvetéseket. Egyebek közt ennek PL/1-ben íródott számítógépes megvalósítása alkotta BÖLCSKEI egyetemi doktori értekezését (1984), majd ennek nyomdokain készült el néhány debreceni „szókinestár”, azaz a tulajdonképpeni írói szótár és a konkordancia közti átmeneti műfajú mű (CsokSzkt. 1–2., ZaySzkt.).

A magyar nyelv nagyszótára 1980-as évekbeli számítógépes újratervelésének (vö. pl. GERSTNER 2006: 13–17 [bő irodalommal]; CSENGERY 2006; PAJZS–MÁRTONFI 2006) hasznos mellékterméke a számítógépes lexikográfia máig egyetlen kismonográfiája (PAJZS 1990), mely elméleti és gyakorlati ismeretekkel egyaránt szolgál. Körülbelül ettől az időtől kezdődően egyre meghatározóbb, mára már lényegében kizárólagos, hogy a magyarországi szótárak informatikai környezetben jönnek létre, rosszabb esetben szövegszerkesztő programokban, optimális esetben TEI-közeli XML-adatbázisként (vö. <http://www.w3.org/XML> és <http://www.tei-c.org>).

A nem írói szótárak köréből mindenképpen ki kell emelni kettőt. Az egyik ELEKFI LÁSZLÓ [1996] „Szókinestárunk nyelvtani alakrendszere” című, eredendően a hagyományos technológiával készült, de számos számítógépes implementációnak, illetve továbbfejlesztésnek alapot adó munkája, amely a magyar inflexió máig legteljesebb leírása, és számos nyelvtechnológiai alkalmazás alapjául is szolgál. A másik a MIgeiSz., amelynek különlegessége, hogy a pusztán konkordancialistánál jóval bonyolultabb szerkezetű, ugyanakkor lényegében tisztán automatikus eljárással készült (csak a válogatásban volt az emberi tudásnak és erőnek szerepe). Ez már mindenképpen a jövő útja: előreláthatólag egyre nagyobb jelentőségük lesz a szótárkészítésben az automatikus és félautomatikus előállítási módú szótáraknak, szójegyzékeknek (néhány ilyen lehetőségéről l. alább).

Mindent összevetve elmondhatjuk, hogy a számítógép megjelenése a lexikográfiában számos jelentős változást eredményezett. Egyfelől – különösképpen az automatikus konkordanciaalkészítés révén – hatékony eszközt kaptak munkájuk megkönnyítésére a lexikográfusok. Másfelől viszont szembe kell nézni azzal, hogy a felhasználói elvárások is megváltoztak, hiszen az írói szótárak konkordanciafunkcióját csaknem kifogástalanul betölti a számítógépes keresés lehetősége. Ezért a lexikográfusok számára az út a nagy hozzáadottérték-tartalmú írói szótárak felé vezet. Ez egyrészt a szemantikai-stilisztikai kidolgozás mélyebb voltát is jelenti, másrészt viszont olyan új funkciók kreálását ösztönzi, amelyek számítógépes támogatással aránylag könnyen megvalósíthatók, és a korábbi szótárakhoz képest újdonságot jelentenek. Ez sokkal több ráfordított erőforrást igényel, és – különösen a gondos szócikkírás és -szerkesztés – meghosszabbítja a szótárak előállításának folyamatát. Mindez persze egyre jobb és jobb szótárakat eredményezhet. A nagy kérdés, hogy ez a fejlődési irány mennyire lehet sikeres a szótárak piacán.

3. Egy konkrét példa: a készülő József Attila-szótár. A Petőfi-szótárral csaknem egy időben felmerült egy József Attila-szótár elkészítésének gondolata (WACHA 1959, 1965). A cél József Attila szépírói (tehát költői, versfordítói és szépprózai) nyelvének szótározása volt (WACHA 1965: 387–388). A munkálatok az anyaggyűjtés és -rendezés végeztével azonban sajnos anyagi okokból megrekedtek. A József Attila nyelvhasználatával, nyelvművészetével (l. különösen TÖRÖK 1976), illetve nyelvszemléletével, nyelvesztétikájával, névvarázselméletével (l. TVERDOTA 1986 [a szerző korábbi kutatásainak irodalomjegyzékével]) kapcsolatos kutatások természetesen folytak azóta is, igaz, nem lexémaközpontú, szótári keretben.

E sorok írója 2011-ben elindított egy újabb projektumot, amelynek célja a József Attila-szótár elkészítése. Az új tervek a teljes József Attila-i életmű szótározására irányulnak, tehát a verseken kívül a szépprózai művek, drámarészletek, értekező próza, levelek, pszichoanalitikus írások, feljegyzések és egyéb dokumentumok is a korpusz részét képezik. A szótár feldolgoz minden egyes fennmaradt kontextusos szót, amelyet József Attila leírt saját szövegeként, legyen akár eredeti alkotás, akár fordítás. A kétes hitelű szövegek, illetve az ezekhez hasonló státusú interjúválaszok (ezek végleges formája a korabeli újságírói gyakorlat ismeretében egyértelműen az interjú készítőjéhez köthető ugyanis, ugyanakkor gyakorta átsüt ezeken a válaszokon József Attila egyéni stílusa, másutt is fellelhető szófordulatai) szintén bekerülnek az anyagba, de a korpusz törzsétől elkülönítve.

Többek között a művelt műfajok nagy számából adódóan József Attila szókészlete igen változatos. Azáltal hogy a szótár gyűjtőköre a teljes életműre kiterjed, lehetővé válik a különféle regiszterek, lexikai rétegek, illetve az egyes műfajok szókészletének összehasonlítása. A szótári kidolgozás módja és formája részletező (lexikai-stilisztikai és nyelvtani minősítések, jelentésstruktúra, értelmezések, példamondatok, a költői eszközök és alakzatok bemutatása stb.). Mindebből következőleg a József Attila-szótár alkalmas lesz egyben a 20. század első fele magyar szókészletének áttekintésére is. Ez különösen hasznos, mivel BALASSA kisebb terjedelmű és színvonalában sem igazán erős szótárán (1940) kívül ebben az időszakban nem készült magyar értelmező szótár. Tehát a József Attila-szótár egyaránt hasznos lesz az irodalom, a nyelvészet, a stilisztika és a retorika kutatóinak is. József Attila szókészletének nagysága, tehát a szótár címszóállománya kb. 25 ezerre becsülhető, amelyet mintegy 300 ezer szóelőfordulásból lehet elvonni.

A címszók közé – a köznyelvi szótáraktól eltérően – bekerülnek a tulajdonnevek, így a szótár bizonyos enciklopédikus funkciók betöltésére is alkalmas lesz. Egyelőre kérdésesnek tekinthető még az igenevek önálló címszói vagy az igei paradigma részeként való kezelése, illetve a nem mai helyesírású alakok besorolása. Az eredeti helyesírás megtartása a tájékozódást nehezíti meg, a mai helyesírás favorizálása a szerzői nyelvhasználattal áll szemben. Mindazonáltal ez a kérdés számítógépes szótár esetén elenyésző problémát okozhat csupán, hiszen a keresés nem szorítkozik a szoros betűrendre.

