

A magyar nyelv nagyszótárának metamorfózisai: a cédulagyűjteménytől az elektronikus adatbázisokig

ITTZÉS NÓRA

A magyar nyelv nagyszótárát (Nsz.)¹ a magyar nyelv tudomány legrégebbi szótári vállalkozásaként, ugyanakkor a magyar lexikográfiában új szótártípus megtestesítőjeként tartja számon a szakma. Újdonsága nem csupán nagyszótár voltában és az ezzel együtt járó korpuszalapúságában van — hiszen történeti jellegű szótár írása nem képzelhető el valamiféle szövegtest, szövegalap, ún. korpusz nélkül —, hanem abban, hogy számítógépes adatbázisként készül. Előadásomnak tehát szükségszerűen két súlypontja, két iránya van: a korpusz és a szótár. Az Nsz. esetében e kettő elválaszthatatlan volt: a korpusz nem önmagában épült, mintegy nyitva hagyva az utat a legkülönbözőbb kutatások számára (ahogy például a Magyar nemzeti szövegtár), hanem kifejezetten az Nsz. számára. Nagyszótári céllal gyűjtötték össze az 5-6 millió betűrendbe rendezett cédulából álló archivális gyűjteményt (NSz.) — miközben persze nem volt még kialakult lexikográfiai koncepció, csak valami homályos, a nagyszótári műfajból táplálkozó kép arról, hogy milyen művekre kell támaszkodnia egy efféle szótárnak —, s az Nsz. számára kezdték el építeni a szótár elektronikus formában tárolt korpuszát, a Magyar történeti szövegtárat (MTSz.) is, eredetileg azzal a szándékkal, hogy a teljes nyomtatott korszakot² felölelje. Az, hogy a korpusz a lexikográfiai kutatás mellett számos más kutatás számára is gazdag forrás lehet, csak ráadás. Az MTSz. és a szótári adatbázis kettős pillérére mára gazdag rendszer épült. Ha azt szeretném most röviden bemutatni, milyen megoldásokat, módszereket tett lehetővé a digitális korszak az Nsz. számára, nem csupán a digitális archiválásról, hanem olyan rendszerről kell beszámolnom, amely a lexikográfiai munka legkülönbözőbb területeit megkönnyíti, magában foglalja. Előadásomban beszélek tehát a (részben) elektronikus korpuszról, e korpusz elektronikus forrásjegyzékéről és az elektronikus adatbázisként készülő szótárról.

¹ A szótárt a munkálatok során többféle címmel is jelölték (A magyar irodalmi és köznyelv nagyszótára, Akadémiai nagyszótár), a mai címváltozat (A magyar nyelv nagyszótára) a szótári sorozat 1. Segédletek és 2. A–azsúrozó kötetének 2006-os megjelenésekor vált véglegessé.

² A korszakhatárt Komjáti Benedek Szent Pál levelei magyar nyelven című műve határozta meg, amely 1533-ban jelent meg Krakkóban Hieronymus Vietor nyomdájában. Az első teljes terjedelmében magyarul nyomtatott könyv Erasmus latin bibliafordítása alapján készült.

