

## *M. Pintér Tibor – P. Márkus Katalin*

### **A Termini magyar–magyar szótár és adatbázis szerkesztési elvei és lexikális elemei**

A Termini Magyar Nyelvi Kutatóhálózatba tömörült kutatóállomásoknak kezdettől fogva egyik kiemelt kutatási témája volt a kisebbségi magyar nyelvváltozatok sajátos szókészletének gyűjtése, vizsgálata, elemzése. Ennek eredményeként a kutatócsoport 2007-ben létrehozta és azóta folyamatosan fejleszti a *Termini magyar–magyar szótár és adatbázist* ([termini.nytud.hu/htonline](http://termini.nytud.hu/htonline)), amely nemcsak pillanatképet mutat a határon túli magyar szókészlet kölcsönelemeiről, hanem időbeli (akár a szinkrón diakroniát érintve) és térbeli vizsgálatokat is lehetővé tesz. Lexikográfiai adattárunkban azokat a magyarországiaktól eltérő, idegen eredetű és belső keletkezésű szavakat és szójelentéseket, valamint állandósult szókapcsolatokat igyekszünk összegyűjteni, valamint régiónkénti megjelölésekkel szótározni, képi és (autentikus) hanganyaggal illusztrálni, amelyeket a Kárpát-medencében élő (elősorban nem magyarországi) magyarok használnak. A szótár így a Magyarországon és a környező hét ország magyarlakta régióiban (Erdély, Felvidék, Vajdaság, Kárpátalja, Horvátország, Muravidék, Órvidék) használt sajátos szavakat, közvetlen és hibrid kölcsönszavakat, jelentésbeli és stílusbeli kölcsönszavakat, tükörszavakat és -kifejezéseket, továbbá idegen mintától független belső keletkezésű szavakat, szókapcsolatokat és az egykori közmagyar szókincs megőrzött régiségeit tartalmazza.

A szótár és adatbázis „életciklusa,” fejlődése immáron többrétű (nemcsak lexikológiai, lexicográfiai, hanem informatikai természetű is; l. pl. Benő és mtsai., 2020, 2021). A szótár online megjelenése utáni egyik fontos mérföldkő 2015 volt, amikor lehetővé vált a szócikkek illusztratív képekkel való kiegészítése, amelynek köszönhetően a szótár eddig több száz képpel gazdagodott, valamint 2019, amikor pedig hanganyagok feltöltésének technikai feltételei is megvalósultak, így néhány példamondat meg is hallgatható (ezek élőnyelvi kutatások hanganyagai).

A szótár és adatbázis anyaga, tartalma évente bővül, javul, valamint a szóanyagának lehetősége van követni a szinkroniában bekövetkezett változásokat (ezek egy-egy szócikk történetében, változásaiban, a szerkesztés folyamatában visszakövethetők). A szótár fejlesztése mögötti tudományos munka folyamatosan merít a gyakorlati fejlesztésekből, a szótárról számos (magyar és idegen nyelvű) tanulmány készül, az utóbbi években részletesebbé vált a kölcsönszavak etimológiájának kidolgozottsága, valamint azok lexikológiai, tipológiai minősítése. Ugyanakkor a tudományos ihletettséggű lexicográfia „hátszelében” elkészült egy árnyalt lexicológiai-lexikográfiai kategorizáció a dialektális, regiszterbeli, stílusváltozati, időbeliségi, érzelmi és gyakorisági dimenziói mentén is. A szótár informatikai megoldásai a maga nemében sokáig nívóként voltak jelen a hazai lexicográfiában, illetve fontos, hogy a fejlesztések terén lehetőség van a folyamatos munkára.

A szótár kizárólag online környezetben jelenik meg, ezért a szótár egyik szócikke sem tekinthető lezártnak, illetve véglegesnek, a benne található adatok bármikor bővíthetők, javíthatók. A digitális megjelenés másik nagy előnye, hogy a szerkesztők számára nem léteznek terjedelmi korlátok. Ennek köszönhetően példamondatok és képek is nagy számban állnak az olvasó rendelkezésére, valamint a szócikkek számára vonatkozóan sincs megkötés – minden, a szerkesztők által szótározni kívánt szó bekerülhet a szótárba. Mindezekon túl az olvasó szempontjából az online megjelenésnél fontos lehet a felület alakíthatósága, testreszabhatósága is, így a keresésnél felmerülhet az igény, hogy csak olyan információ jelenjék meg a képernyőn,