3.1. József Attila alkotói módszerének jellemző vonása a szövegváltozatok nagy száma, tehát a feldolgozás során ezek kezelésében különös gonddal kell eljárni. Erre igen alkalmas, ha a számítógépes konkordancia alapjául ún. genetikus

szöveg szolgál. Ez összesítve tartalmazza az összes változatot, így lépésről lépésre követi a szöveg létrejöttét, geneziséét. (A tanulmányok, cikkek esetében egyik kéziratból másikba „vándorló” mondatoknak egyetlen szöveggkonglomerátumként való feldolgozása alighanem lehetetlen. A számítógépes szövegelemzés azonban ezen szövegek ismétlődő részeinek feltárására is lehetőséget nyújt.) A genetikus szöveg használata lehetővé teszi a főszöveg és a variánsok egységes kezelését, így a konkordanciában minden egyes szó a releváns környezetében állhat, és a gyakorisági statisztikák is minden egyes szóelőfordulást csupán egyszer vesznek figyelembe. (A munkálatok során mindvégig eltekintünk az írásjelhasználati variánsoktól, a pusztán hosszúság-rövidségbeli változatoktól, valamint a szórendi cseréktől. Eltekintünk továbbá a – nem véglegesnek szánt szövegváltozatokban előforduló – be nem fejezett szótöredékektől, valamint az olyan korábbi változatoktól, amelyek csak ragozásban térnek el valamely későbbi változattól, és a korábbi változathoz szintaktikailag illeszkedő mondat vagy mondatdarab le sem íródott, így tehát jól formált mondat egység mint minimális kontextus nem járul az adott szóhoz. Az efféle esetek vizsgálata olyan mikrofilológiai feladat, amelynek a kiszolgálása túlmutat egy írói szótár keretein és lehetőségein.)

Tekintsük példaként az alábbi töredéket, amely négy változatban maradt fenn:

- (1) a) *Az árnyékból elindul lassan / hús szárnyakkal a szürkület.*
- b) *Az árnyékból elindul lassan / csapatosan a szürkület.*
- c) *Az árnyakból – csipogva – fiókáit / összegyűjti a szürkület.*
- d) *A csipogó árnyakból fiókáit / összegyűjti a szürkület.*

Ezek genetikus reprezentációja a következő:

- (2) $A[z \text{ árnyékból elindul lassan} / [h\acute{u}s \text{ szárnyakkal} | csapatosan]] \{ \{^1 z\} \{^2 \text{csipogó}\} \text{ árnyakból} \{^1 - \text{csipogva} -\} \} \text{ fiókáit} / \text{összegyűjti} \} a \text{ szürkület.}$
(Ahol: [...] – a variánsblokk határolója; | – az egyes variánsok határolója; $\{^{\#} \dots\}$ – összeindexelt variánsok; / – sortörés.)

Az algoritmus működésének illusztrálására nézzük a **csipogva** konkordanciatételének elkészítését.

Első lépésben vesszük az aktuális szó határolókkal jelölt legkisebb kontextusát, jelesül az $\{^1 \dots\}$ csoportot, majd töröljük az összes konkurens variánst:

- (3) a) $A[z \text{ árnyékból elindul lassan} / [h\acute{u}s \text{ szárnyakkal} | csapatosan]] \{ \{^1 z\} \{^2 \text{csipogó}\} \text{ árnyakból} \{^1 - \text{csipogva} -\} \} \text{ fiókáit} / \text{összegyűjti} \} a \text{ szürkület.}$

Ezután a kiválasztott rész határolókkal jelölt legkisebb kontextusát vesszük, jelesül a második belső [...] csoportot, majd – ha van – töröljük ennek a konkurens variánsát:

- (3) b) $A[z \text{ árnyékból elindul lassan} / [h\acute{u}s \text{ szárnyakkal} | csapatosan]] \{ \{^1 z\} \{^2 \text{csipogó}\} \text{ árnyakból} \{^1 - \text{csipogva} -\} \} \text{ fiókáit} / \text{összegyűjti} \} a \text{ szürkület.}$

A harmadik lépésben kiválasztjuk az álló virgula releváns oldalát, majd töröljük a konkurens variánst:

- (3) c) *A[z árnyékból elindul lassan / [hűs szárnyakkal|csapatosan] \{ \{ 'z \} \} \{ 'csi-
pogó \} árnyakból \{ ' - csipogva - \} \} fiókáit / összegyűjti] a szürkület.*

Végezetül hozzákapcsoljuk a kiválasztott részhez a zárójeleken kívüli szövegrészeket:

- (3) d) *A[z árnyékból elindul lassan / [hűs szárnyakkal|csapatosan] \{ \{ 'z \} \} \{ 'csi-
pogó \} árnyakból \{ ' - csipogva - \} \} fiókáit / összegyűjti] a szürkület.*

Nézzünk egy másik példát. Ebben olyan szóhoz (pl. a **szürkület**-hez) gyárt az algoritmus konkordanciatételt, amely kívül van az összes zárójelen. Ilyenkor az ultima manus/editio elv alapján a [...] és {#...} csoportokból az utolsó változatot kell kiválasztani (ha a kritikai kiadás alapszövege meggyőző érvek alapján eltér ettől az elvtől, azt fordított indexeléssel lehet kezelni), pontosabban először a nem utolsókat kell törölni, majd a konkurens nélkül maradókat kiválasztani. Lépésenként tehát:

- (4) a) *A[z árnyékból elindul lassan / [hűs szárnyakkal|csapatosan] \{ \{ 'z \} \} \{ 'csi-
pogó \} árnyakból \{ ' - csipogva - \} \} fiókáit / összegyűjti] a szürkület.*
 b) *A[z árnyékból elindul lassan / [hűs szárnyakkal|csapatosan] \{ \{ 'z \} \} \{ 'csi-
pogó \} árnyakból \{ ' - csipogva - \} \} fiókáit / összegyűjti] a szürkület.*
 c) *A[z árnyékból elindul lassan / [hűs szárnyakkal|csapatosan] \{ \{ 'z \} \} \{ 'csi-
pogó \} árnyakból \{ ' - csipogva - \} \} fiókáit / összegyűjti] a szürkület.*
 d) *A[z árnyékból elindul lassan / [hűs szárnyakkal|csapatosan] \{ \{ 'z \} \} \{ 'csi-
pogó \} árnyakból \{ ' - csipogva - \} \} fiókáit / összegyűjti] a szürkület.*
 e) *A[z árnyékból elindul lassan / [hűs szárnyakkal|csapatosan] \{ \{ 'z \} \} \{ 'csi-
pogó \} árnyakból \{ ' - csipogva - \} \} fiókáit / összegyűjti] a szürkület.*
 f) *A[z árnyékból elindul lassan / [hűs szárnyakkal|csapatosan] \{ \{ 'z \} \} \{ 'csi-
pogó \} árnyakból \{ ' - csipogva - \} \} fiókáit / összegyűjti] a szürkület.*

3.2. A s z ó c i k k e k – mint általában – három részből állnak: szócikkfejből, szócikktörzsből és szócikkvégből.

3.2.1. A s z ó c i k k f e j b e n találhatóak – egyebek mellett – a szócikkre (valamint alakváltozataira, illetve paradigmikus alakjaira) vonatkozó statisztikai adatok. Minthogy a feldolgozás egésze elektronikus korpuszon alapul, lehetőség nyílik több szempontú statisztikákra: az egyes gyakorisági értékek lebonthatóak, illetve összesíthetők a következők alapján: alakváltozat, paradigmabeli hely (a konkordanciához szükséges lemmatizálásnak, valamint a paradigmikus alakok meghatározásának alapjául számítógépes morfológiai elemzés szolgál), keletkezési idő, műfaj (hasonló, de egyszerűbb felbontáshoz l. ČERMÁK 2008; ČERMÁK–CVRČEK 2010). (A gyakorisági értékek meghatározásakor megfontolandó

az SzGySz.-ban használt ún. diszperzió és módosított gyakoriság, más néven úzus esetleges használata is; vö. FÜREDI 1983: 342–343¹; SzGySz. 2.)

A négy független és egy (vagy több) függő változó megjelenítésére a nyomtatott szótár természetesen nem alkalmas (az egyszerűsített változat pontos megtervezése a sajtó alá rendezés során lesz esedékes, ez a korábbi magyar írói szótárak, különösképpen a Petőfi-szótár eljárásához fog hasonlítani; l. 1. ábra), a leendő elektronikus szótárban viszont mód nyílik a sokdimenziós grafika megvalósítására, illetve különféle szűrések, nézetek vizualizálására.

