

Nyelv négy fal között

Év végére már 2010-es szövegek példamondataival is dolgozhatnak a nagyszótár szerkesztői

Ha most írnák a görög mítoszokat, akkor Szisziüposz nem követ görgetne fel a hegyre. Sokkal inkább lenne nyelvész, aki milliónyi cédula között vadássza a nyelv folyton változó és átalakuló szavait, miközben azok folyton tovább röppennek. Majdhogynem kétszáz éve készül a magyar nyelv nagyszótára.

Papp Sándor Zsigmond

Ki akarna bányamérnök lenni több száz méterrel a föld alatt, vagy olajfúró munkás egy tornyon, valahol a háborgó óceánon? Esetleg szótárszerkesztő, aki több százezer szövegszóból halássza ki a nagyszótárba bekerülő címszavakat? Pedig ha eltölt néhány órát a szótár méreteihez képest meglepően törékeny főszerkesztővel, Ittész Nórával, akkor könnyen lehet, hogy már másnap keresgélni kezd kedvenc lapjában vagy könyvében a ritkább kifejezések után. Mert igenis van ebben valami lelkesítő.

Ez a történeti jellegű értelmező szótár a magyar nyelvtudomány eddigi legnagyobb szótári vállalkozása. Címszavainak jelentéseit példamondatokkal gazdagon illusztrálva mutatja be. A tavalyelőtt megjelent ötödik kötet a c (cé, cz, tz) betű címszavától a „dézsmaszédőig” gyűjtötte össze az elmúlt 230 év szókészletét. A kötet 4122 szócikkben dolgozza föl a szavakat, szókapcsolatokat, használatukat 35 157 példamondattal érzékelteti.

Ha minden jól megy, jövőre jön a hatodik, az igen fontos „egy” szavunkkal. Optimális esetben még húsz vagy harminc év kellene a végéig (húsz fölött lenne a megjelent kötetek száma, hiszen csak a kezdőbetűs szavainknak három kötet kell), ám Ittész Nóra erre is csak megengedően mosolyog: igazából több évtized kell egy efféle munka befejezéséhez. Mármint a printverzió elkészüléséhez, mert a folvama-

Végül a Magyar Tudományos Akadémia úgy döntött, hogy ideje kézbe venni a szervezést, így már célzott gyűjtés is folyt az irodalmi és köznyelvből. Több ezer mű áttekintése után jegyezték ki az érdekes és ritka szavakat az erre képzett szakemberek. A gyűjtés a hatvanas évekig tartott, és öt-hatmillió cédula gyűlt össze. Szinte kezelhetetlen mennyiség. Elég nagy fejtörést is okozott a nyelvtudományi szakembereknek, hogy mit kezdjenek velük. Végül az ötvenes években elkezdődött a cetlik betűrendbe rendezése, előszerkesztése. Eleinte mindössze egy-két ember dolgozott ezen, és nem is lettek sokkal többen, így nem csoda, hogy csak 2005 után fejeződött be a rendezés. (A cédulák különben nagy szerepet játszottak a hétköztetes A magyar nyelv értelmező szótára összeállításában is.)

A nyolcvanas években úgy tűnt, a technika végleg elsepri a hirtelenjében kőkorszaki módszernek tűnő cetlizést. Az MTSZ megalapításával elektronikus korpuszt kezdtek építeni, egész műveket gépeltek le (később szkenneltek), kezdve Komjáthy Benedek 1533-ban megjelent Szent Pál leveleitől, amely az első, teljes egészében magyar nyelvű kiadvány. De a Károli Biblia részleteit is elektronizálták, mint ahogy Mi-

kes leveleiből is néhány passzust. Ám hamar kiderült, hogy ez sem teljesen célravezető, hiszen nem old meg minden nehézséget. Bár sokkal könnyebbé tette a keresést, az ismétlésekkel nem tudott megbirkózni. Képzelték csak el, hányszor fordulhatnak elő a szövegtárban a hátrazott névelők. (Segítünk: a mai korpuszban több mint hárommilliószor.) Később már csak szövegrészleteket rögzítettek a legkülönbözőbb stílusrétegekből. Szakemberek választották ki a reprezentatív könyveket és passzusokat, amelyek hűen tükrözik az adott korszak szóhasználatát. A XX. századért például a József Attila-szakértő Tverdo-ta György felelt. De az is sokat segített, hogy a munka felgyorsítása érdekében lemondtak a régi szövegekről, 1772-nél húzták meg a határt. Az eddigi kötetek így 1772 és 2000 között keletkezett, nyomtatásban megjelent szövegek példamondataival dolgoztak. Annak érdekében, hogy a szótár naprakész legyen, most épp a korpusz kibővítésén dolgoznak: 2001 és 2010 között keletkezett szépirodalmi és tudományos szövegekkel szeretnék kiegészíteni, a szótárírásban sokat segítenek a sajtóban megjelent cikkek. Az idézett forrásokat rendszerre innen kerül be a legtöbb pél-