1. A korpusz

Amikor az 1980-as évek közepén Herman József igazgatósága alatt a nagyszótári munkálat tetszhalott állapotából újraindult, megszületett a döntés is: a szótár számítógépes adatbázisként készüljön, alapját számítógépen tárolt szöveggyűjtemény jelentse. A nagyszótári munkaközösség belső szóhasználatában azonban ma korpuszként nem csupán a MTSz.-t, vagyis a modern korpusznyelvészet tapasztalatai alapján eltervezett, bár végső megvalósításában messze nem tökéletes szöveggyűjteményt emlegetjük, amely az adott korszak írásbeliségének sajátosságait, műfaji arányait figyelembe véve válogatott szövegeket, hosszabb szövegrészeket tartalmaz. Számos (például szógyakorisági, pragmatikai) kutatásban nyilvánvalóan kizárólag az ilyen típusú par excellence korpuszok használatától várható megbízható, hiteles eredmény. A korpuszalapú lexikográfia azonban éppen e klasszikus korpuszok statisztikai jellemzőiből adódóan számos problémával küzd. Az ilyen korpuszokban mindig nagyon magas a hapax legomenonok száma (minél nagyobb a szövegtest, annál több benne az egyszer használt szó) — ezek szótárázása rendszerint nem szükséges —, ugyanakkor bizonyos szavak előfordulási gyakorisága komoly elvi és gyakorlati nehézségek elé állítja a szótárírókat. Egy konkrét példa: az Nszt. *az* szócikkének írásakor kiderült, hogy az akkori korpuszban (amely 2006-ra bővült 27,5 millió szövegszóra) több mint 3 millió adata van e lexémának. E címszó esetében az Nszt.-nek az a sajátossága is különleges megoldásokat tett szükségessé, hogy a szótár minden értelmezett szócikk-egység — jelentés, jelentésárnyalat, értelmezett szókapcsolat — esetében megadja a korpuszbeli első előfordulást. A két — mutató névmási és névelői — szófaj több mint 80 értelmezésének illusztrálásához kellett megtalálni a legkorábbi adatot: tút a szénakazalban. Ahhoz tehát, hogy egy korpusz megbízhatóan tartalmazzon használható adatot egy általános értelmező szótár szócikkének címszavához, értelmezendő jelentéséhez, lényegesen nagyobb terjedelmű korpuszra volna szükség, mint az MTSz. Amikor az 1980–90-es évek fordulóján elkezdődött a próbaszócikkírás, nagyon hamar világossá vált, hogy a célzott gyűjtéssel összeállított, mintegy 600 ezer szótári szó adatait felölelő cédulaanyag megkerülhetetlen, és nagy valószínűséggel található benne adat azokra a lexémákra is, amelyek a 160–170 ezer címszavas MTSz.-ből hiányoznak. Így ma e kibővített korpusz részének tekintjük az archivális cédulagyűjteményt, akárcsak az ún. kiegészítő szöveggyűjteményt, amely számos kereskedelmi forgalomba került cédé szöveges adatbázisát foglalja egybe. E cédék felhasználásának szükségessége nagyjából a 2000-es évek elején, a szerkesztési szabályzat elkészülte után vált nyilvánvalóvá, amikor megindulhatott a betűrend szerint haladó módszeres szócikkírás. E cédék eredetileg nem a szótári munka számára készültek, de összeválogatásukban és filológiai kezelésükben nagyszótári szempontok érvényesültek. E több százmillió szövegszónyi kiegészítő szöveggyűjtemény most már — legalábbis a szóelőfordulásokat illetően — többé-kevésbé kielégíti a szótárírás igényeit.

A lexikográfiai elemzés alapját és az Nszt. példamondatainak forrását jelentő korpusz 1772 és 2000 között keletkezett publikált (nyomtatott vagy elektronikus adathordozón közzétett) szövegeket, szövegrészleteket tartalmaz, mostani állapotában jelentős túlsúlya van a szépirodalmi szövegeknek. A szövegrészletek terjedelme nyilvánvalóan a cédulaanyag esetében a legszűkebb, hiszen itt a néhány szavastól a néhány mondatosig terjedő szöveggörnyezet szerepel. A szótár sem kézirat, sem online forrásokat nem dolgoz fel, mivel meghatározó filológiai elve, hogy adatai ellenőrizhetők, az eredeti forrásból visszakereshetők legyenek.