amelyet a felhasználó látni szeretne (például csak az ekvivalensek, csak bizonyos régió adatai, illetve akár több szótár találatainak együttes listázása). Ezen és hasonló igények alakítják a szótár felületét is, melyben elsődleges szempont az adatok egyszerű áttekinthetősége, valamint a megjeleníteni kívánt információk személyre szabhatósága. Az adatbázis-alapú szótárak nagy előnye, hogy olyan lehetőségeket is kínálnak, amelyeket a hagyományos papírszótárak nem, vagy csak nehézkesen; például lehetséges a benne lévő összes adat egy időben történő tanulmányozása, akár statisztikák készítése (ezekről pl. M. Pintér 2017a, 2017b).

A 2024. május 26-i állapot szerint a szótár 5431 szócikket tartalmazott. A valamivel kevesebb mint öt és félezer szócikk miatt a szótár legfeljebb kisszótár lehetne, azonban a rendkívül részletes szócikkszerkezet miatt a szócikkekben található adatok a kisszótárakhoz képest jóval több nyelvészeti és laikus felhasználói lehetőséget biztosítanak. A szótár a magyar nyelv Kárpát-medencei kontaktusváltozatainak sajátos lexikális elemeit tartalmazza. A határon túli nyelvváltozatok standard és szubstandard elemeire fókuszál (azaz a határon túli nyelvváltozat egészére, nem pedig a határon túli régiók földrajzi nyelvjárásaira jellemző lexikára), amelyek között megvan a kölcsönös érthetőség. Ennek megfelelően a szótár anyaga az alábbi régiókat fedi le: szlovákiai magyar nyelvváltozatok (a szótárban Fv kóddal jelölve), erdélyi magyar nyelvváltozatok (Er), kárpátaljai magyar nyelvváltozatok (Ka), muravidéki magyar nyelvváltozatok (Mv), drávaszögi magyar nyelvváltozatok (Hv), örvidéki magyar nyelvváltozatok (Öv), vajdasági magyar nyelvváltozatok (Va); magyarországi magyar standard (Hu), az összes régiót egységesen lefedő Kárpát-medencei nyelvváltozat (Km).

A szótár elsősorban az élő nyelvet kutatóknak készül. A szociolingvisztikai cél mellett azonban gyakorlati hasznosíthatóságát is fontos hangsúlyozni: a szótár szerkesztői ugyanis szeretnék lehetővé tenni más felhasználóknak is, hogy megismerjék a magyar nyelvváltozatok sajátos szavait, jelentéseit (Lanstyák és mtsai. 2010: 44). A szóanyag lexikológiai és lexikográfiai feldolgozása (a szócikkek szerkezeti felépítése, a szócikkek közti kapcsolatok kidolgozottsága) lehetővé teszi, hogy a szótárt többféle felfogásban lássuk: ilyen „használati rétegek” például a következő, szótárhoz kapcsolható koncepciók: mai magyar nyelv értelmező szótára, általános szótár, idegen szavak szótára, illetve adatbázis-alapúság. A szótár által feldolgozott szóanyag elsősorban a szinkrón nyelvhasználatot tükrözi, így neologizmusokat is tartalmaz. Ezenkívül a szótárban megtalálhatóak a két világháború közti időszak és a szocializmus tipikus szavai is. A szótár szócikkállományát tekintve fő szempont, hogy a szótárba bekerülő szónak, szókapcsolatnak vagy jelentésnek a határon túli magyar nyelvváltozatok egy nagyobb földrajzi területen elterjedt egységnek kell lennie, nem korlátozódhat kisebb földrajzi vagy bármely stilisztikai, esetleg regiszterbeli változathoz. Mivel a szótárba kerülő szavak, jelentések tipikusan valamelyik államnyelv sajátos egységei, amelyek azonban az adott régió magyar nyelvhasználatának mára szerves részei, ezért a szótár felfogható egyfajta idegen szavak szótárának is. A címszavak között olyanok is fellelhetők, amelyek Magyarországon is használatosak, azonban sajátos, az adott kisebbségi nyelvváltozatra jellemző, a magyarországitól eltérő jelentésben. A szótár naprakészségét az adja, hogy adatbázisként működik: a használók lekérdezéseken keresztül férnek hozzá a keresett szóhoz, jelentéshez. Az adatbázis-jellegnek több olyan előnye is van, mely biztosítja a könnyebb szerkeszthetőséget és az egyszerűbb, igényre szabott megjeleníthetőséget. Ez lehet a kulcsa a szótár folyamatos és sikeres működésének.