1. ábra

A *lány* ~ *leány* gyakorisági statisztikája néhány írói szótárban:

a) Juhász Gyula-szótár (JuhászSz.); b) Petőfi-szótár (PetőfiSz.);

c) Balassi-szótár (BalassiSz. – teljes szócikk); d) Csokonai-szókincstár I.

(CsokSzkt. 1. – teljes szócikk)

- a) **lány** ~ **leány** fn [V: 73 ~ 29]
-t -ban -ról -ra -nak -nyal, -ok, -a
-át -ával, -ai; -nyal, -ok, -a
- b) **leány** fn 554 | **leány** 299, **lány** 22, **lyány** 233 :
-ő 258 (b 18), -ők 22 (b 4); -om 22 (t 1, b 6), -od 13 (b 3), -a 57 (b 7), -nek 1, -tok 1; -aid 1, -ai 6 ~ -i 2, -aitok b | -s 51, -okat 1, -omat 5, -odat 5, -át 12, -ait 2, -aitokat 1; -ban 1; -ból 1; -ért 4, -ért 1; -hoz 7, -okhoz 1, -ához 4; -omként 1; -nak 17 (b 5); -odnak 2, -ának 5 (b 3), -ainak b 1; -nál 1, -oknál 2, -ainál 1; -ra 5, -okra 1, -odra 1, -ára 1; -tól 5 ~ -tul 1, -októl 1, -ától 1, -ától 1; -nyal 15, -okkal 2, -ával 4, -ával 1 | -s 4 | -élől 1
- c) **leány** ~ **lány** fn (12+2)
1. Fejnőtt, nem házas nő; hajadon. Az *Angelica* egy igen szép *leány* vala (K52/7). 2. Vki-nek nőnemű gyermeke. Hol *Diomedesszel* az *ferdős leánya*? (30/69) 3. Fiatal nőnemű (állat). *Fulgosius* egy *sasról* ír ilyen *csudát*, *Ki úgy szerette, együtt nevelte lányát*, *Hogy csak annak vitt minden fogott madarát*, *Sőt leány haláláért meg is ölte magát* (30/18–20).
leány (13/cím, 35/cím), *leán* (60/1), *lányok* (3/22), *leányok* (62/5), *leányom* (K27/7), *leánt* (30/59, K8/3), *leányoknak* (K47/28), *leányról* (85/cím). Ő: *szolgáló*.
- d) **lány** fn (3)
Tsak Loderdum édes *Lányom* hadd menjen (15/1027).
Lyány (4/13), *Lyányom* (4/161).
L. még *leány!*
- leány** fn (59)
De *mellyk* az, az annyira egygyűjű *Leány*, a' ki kijövéen a' bőlisből azonnal ne tudja azt a' mester-séget, hogy széfnek lásson, hogy teitzetes légyen, hogy kedveltetésével gyilkoljon, és hogy tudja mi-tosoda fegyver sebesít, *mellyk* ő' s' *mellyk* gyógyít meg, és hoz vissza az életre? (7/662) — Tsak credj édes *Leányom*, tudod a' műket mondottam, vilesd véghez majd mi addig egyet sétálunk az urral a' kertbe (15/601). — ...észe vettem, hogy a *Leány* kész akár *melly* *Legényért* az Asszonyt elárulni (16/558). Ezt szálla a' *Hizelkedés*, a' *melly* a' *Gazdagság* és a' *Szegénység* *Leánya* (1/3284). Ezek a' *lármas* fergetegek, a' *mellyek* körülöttem keringenek, nem az *Otzeán*' halálós *leányi* é? (13/722)
leány (1/772, 773, 7/149, 13/587, 735), *Leány* (7/294, 325, 638, 985, 1037, 1086, 1139, 15/930, 16/569), *Leányok* (7/121, 1346, 12/205), *Leányom* (4/35, 6/131, 15/526, 561, 582, 584, 891, 896, 902, 904, 993, 16/978), *leány* (1/3, 16/741), *Leánya* (7/152, 12/6, 352, 588, 13/5, 15/5), *Leány* (7/298, 676, 15/944), *Leányát* (1/2476, 7/341), *Leányikat* (5/172), *Leányak* (7/657, 658, 852), *Leányba* (6/176), *Leányak* (15/122), *Leány[na]k* (16/818), *Leányok[na]k* (16/821), *Leányomnak* (15/553), *Leányinak* (5/395), *Leányommal* (1/522, 15/153).
Ő: *követ*-, *paraszt*-, *szoba*- . L. még *lány!*

¹ Mivel a hivatkozott forrásban az egyik képlet sajnálatos nyomdatechnikai hibával jelent meg, helyénvaló itt megadni számítási módjukat:

$$F_m = F \cdot D; D = 1 - \frac{V}{\sqrt{n-1}}; V = \frac{S}{f}; S = \sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{f} - f_i)^2}; \bar{f} = \frac{1}{n} \sum_{i=1}^n f_i -$$

ahol F_m : módosított gyakoriság, F : abszolút gyakoriság, D : diszperzió, V : variációs együttható, n : az osztályok száma, S : négyzetes eltérés, \bar{f} : osztálygyakoriságok átlaga, f_i : osztálygyakoriság (FÜREDI 1983 terminológiájával).

A szócikkfej további adatokat tartalmaz: a címszó első előfordulásának évét József Attila életművében, első ismert előfordulását a magyar nyelvben, illetve rövid etimológiai összefoglalót (ez további lexikológiai statisztikai vizsgálatoknak szolgálhat alapjául). Arra is itt utalunk, hogy az adott lexéma mennyire „József Attila-i”: 1. József Attila saját alkotása (tehát korábról ismeretlen); 2. egyáltalán nincs nyoma, hogy más (akár későbbi) szerző használta volna (eltekintve az idézetszerű használatától); 3. az adott szó hapax legomenon, tehát egyetlen, József Attila-i előforduláson kívül tetszőleges magyar szövegben sem lelhető fel (szintén nem idézetszerű) előfordulása a címszónak. Még egy érdekes mutató található a szócikkfejben: azt jelzi, hogy a lexéma gyakorisága szignifikánsan magasabb József Attila életművében, mint a kontrollkorpuszban. Ez utóbbi a kor legrangosabb irodalmi folyóiratának, a Nyugatnak a teljes szövege. (A módszerről bővebben l. ČERMÁK 2008; ČERMÁK–CVRČEK 2010.)

3.2.2. A szócikktörzsnek, a szócikk lényegi részének elkészítéséhez alapul a számítógéppel előállított, a keletkezési évet, azon belül pedig a műfajokat is figyelembe vevő konkordancia szolgál (ahogy a Mikes-szótár esetében is, vö. Kiss M. 2012), az adatok csoportosítása, jelentésekbe sorolása azonban filológusi-lexikográfusi szaktudást igényel.

A szócikktörzsek tartalma a történeti-értelmező szótárak megszokott rendjét követi (részletes jelentésstruktúra lexikai-stilisztikai minősítésekkel, szövegelemzésen alapuló értelmezésekkel, idézetekkel), beépítve a magyar szótári hagyománynak, illetve legkiterjedtebb megvalósítójának, „A magyar nyelv nagyszótárá”-nak (Nsz.) lehető legtöbb vívmányát. Ebben a munkafázisban hasznosíthatók a József Attila szókészletére és szóhasználatára irányuló korábbi nyelvészeti kutatások (különösen TÖRÖK GÁBOR munkái), valamint – a kiterjedt és erősen összefüggő motívum- és szimbólumrendszer okán – a József Attila költői életművének motívumrendszerére vonatkozó kutatások (pl. RIGÓ BÉLA kéziratosa közlése).