Annak ellenére, hogy a történeti szempont érvényesítése és ezzel együtt az adatolás az értelmező szótárak legteljesebb típusát képviselő Nszt.-ben az első pillanattól kezdve megkérdőjelezhetetlen volt, még a szótári céllal létrejött gyűjteményekre is rányomta bélyegét, hogy tényleges lexikográfiai koncepció nélkül és bizonyos szempontból filológiailag is átgondolatlanul történt a korpuszépítés. Most nem a kézi rögzítéssel — legyen szó akár kézírásos vagy írógépes cédulákról, akár számítógépen tárolt szövegekről — szükségszerűen együtt járó szöveghibákról beszélek, hanem a filológiai elvek változásairól és átgondolatlanságáról, ami a cédulagyűjtemény megindulásánál csaknem száz évvel későbbi MTSz. esetében is feltűnő volt. A cédulákon a forrásadatok megjelölésének pontatlanságai vagy utólagos azonosíthatatlanságai jelentenek gondot, de korábban sok kívánnivalót hagyott maga után az MTSz. filológiai megbízhatósága is, elsősorban abból adódóan, hogy nem gondolták át előzetesen, hogy a szótári szöveg példamondataiban milyen szempontok szerint és milyen adatokkal tüntetik majd fel a forrást. A kereskedelmi forgalomba került szöveges cédék anyagát pedig mindannyian ismerjük: amennyire megkönnyítheti a kutatást nagy mennyiségű szöveg számítógépes kereshetősége és kezelése, a digitalizálás gyermekbetegségeit mutató szövegek vizsgálatokor a filológiai pontosság elérése rendszerint nagyon komoly háttérmunkát igényel. Arról, hogy ez a háttérmunka ténylegesen mit jelentett az Nszt. esetében, a későbbiekben még szólok.

1.1. Az archivális cédulagyűjtemény

Jóllehet az NSz. esetében a leghagyományosabb rögzítési formáról van szó, amelynek semmi köze nincs a digitális archiváláshoz, néhány szót mégis szólnom kell e gyűjteményről, nem csupán azért, mert a szótárírás szempontjából egységes rendszerként kezeljük a három szövegtárat. Ez a hagyományos adatgyűjtés 1898-ban kezdődött meg, eleinte spontán módon, majd az Akadémia szervezésében és irányítása alatt folyt a cédulázás egészen a 60-as évek elejéig. E gyűjteményben egyébként az 1750 és 1772 közötti időszak forrásaiból is található cédulák, hiszen a korai tervek szerint az Nszt. a Nyelvtörténeti szótár (NySz.) utáni időszakot dolgozta volna fel. A mintegy 5-6 millió cédula betűrendbe rendezése kisebb-nagyobb megszakításokkal 2005-ig folyt, de egy néhány évvel ezelőtti pincefelújításkor is találtunk több százezer rendezetlen, illetve azonosíthatatlan forrású cédulát. E hatalmas cédulagyűjtemény címszólistáját

annak idején — méltó, hogy név szerint említettessék — Gölniczki Margit, Kiss Lajos titkár- és gépírónöje legépelte. Az elmúlt években sikerült digitalizálnunk ezt a két hasámban mintegy 7300 oldalnyi betűrendes jegyzéket, amely az 1750 és 1960 közötti magyar szókészlet talán leggazdagabb mutatója, és amely néhány éve hozzáférhető az Nszt. honlapján (<http://nszt.nytud.hu/cszej.html>).

