

A TERMINI „HATÁRTALAN” SZÓTÁRA¹

THE ‘BORDERLESS’ DICTIONARY OF THE TERMINI RESEARCH NETWORK

Benő Attila¹, Juhász Tihamér², Lanstyák István³

¹egyetemi tanár, Babeş-Bolyai Tudományegyetem Kolozsvár, Románia, Szabó T. Attila Nyelvi Intézet, Kolozsvár, Románia
attilabe@yahoo.com

²informatikus, nyelvész, JH Media Group, Atlanta, GA, USA, Szabó T. Attila Nyelvi Intézet, Kolozsvár, Románia
balamber@gmail.com

³egyetemi tanár, Comenius Egyetem, Pozsony, Szlovákia, Gramma Nyelvi Iroda, Somorja, Szlovákia
istvan.lanstyak@uniba.sk

ÖSSZEFOGLALÁS

Tanulmányunk a Magyarországgal szomszédos államokban beszélt magyar nyelvváltozatok sajátosságait regisztráló *Termini magyar–magyar szótár és adatbázis* szerkesztési elveiről és hasznosíthatóságáról szól. A hét országban szerkesztett, elektronikus, internetes lexicográfiai adatbázis lehetővé teszi, hogy a mai magyar nyelvet érő, több nyelvből származó, közvetlen és közvetett kontaktushatásokat, megőrzött lexikai régiségeket kontextusban vizsgálhassuk. Mivel a szócikkek információban gazdag ismeretanyagot jelenítenek meg a lexikai egységek stílusértékéről, denotatív jelentéséről, használati köréről és etimológiai vonatkozásairól, az adatbázis jól hasznosítható a szociolingvisztikai, lexikológiai, leíró nyelvészeti és stilisztikai kutatásokban is.

ABSTRACT

The article presents the principles of the editorial and utilisation possibilities of *Termini Online Dictionary and Database*, which registers the characteristics of the varieties of Hungarian spoken in the states neighbouring Hungary. The lexicographical database is edited online in 7 countries, and it makes possible to study direct and indirect linguistic lexical influences, as well as lexical archaisms in context. As the lexicographical items of the dictionary contain abundant information about the stylistic value, referential meaning, usage and etymology, the database can also be used efficiently in sociolinguistic, lexical, descriptive linguistic, and stylistic studies.

¹ Kutatásainkat az MTA támogatja határon túli ösztöndíjprogramja révén, részben közvetlenül, részben a Szabó T. Attila Intézet, illetve a Fórum Kisebbségkutató Intézet keretében működő Gramma Nyelvi Iroda közvetítésével. Lanstyák Istvánnak az írásunk témájába vágó kutatásait ezen kívül a szlovák Kutatás- és Fejlesztéstámogató Ügynökség is támogatta és támogatja: korábban az APVV-0689-12 sz., jelenleg az APVV-17-0254 sz. szerződés alapján.

Kulcsszavak: internetes szótár, lexikográfiai adatbázis, magyar nyelvváltozatok, nyelvi kontaktusok, stílus, etimológia, statisztika

Keywords: online dictionary, lexicographical database, Hungarian language variants, language contacts, style, etymology, statistics

1. BEVEZETÉS

Írásunkban a *Termini magyar–magyar szótár és adatbázis* sajátosságait, szerkesztési elveit, elektronikus jellegéből fakadó előnyeit ismertetjük. A szótár eddigi általánosabb jellegű bemutatásaihoz képest (Lanstyák et al., 2010; Benő, 2017; Sztók, 2017) nagyobb hangsúlyt fektetünk a szótár sajátosságainak leírására, adatbázis jellegének kiemelésére és a különböző kutatásokban való hasznosíthatóságára.