Dilemmát jelent a jelentésszerkezetek kidolgozása során, hogy ezek milyen mértékben tagolódnak, illetve vonódnak össze. Tudni való, hogy minden egyes szóelőfordulásnak eltérő mértékben a jelentése az összes többitől, különösen, ha számot adunk a stilisztikai jelentésről is. Így tehát egy írói szótár esetében elviekben minden szóelőfordulást önálló jelentésárnyalatként kellene kezelni. Ez azonban lehetetlenné tenné, hogy meglássa a szótár használója a fától az erdőt. Ha ezt szeretnénk elősegíteni, illetve a szókészlet elmebeli megjelenését próbáljuk modellálni, akkor sokkal inkább az elnagyoltabb, alulspecifikált értelmezések, illetve ezzel együtt a kisebb jelentésszám vezet eredményre. A két igény egyaránt jogos, így a középút látszik leginkább célravezetőnek.

A jelentések elhatárolásában előnyben kell részesíteni a különféle formális tesztek tanúságát a nyelvtől távolabb álló, logikai alapú, fogalmi szempontú eljárásmodoktól. Ha tehát például egy lexéma két előfordulásában különféle szinonimák jelennek meg mint lehetséges helyettesítők, bizonyosan különböző jelenté-

sekről van szó. Vagy eltérő vonzatszerkezet szintén eltérő jelentésre utal – hiszen a jelentésnek nem pusztán fogalmi-szemantikai, hanem grammatikai vetülete is van. Az ilyen tesztek szinkron értelmező szótár készítésekor a szótárkészítő anyanyelvi kompetenciájára alapozva biztos alkalmazható, írói szótár (mint minden zárt korpuszú szótár) esetében azonban szükség van pótkompetencia kialakítására (vö. FORGÁCS 1993–1994). Ez nem egyszerű feladat, a stilisztikai pótkompetenciáé különösen korlátozott sikert ígérhet csupán. Mindazonáltal a szótár szemantikai és stilisztikai keretének nemcsak a formális módszereken kell alapulnia (ez a lexikai-stilisztikai minősítések esetében is lehetséges bizonyos határok közt), de szükség van a korszerű kutatások egyéb jellegű felhasználására is (pl. a kognitív metaforaelmélet bizonyos eredményei is beépíthetők lennének a stilisztikai leírásba). A József Attilára oly jellemző komplex képeket szintén dokumentálni kell, gazdag leírásuk azonban minden bizonnyal szétfejténe a szótári műfaj kereteit.

Az értelmezésekhez – a tervek szerint – kialakítandó egy zárt nyelvtanú és szókészletű értelmezőnyelv, amely módot nyújthat későbbi számítógépes lexikológiai kutatásokra, így például József Attila szókészlete formális belső hálózatának, összefüggéseinek feltárására. Ez a technológia ugyanakkor alkalmazható teszi a szótári értelmezéseket gépi fordításra: nem csupán ezen írói szótár, de általában a jövőben ilyen értelmezőnyelvet használó bármely szótár esetében. A zárt értelmezőnyelv alkalmazása továbbá lehetséges interfészt jelenthet az emberek, illetve a számítógép számára készített (pl. Wordnet) szótárak között. E két alapvető szótártípus közti átjárhatóság hiánya vagy nehézsége egyre égetőbb probléma a lexikográfiai, illetőleg nyelvtechnológiai fejlesztésekben.

3.2.3. A szócikkek végén belső és külső utalók állnak, ezek összeállításában is lehet bizonyos mértékben informatikai segítségre támaszkodni. A belső utalók az összetételek tagjaira és a különféle frazeológiai alakulatok vezérszavára (tehát a kidolgozás helyére) utalnak. A külső utalók megmutatják a címszó helyét a magyar szókészlet egészében, mind történeti, mind leíró szempontból. Arról tájékoztatnak tehát, hogy az adott szó előfordul-e más szótárakban: értelmező szótárakban (CzF., Ball., Balassa, ÉrtSz.), a szerzőhöz köthető kétnyelvű szótárakban (itt különösen KELEMEN B. [1912] magyar–német és THIBAUT–KOVÁTS S. [1901] magyar–francia szótárára érdemes figyelmeztetni, ezek voltak meghatározók József Attila számára – gondoljunk csak arra, hogy franciául a szótár sorról sorra való bebeflázásával próbált megtanulni), általános (MTsz., ÚMTsz.), illetve a József Attila nyelvi életrajzához kapcsolható (ÖcsödiSzólKm., MakóiTsz., SzegSz.) tájszótárakban, egyéb – 19. és 20. századi életműveket feldolgozó – írói szótárakban (JuhászSz., PetőfiSz., JókaiSz., RadnótiSz.), valamint egyéb olyan szótárakban, amelyek megtalálhatók voltak József Attila könyvtárában, és így hatással lehettek a szókészletére és szóhasználatára (a korábban említetteken kívül TolnaiMagySz.³).

3.3. A szótárt függelékek fogják kiegészíteni. Ezek megvalósításában meghatározó szerepet játszanak különféle megszokottabb, illetve újszerűbb nyelvtechnológiai eljárások. A függelékek két típusba sorolhatók aszerint, hogy a teljes vagy csak a költői életmű feldolgozásával állnak elő.

1. A teljes korpusz alapján készülő első függelék az esetek és inflexió-s toldalékok statisztikai feldolgozását adja. Ez a paradigmatiszta alakok szócikkfejbeli bemutatásának melléktermékeként állítható legegyszerűbben elő. Ez a statisztika is számon tartja a keletkezési időt és a szöveg műfaját.

2. Szintén a teljes korpusz alapján (végső soron a szótár törzsanyagának transzformálásával) jön létre a József Attila-életmű csökkenő gyakoriság szerinti szójegyzeke. Az ilyen listák az írói szótárakban általánosnak mondhatók. A rendezés során itt is figyelembe vehető a módosított gyakoriság, illetve a kontrollkorpuszhoz viszonyított gyakoriság.

3. Noha természetesen a szótári szócikkekben értelmezett egységként gondosan ki lesznek dolgozva a szerzőre jellemző frazeologizmusok, hasonló jelenséget más fénytörésben mutat be a kollokációjegyzék, azaz József Attila legkedveltebb szó-*n*-eseinek gyakorisági listája (vö. ČERMÁK 2008; ČERMÁK–CVRČEK 2010). Itt lehetőség nyílik a törzsanyagban kidolgozott kollokációk megjelölésére is. Ez a függelék is a teljes korpusz alapján készítendő el.

4. Minthogy a magyar rím rendszerint nem tiszta rím, hanem asszonánc, tehát a rímelő szótag(ok)nak nem a teljes azonosságán, hanem – bizonyos szabályok szerinti – hasonlóságán alapul, valamint annak következtében, hogy ragozó nyelv a magyar, a rímek változatossága a magyarban szinte korlátlan (ugyanakkor az aránylag kevés tiszta rímnek a nyugat-európai költészetekben szokásos ismételtetése a magyarban csikorog). Ez okból mindmáig nem készült, mert elvileg sem készülhet komolyabb terjedelmű, a szó klasszikus értelmében vett rímiszótár. Lehetséges ugyanakkor elkészíteni egy (vagy több) szerző megvalósult rímeinek szótárát. Ez hatékonyan előállítható nyelvtechnológiai eszközökkel, emberi kéz érintése nélkül. Ez a függelék a költői életművön alapul, s az algoritmus próbája lehet egy szélesebb körű magyar rímiszótárnak is (vö. MÁRTONFI 2008). (2. ábra.)