Magának az archivális cédulagyűjteménynek a digitalizálása is hosszú évek óta téma, sőt egy évvel ezelőtt komoly ígéretté vált: a Nyelvtudományi Intézet és a Magyar Nemzeti Digitális Archívum és Filmintézet, valamint az Arcanum Adatbázis Kft. megállapodott arról, hogy a MaNDA közfoglalkoztatottai az Arcanum technikai felszerelésével digitalizálják az NSz. céduláit. Az Arcanum korábban már digitalizált cédulagyűjteményeket, például az Egyetemi Könyvtár katalógusát. Természetesen e képként rögzített anyagot csak kézi munkával lehetne kereshetővé tenni, hiszen a karakterfelismertető programok ez esetben alkalmazhatatlanok, és ilyen mennyiségű cédula szövegének kézi átírása is reménytelen feladat. A gyakorlatban viszont jól alkalmazható megoldás, hogy bizonyos számú (50-100) cédulánként kereshető formában beírják a címszót, így a kisebb halmazban már könnyedén megtalálhatók a keresett szó képként rögzített cédulái is. A nagyszótári cédulaanyag ilyen rögzítése azonban még az Arcanum technikai felszerelésével is túlságosan nagy feladatnak bizonyult: elsősorban a cédulák fizikai paramétereinek egyenetlenségei nehezítették meg a munkát. A rendelkezésre álló idő alatt csak a gyűjteménynek mintegy 20 százalékát sikerült digitalizálni, s ahhoz, hogy ez interneten hozzáférhető és kereshető legyen, komoly utómunka szükséges. Az, hogy az intézetnek a jövőben sikerül-e, és milyen módon folytatnia az együttműködést az Arcanummal, egyelőre teljesen bizonytalan.

Az archivális cédulaanyagról szólva még egy mozzanatot ki kell emelnem. Ahogy említettem, e cédulák magukon viselik a kézi rögzítés gyengeségeit: el- és átírások, a grafémahasználat tükrözésének egyenetlensége, több mondatból kompilált vagy túlságosan szűk szövegekörnyezet, bibliográfiai pontatlanságok, oldalszámhiányok stb. nehezítik a szótárírói munkát. Mivel e szövegtárból származik a korai adatok jelentős része, igen nagy azoknak a példamondatoknak a száma, amelyeket a kicédulázott forrásból (az Országos Széchényi Könyvtárban, a Fővárosi Szabó Ervin Könyvtárban stb.) ellenőrizni kell. E munka megkönnyítésére az elmúlt években a szótár munkatársai közel kétezer tételt gyűjtöttek össze az Nszt. internetre felkerült digitalizált forrásai közül, nem utolsósorban a GoogleBooksról, jelentős tapasztalatra téve szert a magyar nyelvű „nemzeti szövegkincs” digitálizálásának anomáliáival kapcsolatban. Ez a — saját erőből karakterfelismertetéssel kereshetővé tett — forrásanyag csak belső szerverünkön, az osztály munkatársai számára hozzáférhető.

1.2. Magyar történeti szövegtár

Az Nszt. elsődleges forrásának az 1985 és 2005 között épített, mai állapotában 27,5 millió szövegszavas MTSz.-t tekintjük, amely az első elképzelések

szerint, ahogy említettem, 1533-tól tartalmazott volna szövegeket. Ha a lexikográfiai szempontok megengedik, a szócikkekben ebből a szövegtárból idézünk. A szöveggyűjtemény műlistáját az adott korszakot jól ismerő irodalomtörténészek állították össze. A szövegeket kézzel rögzítették, a szöveg-, sőt betűhűség érdekében a mellékjeles betűket a latin ábécé betűit számokkal kombináló ún. Prószéky-kódokkal helyettesítve. A rögzítés alapjául általában az első vagy a kritikai kiadások szolgáltak. A szöveg címkézése kezdetben egy erre a célra kidolgozott kódrendszerrel történt, később a kódolást a TEI (Text Encoding Initiative) rendszeréhez igazították. Cél volt a gépi lekérdezés biztosítása is, ezért a 80-as évek végén megkezdődött a rögzített szöveg nyelvtechnológiai feldolgozása a MorphoLogic Kft. munkatársai segítségével. Morfológiai elemző program készült, ennek tesztelése az MTSz. szövegein történt, az együttműködés a ma már általánosan használt helyesírás-ellenőrző program javításához, pontosításához is segítséget nyújtott. Feladat volt az elemzett korpusz ún. egyértelműsítése is (amely segítette annak eldöntésében, hogy például a homonimák közül melyik szó paradigmatis adata szerepel a szövegben). Ma már elmondható, hogy a magyarországi korpusznyelvészeti kutatások bölcsője az Nszt. volt, számos fejlesztéshez ennek korpusza nyújtotta az alapot.³ Természetesen ebben a munkában is megkerülhetetlen volt a kézi beavatkozás és ellenőrzés. Ebben a munkafolyamatban az is kiderült, hogy a régi szövegek számítógépes feldolgozásának egy agglutináló nyelv esetében olyan nehézségekkel kell megküzdenie, amely beláthatatlan messzeségbe tolná a tényleges lexikográfiai munka megkezdését és a szótár megjelenését, ezért döntés született az 1772-es kezdő dátumról.⁴