A magyar nyelv Trianon utáni szétfejlődésével kapcsolatos munkálatok **határtalanítás** néven ismeretesek. Ezen eredetileg azokat a nyelvészeti – lexikológiai, lexikográfiai és korpusznyelvészeti – tevékenységeket értettük, melyek célja az, hogy az újonnan készülő vagy átdolgozott magyarországi nyelvészeti kiadványokban jelentőségüknek megfelelően jelenjenek meg a külső régiók magyar nyelvének változatai, azaz a címük szerint a „magyar” nyelvet vagy annak valamely részrendszerét bemutató kiadványok és számítógépes programok (egy- és kétnyelvű szótárak, nyelvtanok, helyesírási, nyelvhasználati, stilisztikai kézikönyvek, helyesírás-ellenőrző számítógépes programok stb.) ne kizárólag a Mo-i² magyar nyelvvel foglalkozzanak, hanem „a” magyar nyelvvel, ahogy azt címükben hirdetik. Később a „határtalanítás” műszó használata kiterjedt a magyar nyelv szétfejlődésével kapcsolatos nyelvalakítási munkálatokra is (például a külső régiók sajátos szóképzleti elemeinek kodifikálására).

A határtalanítás egyik legfontosabb területe a **szóhatártalanítás**; ennek eszköze az írásunk témájául szolgáló Termini-szótár és -adatbázis.³

² Írásunkban az egyes Kárpát-medencei régiók rövidítései és azok sorrendje alapvetően a Termini-adatbázis gyakorlatát követi: Er: Erdély, Fv: Felvidék, Hv: Horvátország, Ka: Kárpátalja, Km: Kárpát-medence vagy közmagyar, Mo: Magyarország, Mv: Muravidék, Óv: Órvidék, Va: Vajdaság.

³ Dolgozatunkban a rövidség kedvéért olykor egyszerűen *adatbázis*ról beszélünk; ebbe mindig beleértjük a szótárt is.

2. A SZÓTÁR ANYAGA

Az adatbázis a külső régiók sajátos lexikai egységeit (szavait, kifejezéseit, állandósult szókapcsolatait), továbbá köznyelvi szavak régióspecifikus jelentéseit és stílusértékeit szótározza.⁴ A sajátos szavak, kifejezések egyrészt a két- vagy többnyelvűségi helyzetből adódó kontaktusjelenségek, másrészt megőrzött régiségek (olyan szavak és kifejezések, amelyek a magyarországi köznyelvből visszaszorultak, de a külső régiókban megmaradtak), ezenkívül akad köztük néhány önálló, a kontaktushatástól többé-kevésbé független új alkotás is.

2.1. A más nyelvekből származó lexikográfiai egységek közvetlen és közvetett kölcsönelemeket jelölnek. A **közvetlen átvételek** olyan kölcsönszavak és -kifejezések, amelyek új lexémaértékű hangsorokkal gyarapítják egy adott régió magyar nyelvváltozatát, és használatuk – akárcsak minden más, egy-egy közösség nyelvi normájának megfelelő elemé – funkcionálisan indokolt, mivel nem ritkán állami és kulturális sajátosságokat jelölnek, amint ezt az alábbi adatok szemléltetik: Er *prefektus* 'a kormányt képviselő) megyefőnök', Fv *bakalár* 'felsőfokú alapképzést elvégzett és alapfokozatot szerzett személy', Hv *bozsicsnica* 'karácsonyi ajándék-pénz', Ka *cseburek* 'darált hússal töltött, kétrét hajtott, olajban kisütött lepény', Mv *borovnicsevec* 'áfonyapálinka', Öv *császárfröccs* 'almdudler és (fehér) bor keverékéből készült ital, fröccs Ausztriában', Va *babura* 'tölteni való húsos paprika'.