Hol is helyezkedik el a magyar értelmező szótárak között? A magyar nyelv értelmező kéziszótárai alapján véve az 1959–1962 között megjelent hétkötetes értelmező szótár (ÉrtSz.) mikrostruktúráját, illetve a szócikkek jelentésszerkezetét alkalmazzák, így például a Magyar értelmező kéziszótár 1972-es és 2003-as kiadása (ÉKsz.<sup>1</sup>, ÉKsz.<sup>2</sup>) is támaszkodott rá. A magyar nyelv értelmező kéziszótárai közül a 2003-ban megjelent, Pusztai Ferenc által szerkesztett szótár volt az első olyan szótár, amely a határon túli magyar nyelv változatainak elemeit szótározta. A szótár ezen sajátosságának köszönhetően a határon túli magyar nyelvváltozatok

standardizált volta szimbolikusan elismertté vált, ma már nem is kérdéses. A magyar értelmező kéziszótár **határon túli** magyar nyelvváltozatokat felölelő anyagának előkészítése céljából létrejött egy ún. „ht-lista” (akkor még csak Word-dokumentumban), amely a mai Termini-szótár elődjének tekinthető. A „ht-lista” szócikkeinek mikrostruktúrája az akkor szerkesztés alatt álló ÉKsz.<sup>2</sup> szócikkeinek szerkezetéhez idomult, mivel a lista célja az volt, hogy magyarországi kiadású, a magyar nyelv standard változatát bemutató szótárak anyagát határon túli lexika bevonásával kiegészítse. Az ÉKsz.<sup>2</sup> után a Tinta Kiadó által megjelentetett Értelmező szótár+ (ÉrtSz.+) lett a következő értelmező szótár, amely szavakat és jelentéseket emelt be a határon túli magyar szókészletből.

A szótár több tekintetben eltér a hagyományos értelemben vett szótárszerkesztési megoldásoktól (például a mikrostruktúra részletessége, a több régióból történő szinkron szerkesztés, a webes megjelenés és szerkesztés, az adatbázis-alapúság). A szótár adatbázisa lexikográfiai szempontból rendkívül gazdag szerkezetű, igyekszik minél pontosabb és részletesebb képet adni a címszó lexikális és konceptuális tartalmáról. A szócikkek mikrostruktúrája három – tipográfiaiailag is jól elkülönülő – részből áll: szócikkfej, szócikktörzs, szócikkkláb. A szócikkfej elemei két kategóriába sorolhatók: azok az elemek, amelyeket a szótár minden egyes szócikke tartalmaz (ezek a kötelező elemek: címszó, szófaj, grammatikai kategóriák, fontosabb nyelvtani alakok), illetve azok az elemek, amelyek csak bizonyos típusú szócikkeknél szerepelnek (kizárólag ott, ahol azok feltüntetése nyelvészetileg indokolt: ezek a fakultatív elemek, például az írásváltozat vagy a kiejtés). A szócikkek összetettségét a szócikktörzs információgazdagsága bizonyítja. Itt is két egymástól elkülönülő információcsoportot találhatunk: a címszó jelentését leíró információkat, valamint a használatot bemutató, illetve a jelentést pontosító példamondatokat. A címszó értelmezését segítik továbbá a fogalomköri minősítések, a régiós / földrajzi nyelvváltozat megnevezései, illetve a stílusminősítések. Egynyelvű, értelmező szótárként az ekvivalensi szerepben álló értelmezés háromféle lehet: szinonimikus, kifejtő vagy körülíró. A szótár példamondatai nemcsak a jelentések minél pontosabb bemutatásának eszközei, mivel a lehető legtöbb példamondaton keresztül pontosítható a jelentés és jelentésárnyalat, valamint a példamondatok különböző típusain keresztül bemutatatható a címszó stilisztikai, nyelvhasználati éréke. A digitális szótárak egyik nagy előnye, hogy a szerkesztőknek nem kell tekintettel lenniük a feldolgozott (szó)anyag mennyiségére. A szótár példamondatainak részletessége, illetve a lehetőség arra, hogy ilyen gazdag tipológiája is megjeleníthető, bizonyítja az online szótárszerkesztésben, illetve megjelenítésben rejlő lehetőségeket. Mivel élő nyelvet feldolgozó, elsősorban szociolingvisztikai kutatásokra használt szótárról van szó, a példamondatok elsődleges felosztása azok hitelességén alapul: a szótárszerkesztők igyekeznek valós, elhangzott vagy leírt példamondatokat használni, illetve kerülnek a szerkesztők általi mondatírást. Ennek tekintetében a szótár példamondatai három nagy típusra oszlanak: (1) hiteles, írott nyelvi példamondat (azaz írott változatban adathozható mondatok), (2) hiteles, beszélt nyelvi példamondat (azaz a beszélt nyelvből adathozható mondatok) és (3) nem hiteles példamondat (melyek forrása nem tisztázható vagy megbízható). A szótár példamondatai nagy többségben hiteles írott és laikus beszélők által konstruált elemek. A példamondatok jelöléseinek egyik további érdekessége, hogy a szerkesztők a beszélő régióbeli hovatartozását, valamint a hiteles példamondatok esetében a forrást is feltüntetik. A szócikk végén található szócikkkláb tartalmazza a címszó etimológiai információit, valamint ugyanahhoz a szócsaláddhoz tartozó, a szótárban megtalálható egyéb szavakra történő utalásokat. A szótár nem titkolt célja a magyar nyelvet érintő lexikológiai kutatások alapanyagának gazdagítása határon túli magyar szókészleti anyaggal, valamint – és ezáltal – Magyarország és a külső régiók szókincsbeli eltéréseinek megismertetése, a lexikai kapcsolatok dokumentálása; emellett kiemelt cél a magyar etimológiai, kontaktológiai és kétnyelvűségi kutatások új dimenziójának megteremtése, illetve a magyar lexikográfiát érintő technológiai újítások bevezetése, meghonosítása (gondolunk itt a hang- és képanyagra).