A 90-es évek végén, a próbaszócikkírás során a kb. 22 millió szövegszavas elemzett korpuszt (<http://www.nytud.hu/hcc/>) használtuk, a 2000-es évek első felében azonban megkezdődött a rendszeres szócikkírás, és megtörtént az MTSz. első jelentős bővítése. Sem pénz, sem személyi kapacitás nem volt arra, hogy a szükséges morfológiai elemző programokat lefuttassuk, és a szükséges ellenőrzéseket végrehajtsuk, ma tehát a szótárírásban az elemzetlen korpuszra szorulunk. Ugyanakkor jelentős előrelépés volt, hogy az MTSz. szövegeit és a szótár kiegészítő szöveggyűjteményének anyagát ma már egységes keresőfelületen tudjuk lekérdezni. Ma ezen a felületen nfo kiterjesztésű fájlokban keressük az egységes adatbázisba rendezett cédés forrásokat, az MTSz.-t, sőt a már megjelent kötetek

³ A nyelvtechnológiai munkálatokhoz vö. PAJZS–MÁRTONFI 2006: 20.

⁴ 1994 áprilisában a nagyszótári munkaközösség Kiefer Ferenc igazgató és jeles lexikográfusok meghívásával megbeszélést tartott a nagyszótári próbaszócikkekről és tipográfiai mintákról, jóllehet korábban még sem a címszám, sem a szerkesztés alapvető elvei, sem a feldolgozott korszak időhatárai nem váltak véglegessé. Ezen az ülésen ismertette Villó Ildikó azokat a tapasztalatait, amelyeket a próbaszócikk írásakor a cédulagyűjtemény és az elektronikus korpusz használhatóságával kapcsolatban szerzett, és ekkor vetődött föl először az 1772-es kezdő korszakhatár. 1997 júniusában újabb szakmai tanácskozássra került sor, amely meghozta a végső döntést: 1772 és 2000 közötti idézetekkel dolgozik a szótár.

anyagát vagy a legkorábbi cédulás források karakterfelismertetéssel létrehozott szövegét is. A program rafinált és sok esetben nehézkes lekérdezést (ún. dzsóker-karakterek és logikai kifejezések használatát) tesz szükségessé a speciális karaktereket alkalmazó, a szöveg történeti jellege miatt morfológiailag sem csupán a mai nyelvhasználatot tükröző anyagban, különösen a több szóból álló szerkezetek (pl. elváló igekötős igék) keresésekor, ugyanakkor hallatlan előnye a gyorsaság.

1.3. Kiegészítő szöveggyűjtemény

Az Nszt. harmadik szövegtára a kiegészítő szöveggyűjtemény, amely kereskedelmi forgalomba került, eredendően nem a lexikográfiai munka számára készült adathordozók meglehetősen vegyes anyagát tartalmazza. Az ezeken az adathordozókon tárolt gyűjteményből csak a szótári munka számára megfelelő adatállományokat emeltük be a korpuszba, s természetesen a szócikkekben sem a cédére, hanem ezekre az adatállományokra hivatkozunk. Vagyis azokra a nyomtatott szövegekre, amelyek alapján az adathordozón rögzítették a szövegeket. Mivel e cédék kiadását a filológiai szempontok alig-alig határozták meg, a szövegek minősége a szokásosnál is nagyobb figyelmet és kritikai megközelítést tesz szükségessé a szócikkírók és a szerkesztők részéről.