2. ábra

Mutatvány a megvalósult magyar rímek szótárából (MÁRTONFI 2008: 201–202)

szikla (14+31)

~ (2): *virága* (?) (Kisfaludy K.: A két hajós, 1822–1826); *szikra* (Arany J.: Toldi szerelme XI/133, 1863–1879); *halikra* (Arany J.: A nagyidai cigányok III, 1851); *alakja* (Vajda: Béla királyfi VI, 1854); *volna* (Vajda: Visegrádon, 1855); *boritva* (Vajda: Az apagyilkos, 1860); *lábaidra* (Juhász Gy.: Kérdés, 1906); *szikra* (Kosztolányi ford.–Rilke: A fogoly, 1921); *síkra* (Kosztolányi ford.–Leconte de Lisle: A jaguár, 1921); *titka* (Dsidá: Új Loreley-vers, 1926); **a szikla**: *csontjaikra* (Arany J.: A rodostói temető, 1848); *szakítja* (Kosztolányi ford.: Válás, 1931); **izzó szikla**: *szállongó szikra* (Vajda: Vasúton, 1885)

sziklába (2): *lába* (Arany J.: Toldi szerelme X/2, 1863–1879)

sziklából (2): *csapjából* (Kisfaludy S.: Somló 55, 1805); *gyomrából* (Kisfaludy S.: Micz bán 4, 1836); *odvából* (Csokonai ford.–Kleist: A tavasz, 1801–1802)

sziklaiddal (1): *diadal* (Arany J.: Murány ostroma I, 1847)

sziklánál (2): *csinál* (Tompa: Tanács, 1860)

szikláira (2,25): *partyaira, vaskapuira, fiaira* (Ányos: Barcsay kapitánynak, 1778); *szállásira* (Uő: Gróf Eszterházi Pálnak pécsi püspökségre lett felemeltetését inneplő versek, 1781)

sziklaival (2): *habjaival* (Kisfaludy S.: A kesergő szerelem IV/38, 1798); *vérítal* (Kosztolányi ford.–Leconte de Lisle: A jaguár, 1921)

sziklája (1): *lakója* (Vajda: Az én szülőföldem, 1847)

sziklájához (2): *apánkhoz* (Petőfi: A táblabíró, 1847)

sziklájáig (2): *Sináig* (Csokonai ford.–Béze: A szarándok, 1796)

sziklájára (3,5): *kalibájára* (Uő: Egy fohászkodás, 1782); *várára* (Uő: Barcsaynak, 1782)

szikláját (2): *pusztáját* (Csokonai ford.–Bürger: Az utolsó szerencsétlenség, 1802–1803)

sziklák (2): *sírták* (Arany J.: A lantos, 1849)

sziklákat (2,17): *[sziklákat], tölgyfákat, kányákat, rókákat* (Verseghy: Rikóti Mátyás II, 1804); *láthat* (Arany J.: Murány ostroma I, 1847); *látok* (Babits: Az Isten és az ördög, 1933); *világnak* (Reményik: Boldog testvérünk, 1935)

sziklákon (2): *határon* (Gyulai: Erdély, 1847)

sziklákra (2): *fúvására* (Petőfi: Lehel vezér, 1848)

sziklán (1,6): *karján* (Kisfaludy K.: Zuárd, 1828–1829); *alján* (Gyulai: Erdély határán, 1851); *villám* (Vajda: Béla királyfi VI, 1854); *villám* (uő: Visegrádon, 1855); *ritkán* (Tompa: Emléklapok, 1856)

sziklát (2,5): *itt lát* (Kosztolányi ford.–Hugo: A költő szomorúsága, 1921); **büszke sziklát**: *dühébe szirkát* (Tompa: Gránit-szobor, 1847)

sziklával (2): *nyavalyával* (Verseghy: Tisztelete Marczibány Istvány Úrnak és Hitvesének, 1806)

Vő: *kőszikla*; ad₁, át, dob₁, felett, fok, ha₁, hit, is, nyom₁, reng, váll, van

sziklabástya (0+1)

sziklabástyán (2): *borostyán* (Babits ford.–Wilde: Charmides 2, 1911)

sziklabérc (1+1)

~ (2): *oltalma kész* (Tompa: Szent László királyról, 1846)

sziklabérc (2): *küszködének* (Tompa: Egy könyv olvasása közben, 1861)

Vő: felett

sziklabolt (2+1)

~ (1,67): *[szikla-bolt], lakás volt* (Kisfaludy S.: Tátika 18, 1805); *anyja volt, drága holt, [szikla-bolt]* (Kosztolányi ford.–Wilde: Shelley sírja, 1921)

sziklaboltba (3): *kikapartam volna* (Arany J.: Toldi szerelme IX/59, 1863–1879)

sziklacsont (0+1)

sziklacsontok (3): *omlok, sziklahomlok* (Komjáthy: A getszermáni kertben, 1885–1888)

sziklacsúcs (0+1)

sziklacsúcsokon (2): *távozom* (Kosztolányi ford.–Leconte de Lisle: A jaguár, 1921)

szikladörgető (2+0)

~ (2): *elő, ébredő* (Kosztolányi: Húsvéti himnusz, 1908 – elő 2×)

sziklaél (1+1)

~ (1): *szél, [szikla-él]* (Tóth Á. ford.–Byron: Childe Harold indulása 12, 1917)

sziklaélen (2): *sötéten* (Reményik: Két fény között, 1927)

Megjegyzendő, hogy lehetséges egy másik fajta magyarrímszótár-típust is létrehozni, ám ez nem hagyományos értelemben vett, hanem virtuális (azaz dinamikus) szótár. Ennek célja, hogy a felhasználó által megadott rímhívó szóra vagy kifejezésre generálja az összes magyar rímet. A működtető algoritmus egyszerű: miután a felhasználó megadta a megrímelni kívánt szót vagy kifejezést, a rímgenerátor előállítja az összes lehetséges rímelő karakterszekvenciát, majd az utána kapcsolt morfológiai elemző csak azokat a lehetőségeket engedi a kimenetre, amelyek megfelelnek a magyar szóalaktan szabályainak. (A <http://rimszotar.poet.hu/> címen feltehetőleg 2010 nyarán megjelent rímgenerátor a minimális dokumentáció alapján lényegesen egyszerűbb, elméletileg nem koherens alapokon áll: „A szótár jelenleg több mint 165 000 szót tartalmaz, melyek részben a Poeten található versekből, részben pedig egyéb szabadon felhasználható forrásokból származnak. A rímpárokat automatikus úton, valamint önkéntesek segítségével állítottuk össze.”)

5. A következő, a költői életművet feldolgozó függelék az előreláthatóan félautomatikus előállítású metrikai adatbázis. Ez az alapvető filológiai adatokon felül tartalmazza az egyes versek versrendszerét (időmértékes, ütemhangsúlyos stb.), strófaszerkezetét, metrikai leírását, rímképletét, az alliteráció mértékét stb. A metrikai adatbázis létrehozásához ritmikai elemző, illetve ebből bizonyos mértékű absztrakció alapján metrikai következtetéseket levonó nyelvtechnológiai eszköz használandó. Az algoritmus kiterjeszhető lesz a magyar vers sokkal szélesebb körének számítógépes feldolgozására. (1. táblázat.) (A <http://www.poet.hu/verselemzo.php> alatti ritmikai elemző lényegesen kisebb tudású annál, mint amivel a kijelölt feladat magas színvonalon megoldható lenne.)