Szisziüposz külföldön

A nagyszótár elkészítése minden nyelv esetében gigászi feladat. Csak néhány példa: a Grimm testvérek 1854-ben fogtak hozzá a német szavak feltérképezéséhez, de csak 1954-ben fejeződött be a munka, 18 kötettel a végén. Az Oxford-szótár első kiadásához 44 év kellett (1884–1928), és tíz kötet, a második kiadás ehhez még

hozzátett két további kötetet és négy pótkötetet. A megújított, online verzióknak '93-ban fogtak neki 34 millió fontos költségvetéssel. A franciák 23 év alatt készültek el a 16 kötetükkel, ennek érdekében külön költségvetésű intézetet alapítottak. S hogy egy szomszédos példát is hozzunk: a románok 104 év alatt hozták tető alá a maguk szerény 37 kötetét. Az újranyomást nemrég kezdték el 18 egységes kötetben, a vállalkozást a Román Nemzeti Bank támogatja.

damondat a nagyszótárba. A sajtó ugyanis friss, tele van köznyelvi fordulatokkal, ugyanakkor irodalmi, tájnyelvi és szaknyelvi szavakat is tartalmaz, ezért megkerülhetetlen a szótár megírásában. A mostani gyűjtés mintegy tizszázalékos bővítést jelentene, és év végéig szeretnék lezárni. Persze a folyamat itt sem áll meg végleg. Jelenleg 27,5 millió szövegszó (mintegy 22 ezer szövegrészlet) alkotja a korpuszt, ám ez eltörpül a több százmillió szövegtárak mellett.

Sokan mondhatnák most, hogy de hát miért nem használják ki az internet adta bő lehetőségeket. A válasz egyszerű: mert megbízha-

atlan. Vagyis visszakereshetetlen. Semmi sem garantálja, hogy a kiválasztott szöveg holnap is fent lesz még, márpedig a forrás fellelhetősége igen fontos szempont. Ezért nem dolgoznak kéziratokkal sem.

Csiklandozó érzés egy saját fejlesztésű szóval bekerülni a magyar nyelv székesegyházába vagy élő múzeumába. Hiszen ez még a szófáragásban kiváló Parti Nagynak, Szentkuthynak és Esterháznak sem sikerülhet csak úgy. A kitalált szó, a forrás pontos megjelenésével csak akkor jöhet számításba a szótár szempontjából, ha elterjed. Nem kell a sarki fűszeresnek is használnia, de sokat segít, ha több helyen is megjelenik nyomtatásban, vagyis nem csupán a költő életműve tartalmazza.

Ám hiába bővült egyre jobban és egyre modernebb formában a szövegtár, így is maradt hiány. Nem is kevés. Ezért is rehabilitálták végül a korábban lesajnált cédulákat, sok ritka szót ugyanis csak ezeken találtak meg. A nyolcvanas évektől Villó Ildikó kezdte el próbaképpen megírni az első címszavakat, máris szembesülve a szerkesztők talán legnagyobb feladatával: a válogatással és mérlegeléssel. A legújabb szavakat nem számítva ugyanis mintegy 600

Jövőre jelenhet meg a hatodik kötet, az igen fontos „egy” szavunkkal

FOTÓ: MÓRICZ-SABJÁN SIMON



otodik kötet a c (cé, cz, tz) betű címszavától a „dézsmaszedőig” gyűjtötte össze az elmúlt 230 év szókészletét. A kötet 4122 szócikkben dolgozza föl a szavakat, szókapcsolatokat, használatukat 35 157 példamondattal érzékelteti.

Ha minden jól megy, jövőre jön a hatodik, az igen fontos „egy” szavunkkal. Optimális esetben még húsz vagy harminc év kellene a végéig (húsz fölött lenne a megjelent kötetek száma, hiszen csak a kezdőbetűs szavainknak három kötet kell), ám Ittész Nóra erre is csak megengedően mosolyog: igazából több évtized kell egy efféle munka befejezéséhez. Mármint a printverzió elkészüléséhez, mert a folyamatos frissítések és kiegészítések miatt a munka sosem fejeződik be. De hát ki tudná begyömöszölni az élő és folyton megújuló nyelvet egy szűk garzonba?