2. Forrásjegyzék-adatbázis

A filológiában jártas embereknek e rövid ismertetés alapján is nyilvánvaló lehet, hogy e háromféle szöveggyűjtemény (viszonylag) egységes kezelése elengedhetetlen feltétele a megbízható szótári munkának, illetve a történeti szótári szempontok érvényesítésének, ugyanakkor számos gyakorlati és technikai nehézséggel jár. Az adatközlés (névformák, címek, időszaki kiadványok hivatkozása stb.) egységesítése azt igényelte, hogy gyakorlatilag pontról pontra ellenőrizni kellett a forrásjegyzék 30 ezernél több bibliográfiai tételét: kialakítani a szerzők és a fordított művek eredeti szerzőinek névformáját, ellenőrizni a gyűjteményes művek, antológiák különböző szövegeinek szerzőségét (pl. a tartalomjegyzékből) és nem utolsósorban a művek keletkezési idejét, mivel ez meghatározó fontosságú egy történeti jellegű szótárban. Meggyőződésem szerint az a filológiai munka, amelyet a szótár munkatársai elvégeztek, igen jelentős eredményeket hozott az irodalomtörténeti kutatás számára is. Egyetlen példa: az eredeti kiadások bibliográfiai adatainak ellenőrzésekor, az egyes művek előszavának végigolvasásakor derült ki, hogy számos 1772-es évszámmal számon tartott mű keletkezési éve téves. 1772 nem a szöveg létrejöttének, hanem kiadásának éve. E tény talán azt jelzi, hogy valamiféle könnyítés történhetett a cenzúravizonyokban, és ezért jelenhetett meg ebben az évben több korábban altagott mű (pl. Sartori Bernát: Magyar nyelven filozófia, Bessenyei György több műve).

Petárdák még mindig robbanhatnak körülöttünk, de legalább már nem puska-poros hordón ülünk: e jelentős filológiai háttérmunka nyomán létrejött egy

ugyan még mindig nem hibátlan, de már megbízható forrásadatbázis az Nszt. három szövegtárának forrás- és névjegyzékeivel. A néhány évvel ezelőtti fejlesztések nyomán a forrásadatbázis XML-formátumban a külső felhasználók számára is hozzáférhető az Nszt. honlapján (<http://nszt.nytud.hu/forr.php>).

3. A szótári szöveg mint adatbázis

A 20. század második fele a nagyszótárirás újabb hullámát indította el, már az elektronikus szótárak időszakában, amelyek közül elsősorban a francia Trésor (1971–1994) és a BATTAGLIA-féle (1961–2002) olasz szótár ötletadó mintaként szolgálhatott a Nyelvtudományi Intézetben újrainduló munkálat számára is.

Nemcsak a szótár szövegtárai, forrásjegyzékei, különböző segédletei tárolódnak tehát számítógépen, hanem maga a szótári szöveg is adatbázisként, XML (= Extensible Markup Language) formátumban, a szövegek tartalmi-logikai egységeinek azonosítására szolgáló nemzetközi szabvány szerint készül, így akár elektronikus adathordozón, akár nyomtatott formában meg lehet jelentetni. A nagyszótári munkaközösség a harmadik szócikkíró programot (XMetal) használja, mióta a próbaszócikkírás, illetve a módszeres lexikográfiai munka folyik.