A szótár elsődleges célja a magyar nyelv határon túli változatainak monitorozása, ugyanakkor az újabb fejlesztések a speciálisan magyarországi szavak és jelentések (ha úgy tetszik, modern reáliák) szótárazására, valamint a Kárpát-medence minden régiójában azonos jelenésben használatos szavak szótárazására is kiterjednek. A magyarországi jelentések, szavak figyelése, folyamatos beépítése éppen Nádor Orsolya munkájának eredménye, mivel a magyarországi vonatkozású elemek kutatását, valamint lektorálását ő végzi.

Miért is lehet fontos a nyelvészet számára? Elsősorban a szociolingvisztikai hozadék miatt, de természetesen a szócikkek mikroszerkezetébe kódolt információk miatt is: így megemlíthetők a stilisztikai, dialektológiai, etimológiai és kontaktológiai információk is. De érdekes kutatási témát biztosít a reáliák kutatóinak is; ezek illusztrálása ma már nem csak lexikális, hanem képi elemekkel is megvalósul. A szótár elsősorban nyelvészek számára jelent érdekes kutatási terepet. Ezt bizonyítja a felhasználói sajátossága is: a szótár nyilvános felületét bárki használhatja, kereshet is benne, azonban a sokrétű adathálót érintő különféle lekérdezések és statisztikák jelenleg csupán a kutatók számára elérhetőek. Ugyanakkor az sem elhanyagolható tényező, hogy az érdeklődő (ha úgy tetszik, laikus) felhasználó tájékozódhat valamely régió speciális szókészletéről, megismerheti a saját régióját érintő szókincs nyelvészeti, szociolingvisztikai tulajdonságait (például a használat fogalomköri, regionális vagy stílusbeli – ezen belül regionális, regiszterbeli, stílusváltozati hovatartozásra utaló, időbeliségre vonatkozó, érzelmi viszonyulásra utaló vagy a relatív gyakoriságra vonatkozó – információkat). Az explicit információk mellett a példamondatok rendkívül gazdag implicit információt tartalmaznak. A szótár fontos elve, hogy nem korlátozza a példamondatok számát, így egy-egy címszóhoz tartozó példamondat többféle implicit (a címszó és a szócikk szempontjából nem címkézett) nyelvi sajátosságot rejt a beszélt nyelv használatáról. A példamondatok nemcsak a címszó denotatív jelentéseinek és jelentésárnyalatainak pontosabb bemutatását szolgálják, hanem – főleg ott, ahol nagyobb számú példamondat található – segítségükkel viszonylag megbízhatóan megítélhető a címszó stilisztikai, nyelvhasználati értéke is.

Érdekes nyelvi adat a jelentésekhez kapcsolódó képi illusztráció, ahol általában kontextusba ágyazva jelennek meg a címszót vagy annak valamely jelentését illusztráló vizuális tartalmak.