1. táblázat

Mutatvány József Attila metrikai adatbázisából

	Cím	Kötet	Keltezés	Versrsz.	Metrikai képlet	Stró-fák	Rímképlet	Str.	Megjegyzés
1	A jámbor tehén	Szépség koldusa	22. 08–09.	im.	5j 5.5j	4 4 3 3	abba baab cdc dcd	4	Petrarca-szonett
2	Éhség	Szépség koldusa	22. 07–08.	im. (szim.)	5j(=5/5) 5.5j(=5/6)	4 4 3 3	abba abba cde ced	4	Petrarca-szonett
3	Megfáradt ember	Nem én kiáltok	23. 08.	im.	4.5t 5j 5.5j 5.5t 6j 6t	4	xaxa	3	
4	Tanulmányfej	Nem én kiáltok	22. 11 vége	im. (szim.)	5j(=5/5) 5.5j(=5/6)	4 4 3 3	abba abba ccd cdc	4	Petrarca-szonett
5	Rög a röghöz	Nem én kiáltok	23. 08.	im.	5.5j+5j	2	ab ab	8	
6	Névnapi dícséret	Nem én kiáltok	22. 07. 29.	im. (szim.)	asc min 3+gly	4	xxxx	4	első aszkl. strófa
7	Útrahívás	Nem én kiáltok	23. 03.13.	im. (szim.)	sap(=4/7) 3+ado	4	xxxx	5	szaphói strófa

	Cím	Kötet	Keltezés	Versrsz.	Metrikai képlet	Stró- fák	Rímképlet	Str.	Meg- jegyzés
8	Sacrilégium	Nem én kiáltok	23. 02. 23.	im. (szim.)	$5j(=5/5) 5.5j(=5/6)$	4 4 3 3	abab abab ccd dcd	4	Petrar- ca-szo- nett
9	Bevezető	Nincsen apám, se anyám	27.05.	im. (szim.)	$(4j=5/3)=amb [8.:+1]$	8	ababccdd	1	ambro- ziánus
10	Szeretnének	Nincsen apám, se anyám	28	im. (szim.)	$5j=5/5$	2	aa	9	felező tízes
11	Tiszta szívvel	Nincsen apám, se anyám	25. 03.	üh. (szim.)	$4/3=3.5t$	4	aabb	4	
12	Két dal	Nincsen apám, se anyám							ciklus
12a	Éji dal	Nincsen apám, se anyám	28. 03.	im. (szim.)	$4.5j+4j=5/4+5/3$	4	abab abab	2	~huitain
12b	Dal	Nincsen apám, se anyám	28. 02. 10.	im. (szim.)	$4j=5/3 4.5j=5/4$	3 5	AbA abbaA	2	triolett
13	Nemzett József Áron	Nincsen apám, se anyám	28 eleje	im.	$3t+3.5t+3t+3.5t$	4	abab	4	
14	Engem temetnek	Medvetánc – Nincsen...	26. 08.	üh. (szim.)	$5/4=4.5j$	4	xaxa	4	Ady- kilences
15	Magyar Alföld	Nincsen apám, se anyám	28	üh. (szim.)	$4/4+3/2+4/4/2+4/3/2$ $=4t+2.5t+5t+4.5t$	4	xaxa	4	
16	Biztató (Kínában lóg a mandarin...)	Nincsen apám, se anyám	27.01.	im. (szim.)	$(4j=5/3)=amb$	4	aaBa	4	ambro- ziánus
17	József Attila (Vi- dám és jó volt...)	Nincsen apám, se anyám	28 eleje	im.	$4j 4.5j$	10 1	abbaaabba a	2	
18	Hangya	Nincsen apám, se anyám	26 nyara	im. (szim.)	$5.5j+4.5t+3j 5.5j+5j$	3 2 3 2	abc ac xxb ab	4	
19	Isten (Hogyha golyóznak...)	Nincsen apám, se anyám	28	üh. (szim.)	$5/4(=4.5j)$	4	aabb	5	Ady-ki- lences
20	Istenem	Nincsen apám, se anyám	28 eleje	üh. (szim.)	$5/4(=4.5j)$	4	aabb	5	Ady-ki- lences
21	Hosszú az Úristen	Nincsen apám, se anyám	28	üh. (szim.)	$6+6+8+6(=$ $3t+3t+4t+3t)$	4	aaxa	6	
22	Reggeli	Nincsen apám, se anyám	25. 03.	üh.	$5/3 5/4 [4.: 2]$	31	vegyes	1	
23	Ülni, állni, ölni, halni	Nincsen apám, se anyám	26 nyara	üh.	4/4	29 4	aabccddeeff...	2	felező nyolcas
24	Fiatalkorú asszonyok éneke	Nincsen apám, se anyám	25. 07–08.	üh.	4-es [sorvégeken néha 3-as] ütemű hosszú sorok	3 3 3 2 3 4	vegyes	6	

6. A szintén a költői életművet feldolgozó, utolsó függelék, a fonéma-statisztika megismétli JÉKEL–PAPP (1974) elemzését, csak ezúttal Ady helyett József Attila versei nyújtják az alapanyagot. A függelék tehát tartalmazza József Attila minden versének, versciklusának, illetve verskötetének fonémastatisztikáját, illetve ezeknek az egymástól való, az egyes fonémák által kifizetett 66 dimenziós vektortérben mért távolságát. Ezek irodalomtudományi-esztétikai értelmezése az Ady-statisztika esetében számos új megállapítás alapjául szolgált. A József Attila-i költői életmű ilyen bemutatásától hasonló várható, továbbá a két életmű összevetése is érdekes tanulságokkal szolgálhat majd.

Kulcsszók: lexikográfia, írói szótár, számítógépes nyelvészet, automatikus és félautomatikus szövegfeldolgozás, József Attila-szótár.

Hivatkozott irodalom

- BACHÁT LÁSZLÓ 1965. A Kölcsey-szótár. *Az MTA I. Osztályának Közleményei* 21: 388–391.
- Balassa = BALASSA JÓZSEF 1940. *A magyar nyelv szótára* 1–2. Grill, Budapest.
- BalassiSz. = JAKAB LÁSZLÓ – BÖLCSKEI ANDRÁS 2000. *Balassi-szótár*. Számítógépes Nyelvtörténeti Adattár 8. Debreceni Egyetem, Debrecen.
- Ball. = BALLAGI MÖR 1873. *A magyar nyelv teljes szótára* 1–2. Heckenast, Pest.
- BánkBSz. = BEKE JÓZSEF 1991. *Bánk bán-szótár*. Katona József Bánk bán című drámájának szókészlete. Katona József Társaság, Kecskemét.
- BENKŐ LÁSZLÓ 1965. A Juhász Gyula-szótár. *Az MTA I. Osztályának Közleményei* 21: 377–381.
- BENKŐ LÁSZLÓ 1979. *Az írói szótár. A szépirodalmi nyelv és stílus lexikográfiai feldolgozása*. Akadémiai Kiadó, Budapest.
- BÖLCSKEI ANDRÁS 1984. *Szövegfeldolgozás*. Egyetemi doktori értekezés. Kézirat. Debrecen.
- BÜKY LÁSZLÓ 2002. *Füst Milán metaforahasználatának alapjai – szótárszerű feldolgozásban*. Szegedi Tudományegyetem, Szeged.
- BÜKY LÁSZLÓ 2004. *Füst Milán metaforahasználatának szövegmondattanából – szótárszerű feldolgozásban*. Szegedi Tudományegyetem, Szeged.
- ČERMÁK, FRANTIŠEK 2008. An author's dictionary. The case of Karel Čapek. In: BERNAL, ELISENDA – DECESARIS, JANET (eds.) *Proceedings of the XIII Euralex international congress (Barcelona, 15–19 July 2008)*. Institut Universitari de Lingüística Aplicada Universitat Pompeu Fabra, Barcelona. 323–332.
- ČERMÁK, FRANTIŠEK – CVRČEK, VÁCLAV 2010. Author dictionaries revisited. Dictionary of Bohumil Hrabal. In: DYKSTRA, ANNE – TANNEKE, SCHOONHEIM (eds.) *Proceedings of the XIV Euralex international congress (Leeuwarden, 6–10 July 2010)*. Fryske Akademy, Afûk, Ljouwert, 592–598.
- CZF. = CZUCZOR GERGELY – FOGARASI JÁNOS 1862–1874. *A magyar nyelv szótára* 1–6. Emich–Athenauem, Pest/Budapest.
- CSENGERY KINGA 2006. Az elektronikus korpusz. In: ITTÉZS NÓRA szerk., 2006. *A magyar nyelv nagyszótára* 1. Segédletek. MTA Nyelvtudományi Intézet, Budapest. 18–19.
- CsokSztk. 1. = JAKAB LÁSZLÓ – BÖLCSKEI ANDRÁS 1993. *Csokonai-szókincstár* 1. Számítógépes Nyelvtörténeti Adattár 5. KLTE, Debrecen.
- CsokSztk. 2. = JAKAB LÁSZLÓ 2011. *Csokonai-szókincstár* 2. Számítógépes Nyelvtörténeti Adattár 12. Debreceni Egyetem, Debrecen.
- ELEKFI LÁSZLÓ [1996]. *Szókincsünk nyelvtani alakrendszere*. Kézirat. (Online alkalmazása [2006]: <http://corpus.nytud.hu/cgi-bin/e-szokincs/alaktan>.)
- ÉrtSz. = BÁRCZI GÉZA – ORSZÁGH LÁSZLÓ 1959–1962. *A magyar nyelv értelmező szótára* 1–7. Akadémiai Kiadó, Budapest.
- FORGÁCS TAMÁS 1993–1994. Zárt korpuszok és pótkompetencia. *Néprajz és Nyelvtudomány* 35: 17–23.
- FÜREDI MIHÁLY 1983. A mai magyar széppróza statisztikai vizsgálata. In: KISS JENŐ – SZÜTS LÁSZLÓ szerk., *A magyar nyelv rétegződése* 1. Akadémiai Kiadó, Budapest. 336–349.
- GÁLDI LÁSZLÓ 1958. Mutatvány a készülő Petőfi-szótárból. *Magyar Nyelv* 54: 322–333.
- GÁLDI LÁSZLÓ 1965a. Írói szótáraink fő kérdései, különös tekintettel a Petőfi-szótárra. Vitaindító előadás. *Az MTA I. Osztályának Közleményei* 21: 369–377.
- GÁLDI LÁSZLÓ 1965b. Mutatvány a Petőfi-szótárból. *Az MTA I. Osztályának Közleményei* 21: 403–409.
- GERSTNER KÁROLY 2006. A magyar nyelv nagyszótárának áttekintő története. In: ITTÉZS NÓRA szerk., 2006. *A magyar nyelv nagyszótára* 1. Segédletek. MTA Nyelvtudományi Intézet, Budapest. 10–17.