A szótári struktúra egységeit jelölőelemek, ún. tagek (vö. angol *tag* 'címke') azonosítják. Ezekhez a logikai címkékhez rendelődnek hozzá a szócikk fizikai megjelenésének paraméterei akár elektronikus, akár hagyományos nyomtatott formában. E formátumban a szócikkek típusa, a szerkezeti elemek sorrendje és hierarchiája előre meghatározható, illetve szükség esetén módosítható, bővíthető. A dokumentumtípusok definiálása és a hozzájuk tartozó séma használata biztosítja, hogy a szócikkek formai tekintetben szigorúan egységes módon készülhessenek. Ez a formátum lehetőséget ad a szótár mennyiségi mutatóinak nyilvántartására, illetve a már kész szócikkek folyamatos karbantartására, ellenőrzésére és nem utolsósorban a folyamatos bővítésére, frissítésére is. Nemcsak a címszavak, a jelentések és a példamondatok bővítését, kiegészítését vagy a hiányzó etimológiai, szótörténeti és frazeológiai információk szótárba emelését teszi lehetővé a későbbiekben, hanem azt is, hogy az Nszt. különböző típusú és műfajú szótárak (például történeti, kézi- vagy szinonimaszótár) alapszótárává válják.

A szótár bővítésének alapja természetesen a korpusz kibővítése lehet akár korhatáron belüli új művekkel — ahogy történt a 2000 utáni években —, akár a korhatár módosításával. Az elmúlt évben meg is indult a MTSz. legújabb bővítése, részben a mostani, 1772 és 2000 korpusz egyetlenségeinek kiigazításával, részben a felső korhatár 2010-re történő emelésével. Gyakorlatilag — néhány szórványos hiánytól eltekintve — befejeződött a szépirodalmi és szaknyelvi szövegek, periodikák kiválogatása, megtörtént a szaknyelvi szövegek rögzítése, megkezdődött az ellenőrzés és a fejlécezés, vagyis a bibliográfiai adatok rögzítése. Terveink szerint 2015-ben az XML-formátumú MTSz. felső korhatára 2010-re emelkedik. Ez természetesen magával hozza azt is, hogy a közeljövőben a

munkaközösségnek nemcsak újabb szótári köteteket kell létrehoznia, hanem meg kell oldania azt is, hogy viszonylag rövid időn belül az eddig elkészült betűk anyagába is bekerüljenek az új korpusz adatai.

A szótári adatbázissal kapcsolatban még egy — reménységünk szerint nem távoli — tervet szeretnék megemlíteni: az online hozzáférést a szótári adatbázishoz. Elképzelésünk szerint ez nem a nyomtatott forma PDF-jének olvasását biztosítja majd, hanem az XML-formátumú adatbázis adta lehetőségeket kihasználva saját felületen keresztül teszi lehetővé a legkülönbözőbb kereséseket.

Az elektronikus szótárak már említett időszakában sokak számára azzal az illúzióval járt, hogy ez feltétlenül és mindenben megkönnyíti majd a szótárírást. A digitalizálás hatalmas eredményei ellenére e korszak szótárainak — köztük az Nszt.-nek — a keletkezéstörténete azonban jól példázza, hogy a technikai változások sem tudják lényegesen meggyorsítani a nagyszótári jellegű lexikográfiai leírás munkáját. Nyelvi elemzés nélkül nincs szótárírás. Azt pedig soha nem fogja helyettünk elvégezni a számítógép.

Irodalom

- BATTAGLIA, SALVATORE szerk. 1961–2002. *Grande dizionario della lingua italiana 1–21*. Torino, UTET.
- Nszt. = ITTZÉS NÓRA főszerk. 2006–2013. *A magyar nyelv nagyszótára. 1–5*. Budapest, MTA Nyelvtudományi Intézet.
- NySz. = SZARVAS GÁBOR–SIMONYI ZSIGMOND 1890–1893. *Magyar nyelvtörténeti szótár 1–3*. Budapest.
- PAJZS JÚLIA–MÁRTONFI ATTILA 2006. A magyar nyelv nagyszótárának számítógépes vonatkozásai. In: ITTZÉS NÓRA főszerk., *A magyar nyelv nagyszótára. 1. Segédletek*. Budapest, MTA Nyelvtudományi Intézet. 20–22.
- Trésor = IMBS, PAUL szerk. 1971–1994. *Trésor de la langue française. Dictionnaire de la langue du XIXe et du XXe siècle. 1789–1960. 1–16*. Párizs, Éditions du Centre national de la recherche scientifique.