– Ha büszkén akarunk fogalmazni, akkor azt is mondhatjuk, hogy lassan már kétszáz éves a mi nagyszótárunk – mondja a főszerkesztő, és nem is túloz nagyon. A Magyar Tudós Társaság ugyanis még 1831-ben hagyta jóvá a nagyszótár első terveit, ám a tényleges szótárírói munka csak a XXI. század elején kezdődött el: a sorozat első kötetei 2006-ban jelentek meg. Mielőtt bárki patópálsággal vádolná a tudósokat, tudnia kell, hogy a nagyszótár nem az a műfaj, amelynek már az elhatározás másnapján neki lehet kezdeni. Előbb ugyanis adatbázisként tárolt szöveggyűjteményre van szükség, nem is akármilyen méretűre: más nyelvek esetében nem ritka, hogy több százmillió szövegszóból merítenek a szótárírók.

Mi ezt sem kapkodtuk el. A Magyar történeti szövegtár (vagy Magyar történeti korpusz) kialakítása csak az 1980-as évek közepén indult el. Ennél korábban kezdték el viszont cetlikén gyűjteni a szavakat amolyan társadalmi kezdeményezésként: vidéki értelmiségiek, papok, tanárok jegyezték fel kedvelt szövegrészleteket vagy mondatokat, amelyekben különös kifejezések szerepeltek. Ittész meg is mutat egyet a rengeteg cetliből. Valaki 1842-ben figyelt fel az „egynéha” kifejezésre Erdélyi János szóhasználatában. Fel is jegyezte: „Mint nem vagyok az, ki / Egynéha valék / Anyám, ki felápolt, / Sírból, ha kielne, / Ismerne-e még?”

advány. De a Károli Biblia részleteit is elektronizálták, mint ahogy Mi-

kek. Az idézett forrásokat rendszerre innen kerül be a legtöbb pél-

internet adta bő lehetőségeket. A válasz egyszerű: mert megbízha-

lással a szöveg szempontjából, na elterjed. Nem kell a sarki fűszeresnek is használnia, de sokat segít, ha több helyen is megjelenik nyomtatásban, vagyis nem csupán a költő életműve tartalmazza.

Ám hiába bővült egyre jobban és egyre modernebb formában a szövegtár, így is maradt hiány. Nem is kevés. Ezért is rehabilitálták végül a korábban lesajnált cédulákat, sok ritka szót ugyanis csak ezeken találtak meg. A nyolcvanas évektől Villó Ildikó kezdte el próbaképpen megírni az első címszavakat, máris szembesülve a szerkesztők talán legnagyobb feladatával: a válogatással és mérlegeléssel. A legújabb szavakat nem számítva ugyanis mintegy 600 ezerből kell kiválasztani azok fontossága és gyakorisága alapján a végül nyomtatásban is megjelenő 110 ezer szócikket. (A szűkítés problémájával más nyelvek is szembesültek, ám a kiadhatóság és kezelhetőség miatt kell a gondos válogatás.) Csak egy példát említsünk. Vajon szerepeljen az összes számnév? De miért lenne olyan fontos a 74, mint mondjuk a 48 vagy az 56? Végül úgy döntöttek, hogy a huszonnégy órás nap miatt 24-ig emelik be a számneveket. (Bár, teszi hozzá Ittész, 13 fölött kicsit másként viselkednek ezek a számnevek. Ritkán mondjuk például azt, hogy „tizenkilenc-re ma várom a Nemzetinél”). Esetleg szóba kerülhet az 1001, hiszen ezt a sok szinonimájaként használjuk, de a 101 hiába Orwell szobája és a kiskutyák, már kevesebb eséllyel indul. (Legfeljebb a „számos” jelentése okán kerülhet be ő is a surranópályán.)

Logikus felvetés lenne, hogy miért erőltessük mindenáron a printverziót, miért nem tároljuk kizárólag elektronikusan a szócikkeket, hiszen az e-szótár a folyamatos javítást, bővítést és kiegészítést is könnyedén lehetővé teszi. Természetesen tervezik, hogy rövid időn belül az online hozzáférés is zökkenőmentes legyen.

– De ki tudja megmondani, meddig él a mai tárolási forma? – veti fel Ittész. Az elmúlt húsz évben nagy utat tettünk meg a flopitól a pendrive-ig. Gondoljunk csak bele, milyen időigényes munka lenne újra és újra a megfelelő hordozóra konvertálni a szótár anyagát. A könyv viszont már pár száz éve méltóságteljesen őrzi a benne rögzített tudást.

Akár a bányá mélyén vagy az olajfúró toronyban is.



Jövőre jelenhet meg a hatodik kötet, az igen fontos „egy” szavunkkal

FOTÓ: MÓRICZ-SABJÁN SIMON