A szótár immár lassan 20 éve működik, szerkezetileg, szótárszerkesztési elveit tekintve szinte változatlanul. Elérkezett az az idő, amikor érdemes lesz főként technikai újításokat bevezetni; a konceptuális újításoknak nyilván határt szab a már meglévő lexikális feldolgozottság, hiszen minden új, a szócikk szerkezetét érintő módosítás a már meglévő szóanyag szerkezetét is befolyásolja, befolyásolná. A változások így elsősorban a felhasználói és szerkesztői felületet érintik majd (jó lenne egy korszerű XML-technológiával készített XML-szerkezet is), illetve a több régióból történő szerkesztés miatt megjelentek következtelenségek: bizonyos részeken szükség mutatkozik a hibajavításra, valamint egységesítő fejlesztésekre (ezek leginkább a szócikkfejlben és szócikkkláiban található információkra vonatkoznak, mivel azok karbantartását lehet legegyszerűbben elvégezni). A szótár és a feladatok folyamatosan változnak, igazodva az aktuális nyelvészeti és technológiai igényekhez. Reméljük, Orsi munkájával továbbra is támogatja a szótár fejlesztését.

## Irodalom

- Benő Attila – Lanstyák István – Juhász Tihamér 2020. A Termini „határtalan” szótára. *Magyar Tudomány* 2020/2: 153–163.
- Benő Attila – Lanstyák István – P. Márkus Katalin – M. Pintér Tibor 2021. A Termini magyar–magyar szótár és adatbázis szerkezeti és szerkesztési sajátosságai. *Magyar Nyelvőr* 145: 378–389.
- Lanstyák István – Benő Attila – Juhász Tihamér 2010. A Termini magyar–magyar szótár és adatbázis. *Régió* 2010/4: 37–58.
- M. Pintér Tibor 2017a. Számszerűsített kétnyelvűség. Adatbányászati technikák haszná a nyelvészetben. In: Misad Katalin (szerk.): *Kétnyelvűség térben és időben*. Szenczi Molnár Albert Egyesület, Pozsony. 125–157.

M. Pintér Tibor 2017b. Adatbányászat a kétnyelvűség mögött. In: Benő Attila – Gúti Erika – Juhász Dezső – Sztókai Szilvia – Terbe Erika – Trócsányi András (szerk.): *Tudományköziség és magyarságtudomány a nyelvi dimenziók tükrében* (A VIII. Nemzetközi Hungarológiai Kongresszus (Pécs, 2016. augusztus 22–27.) három szimpóziumának előadásai). Termini Egyesület, Törökbálint. 158–167.

### Szótárak

ÉKsz.<sup>1</sup> = *Magyar értelmező kéziszótár*. 1972. Juhász József – Szőke István – O. Nagy Gábor – Kovalovszky Miklós (szerk.). Akadémiai Kiadó, Budapest.

ÉKsz.<sup>2</sup> = *Magyar értelmező kéziszótár*. 2003. Pusztai Ferenc (főszerk.). 2., átdolgozott kiadás. Akadémiai Kiadó, Budapest.

ÉrtSz. = *A magyar nyelv értelmező szótára 1–7*. 1959–1962. Bárczi Géza – Országh László (főszerk.). Akadémiai Kiadó, Budapest.

ÉrtSz.+ = *Értelmező szótár +. Értelmezések, példamondatok, szinonimák, ellentétek, szólások, közmondások, etimológiák, nyelvhasználati tanácsok és fogalomkörü csoportok 1–2*. 2007. Eőry Vilma (főszerk.). Tinta Könyvkiadó, Budapest.

*M. Pintér, Tibor – P. Márkus, Katalin*

### **Principles of the editorial process and lexical data of the *Termini Online Hungarian–Hungarian Dictionary and Database***

The *Termini Online Hungarian–Hungarian Dictionary and Database* describes the lexicon of the Hungarian language as it is spoken in the countries surrounding Hungary. It is considered to be a general dictionary of present-day Hungarian. Each entry contains authentic sample sentences to illustrate the use of the headword, making it possible to examine the special use of a word or construction in a grammatical and pragmatic context. The lexicographical database is edited online in eight countries. The study outlines some main tasks of lexicography management regarding the dictionary as well as the life cycle of the editing process. The fact that editing is done online makes it possible for the dictionary to expand, as a result of activity in eight countries. In the study peculiarities of the editing process and of the dictionary are highlighted, concerning the following topics: dictionary structure, database character and the use of the dictionary for linguistic research purposes.