- GoetheW. = BERLIN-BRANDENBURGISCHEN AKADEMIE DER WISSENSCHAFTEN – AKADEMIE DER WISSENSCHAFTEN ZU GÖTTINGEN – HEIDELBERGER AKADEMIE DER WISSENSCHAFTEN (1966–). Goethe-Wörterbuch 1–. Kohlhammer, Stuttgart–Berlin–Köln–Mainz. = <http://gwb.uni-trier.de/>
- JAKAB LÁSZLÓ 2002. *A Jókai-kódex mint nyelvi emlék – szótárszerű feldolgozásban*. Számítógépes Nyelvtörténeti Adattár 10. Debreceni Egyetem, Debrecen.
- JAKAB LÁSZLÓ – BÖLCSKEI ANDRÁS 1988. *A XVI. századi orvosi könyv szóalakmutatója*. Számítógépes Nyelvtörténeti Adattár 4. KLTE, Debrecen.
- JAKAB LÁSZLÓ – KISS ANTAL 1978. *A Jókai-kódex ábécérendes adattára*. Számítógépes Nyelvtörténeti Adattár 1. KLTE, Debrecen.
- JAKAB LÁSZLÓ – KISS ANTAL 1980. *A Jókai-kódex szókinszének szófaji megoszlása*. Számítógépes Nyelvtörténeti Adattár 2. KLTE, Debrecen.
- JAKAB LÁSZLÓ – KISS ANTAL 1983. *A Birk-kódex ábécérendes adattára*. Számítógépes Nyelvtörténeti Adattár 3. KLTE, Debrecen.
- JAKAB LÁSZLÓ – KISS ANTAL 1994. *A Guary-kódex ábécérendes adattára*. Számítógépes Nyelvtörténeti Adattár 6. KLTE, Debrecen.
- JAKAB LÁSZLÓ – KISS ANTAL 1997. *Az Apor-kódex ábécérendes adattára*. Számítógépes Nyelvtörténeti Adattár 7. KLTE, Debrecen.
- JAKAB LÁSZLÓ – KISS ANTAL 2001. *A Festetics-kódex ábécérendes adattára*. Számítógépes Nyelvtörténeti Adattár 9. Debreceni Egyetem, Debrecen.
- JÉKEL PÁL – PAPP FERENC 1974. *Ady Endre összes költői műveinek fonémastatisztikája*. Akadémiai Kiadó, Budapest.
- JókaiSz. = BALÁZS GÉZA et al. 1994. *Jókai-szótár 1–2*. Unikornis, Budapest.
- JuhászSz. = BENKŐ LÁSZLÓ 1972. *Juhász Gyula költői nyelvének szótára*. Akadémiai Kiadó, Budapest.
- KÁROLY SÁNDOR szerk., 1965. *A Bécsi Kódex nyelvtana szótárszerű feldolgozásban. Morfológiai rész*. Akadémiai Kiadó, Budapest.
- KARPOVA, OLGA [КАРПОВА, ОЛЬГА] 1989. *Словари языка писателей*. Московский Полиграфический Институт, Москва.
- KARPOVA, OLGA 2011. *English author dictionaries (the XVIth–XXIst cc.)*. Cambridge Scholars Publishing, Cambridge.
- KELEMEN BÉLA 1912. *Magyar és német nagy kézi szótár, tekintettel a két nyelv szólásaira 1. Magyar–német rész*. Athenaeum, Budapest.
- KELEMEN JÓZSEF 1966a. Gépi adatgyűjtés és adatfeldolgozás a lexikográfia szolgálatában. In: ORSZÁGH LÁSZLÓ szerk., *Szótártani tanulmányok*. Tankönyvkiadó, Budapest. 29–54.
- KELEMEN JÓZSEF 1966b. Lyukkártyarendszerű adatgyűjtés és -feldolgozás a nyelvtudományban. *Nyelvtudományi Közlemények* 68: 343–380.
- KISS JÓZSEF 1965. [Hozzászólás az írói szótárak vitájához.] *Az MTA I. Osztályának Közleményei* 21: 392–395.
- KISS MARGIT 2012. A digitális Mikes-szótár. *Magyar Tudomány* 173: 279–284.
- KISS MARGIT – TŰSKÉS GÁBOR 2011. A Mikes-szótár. Kutatási beszámoló. *Magyar Nyelvőr* 135: 313–323.
- LADÓ JÁNOS 1962. Írói szótárak és „nyersszótár”-ak. *Magyar Nyelv* 58: 192–198.
- LEHR ALBERT 1882. *Arany János: Toldi. Költői elbeszélés. Nyelvi és tárgyi bő magyarázatokkal ellátta L. A. Franklin-társulat*, Budapest.
- LEHR ALBERT 1905. *Arany János: Toldi estéje. Költői elbeszélés hat énekben. Magyarázta L. A. Franklin-társulat*, Budapest.

- MakóiTsz. = H. KOVÁCS MIHÁLY 1926. *Tájszógyűjtemény [Makóról]*. A Magyar Nyelvtudományi Társaságnak az MTA Nyelvtudományi Intézetben, az Új magyar tájszótárnál letétbe helyezett kéziratos anyaga.
- MARTINKÓ ANDRÁS 1965. [Hozzászólás az írói szótárak vitájához.] *Az MTA I. Osztályának Közleményei* 21: 396–397.
- MÁRTONFI ATTILA 2008. Egy magyar rímszótár terve. In: BARTÓK ISTVÁN et al. szerk., „*Mielz valt mesure que ne fait estultie*”. *A hatvanéves Horváth Iván tiszteletére*. Krónika Nova Kiadó, Budapest. 198–204. = http://arsmetrica.elte.hu/articles/martonfi/0804_martonfi.pdf
- MATTAUSH, JOSEF 1990. Das Autoren-Bedeutungswörterbuch. In: HAUSMANN, FRANZ JOSEF et al. Hrsg., *Wörterbücher – Dictionaries – Dictionnaires. [...] An international encyclopedia of lexicography [...] 2*. Walter de Gruyter, Berlin – New York, 1549–1562.
- MlgeiSz. = SASS BÁLINT et al. 2010. *Magyar igei szerkezetek. A leggyakoribb vonzatok és szókapcsolatok szótára*. A Magyar Nyelv Kézikönyvei 21. Tinta Könyvkiadó, Budapest.
- MTsz. = SZINNYEI JÓZSEF 1893–1901. *Magyar tájszótár* 1–2. Hornyánszky, Budapest.
- Nszt. = ITTÉZS NÓRA szerk., 2006–. *A magyar nyelv nagyszótára* 1–. MTA Nyelvtudományi Intézet, Budapest.
- ÖcsödiSzólKm. = ENYEDI ANTALNÉ 1992. *Öcsödi szólások és közmondások*. Magyar Csoportnyelvi Dolgozatok 51. ELTE BTK – MTA Nyelvtudományi Intézet, Budapest.
- PAJZS JÚLIA 1990. *Számítógép és lexikográfia*. Linguistica. Series A. Studia et Dissertationes 4. MTA Nyelvtudományi Intézet, Budapest.
- PAJZS JÚLIA – MÁRTONFI ATTILA 2006. A magyar nyelv nagyszótárának számítógépes vonatkozásai. In: ITTÉZS NÓRA szerk., 2006. *A magyar nyelv nagyszótára* 1. Segédletek. MTA Nyelvtudományi Intézet, Budapest. 20–22.
- PAPP FERENC 1964. A magyar szókincs lyukkártya-rendszerű feldolgozása. *Magyar Nyelvőr* 88: 457–464.
- PAPP FERENC 1965. A Magyar Szóvégmutato Szótár és a vele kapcsolatos munkálatok. *Magyar Nyelv* 61: 187–200.
- PAPP FERENC 1975. Konkordancia: írói szótár előkészítése számítógépen. *Magyar Nyelvőr* 99: 351–355
- PAPP FERENC 2000. *A debreceni thesaurusz*. Linguistica. Series C. Relationes 11. MTA Nyelvtudományi Intézet, Budapest.
- PetőfiSz. = GÁLDI LÁSZLÓ szerk., 1973–1987. *Petőfi-szótár. Petőfi Sándor életművének szókészlete* 1–4. Akadémiai Kiadó, Budapest.
- RadnótiSz. = BEKE JÓZSEF 2009. *Radnóti-szótár. Radnóti Miklós költői nyelvének szókészlete*. Argumentum, Budapest.
- SCHLAPS, CHRISTIANE – SCHARES, THOMAS 2008. Lemmatisierungspraxis und -problematik im Autorenwörterbuch am Beispiel des Goethe-Wörterbuchs. In: BERNAL, ELISENDA – DECESARIS, JANET (eds.) *Proceedings of the XIII Euralex international congress (Barcelona, 15–19 July 2008)*. Institut Universitari de Lingüística Aplicada Universitat Pompeu Fabra, Barcelona. 599–606.
- J. SOLTÉSZ KATALIN 1965. Nyelvstatisztikai vizsgálatok a Petőfi-szótár anyaga alapján. *Nyelvtudományi Közlemények* 67: 345–354.
- SZATHMÁRI ISTVÁN 1965. [Hozzászólás az írói szótárak vitájához.] *Az MTA I. Osztályának Közleményei* 21: 397–403.
- SzegSz. = BÁLINT SÁNDOR 1957. *Szegedi szótár* 1–2. Akadémiai Kiadó, Budapest.
- SzGySz. = FÜREDI MIHÁLY – KELEMEN JÓZSEF 1989. *A mai magyar nyelv szépprózai gyakorisági szótára (1965–1977)*. Akadémiai Kiadó, Budapest.

- THIBAUT, M. A. [HAAS, JOHANN GOTTFRIED] – KOVÁTS S. JÁNOS 1901². *Francia–magyar és magyar–francia szótár 2. Magyar–francia rész*. Franklin-társulat, Budapest.
- ToldiSz. = Pásztor Emil 1986. *Toldi-szótár*. Arany János Toldijának szókészlete. Tankönyvkiadó, Budapest.
- TolnaiMagySz.³ = TOLNAI VILMOS szerk., 1928³. *Magyarító szótár a szükségtelen idegen szavak elkerülésére*. Eggenberger, Budapest.
- TOLNAI VILMOS 1917. *Szómutató Lehr Albert Arany-magyarázataihoz*. Franklin-társulat, Budapest.
- TÖRÖK GÁBOR 1976. *József Attila-kommentárok*. Gondolat, Budapest.
- TVERDOTA GYÖRGY 1986. József Attila nyelvészeti kájának forrásai. *Irodalomtörténeti Közlemények* 90: 373–392.
- ÚMTsz. = B. LŐRINCZY ÉVA szerk., 1979–2010. *Új magyar tájszótár* 1–5. Akadémiai Kiadó, Budapest.
- VégSz. = PAPP FERENC 1969. *A magyar nyelv szövegmutató szótára*. Akadémiai Kiadó, Budapest.
- VINOGRADOV, VIKTOR VLAGYIMIROVICS [ВИНОГРАДОВ, ВИКТОР ВЛАДИМИРОВИЧ] ред. 1956–1961. *Словарь языка Пушкина* 1–4. Государственное Издательство Иностраных и Национальных Словарей, Москва.
- WACHA IMRE 1959. Mutatvány a készülő József Attila-szótárból. *Magyar Nyelvőr* 83: 274–288.
- WACHA IMRE 1961. A magyar írói szótárak kérdése. *Magyar Nyelvőr* 85: 189–203.
- WACHA IMRE 1962. Az Ibsen-szótár szerkesztési tanulságai. *Magyar Nyelv* 58: 199–206.
- WACHA IMRE 1965. A József Attila-szótár. *Az MTA I. Osztályának Közleményei* 21: 381–388.
- WEISZ DEZSŐ [1940 e.]. *A Debreceni Codex nyelve és szókincse*. Kézirat. Debrecen.
- ZaySzkt. = JAKAB LÁSZLÓ – BÖLCSKEI ANDRÁS 2003. *Egy XVI. századi emlékirat szókincstára. Zay Ferenc: Az Landor feyrwar el wezessenek oka e woth es igy esewth (1535 k.)*. Számítógépes Nyelvtörténeti Adattár 11. Debreceni Egyetem, Debrecen.
- ZrínyiSz. = BEKE JÓZSEF szerk., 2004. *Zrínyi-szótár. Zrínyi Miklós életművének magyar szókészlete*. Zrínyi-könyvtár 5. Argumentum Kiadó, Budapest.

Computer and author dictionary – with special regard to the author dictionary on Attila József’s oeuvre in hand

The two most important predecessors of the genre of the author dictionary are concordances and glossaries – from the ancient world. It was in the 19th and 20th centuries (in Hungary chiefly from the 1950’s) that the modern author dictionary came into existence, with backgrounds from historical-explanatory dictionaries, on a corpus linguistic basis. In parallel with this, from the 1960’s, computational linguistics developed, and this fact afforded remarkable possibilities to producing diverse dictionaries, close-structured vocabularies, and concordances. Such works – mainly based on earlier literary monuments – were compiled above all in Debrecen. Nowadays, language technology tools enable and increasingly support the automatic (or at least semi-automatic) way of the dictionary compilation. Because of this, lexicographers need to rethink and redefine their duties. The main part of the study reports on such an effort, and shows – considering both the entries and the appendices – the language technology solutions applied in the author dictionary on Attila József’s oeuvre.

Keywords: lexicography, author dictionary, computational linguistics, automatic and semi-automatic text processing, an author dictionary on Attila József’s oeuvre.

MÁRTONFI ATTILA
MTA Nyelvtudományi Intézet