A **közvetett nyelvi hatások** jelentésbeli kölcsönszók, tükörszók és tükörszerkezetek kialakulásában nyilvánulnak meg. A jelentéskölcsönzés abban mutatkozik meg, hogy a már korábban meglévő szó az államnyelvi hatás eredményeként egy újabb jelentésben is használatos lesz, vagy pedig jelentése kitágul: Er *régiség* 'szolgálati idő', Fv *akadémia* 'közgazdasági szakközépiskola', Ka, Hv, Mv *dolgozik* 'működik', Öv *katedra* 'templomi szószék', Er, Va *kamion* 'tehergépkocsi'. A tükörszavak, tükörszerkezetek szintén a közvetett nyelvi hatás formáit jelzik, és átadó nyelvi lexikai mintát követve jönnek létre: Fv *alapiskola* 'általános iskola', Er, Fv, Ka *családorvos* 'házi orvos', Er *iskolacsoport* 'iskolaközpont', Er, Fv, Va, Ka, Hv *fizikai személy* 'természetes személy', Er, Fv *anyasági szabadság* 'az anya által igénybe vehető gyereknevelési szabadság', Fv, Va *rekreációs futás* 'kocogás' (Benő, 2013; 2014, 81–87).

2.2. A kisebbségi nyelvváltozatok sajátos lexikai elemei nem szükségszerűen a többségi nyelv közvetett hatásának eredményei. Olykor a mai magyarországi köz-

⁴ A szótár szerkesztői: Benő Attila (Erdély), Lanstyák István (Felvidék), Lehocki-Samardžić Anna (Horvátország), Márku Anita (Kárpátalja), Kolláth Anna (Muravidék), Szoták Szilvia (Őrvidék), Molnár Csikós László (Vajdaság). Műszaki szerkesztő: Juhász Tihamér (Erdély). Régiófüggetlen munkatárs: Kitlei Ibolya (Magyarország).

nyelvtől eltérő szavak, kifejezések megőrzött régiségek is lehetnek. Mi azokat a lexikai elemeket tekintjük megőrzött régiségeknek, amelyeknek korábbi általános elterjedtségét nyelvtörténeti adatok is igazolják, több külső régióban is használatosak, és jelentésük közismert. A szótár anyaga több hasonló megőrzött régiséget adhat: Er, Fv *altiszt* 'tiszthelyettes', Er, Fv, Va *hajtási jogosítvány* 'vezetői jogosítvány', Er, Fv *kihágás* 'szabálysértés', Er, Fv, Ka *születési bizonyítvány* 'születési anyakönyvi kivonat'. A megőrzött régiségnek tekinthető szavak és kifejezések egy része olyan lexikai elem, amely az 1950-es évekig a magyar nyelvterület egészén köznyelvi, normatív jellegű volt (Benő, 2014, 87–91.).

2.3. A lexikográfiai adattár szócikkei három fő szerkezeti egységből állnak: 1. szócikkfej, 2. szócikktörzs, 3. szócikkkláb. A **szócikkfej** olyan információkat jelenít meg, amelyek az egész szócikkre és valamennyi jelentésre érvényesek. Ezek a következők: 1. a címszó szótári alakja, 2. a címszó kiejtése, 3. a címszó írásváltozata(i), 4. szófaji minősítés, 5. alaki felépítés, 6. a kontaktusjelenség típusa, 7. a címszóra vonatkozó bibliográfiai adatok.⁵ A **szócikktörzsben** található az egyes jelentésekre vonatkozó tudnivalók: 1. fogalomkörü minősítések; 2. felségjelzések; 3. stílusminősítések (dialektusbeli, regiszterbeli, stílusváltozati hovatartozás, időbeliség); 4. értelmezések; 5. az adott jelentést szemléltető példamondatok (a példamondatok típusának meghatározásával). A **szócikkklábban** jelenik meg a megjegyzés helye és az adott szóhoz kapcsolódó más szócikkekre való utalás, amely a címszó szótározott alakváltozataira, más régióbeli analóg szavakra vagy a szócsalád más tagjaira vonatkozhat. Ugyancsak a szócikkklábban láthatók a szó eredetére vonatkozó információk. (Ezek az etimológiai vonatkozások valamennyi régióra külön-külön jelennek meg, az indító és a közvetítő nyelv meghatározásával.)

A szótár régióspecifikus keresőjével megállapítható az egyes régiók szócikkeinek és szójelentéseinek a száma. 2017-ben az adattár 1221 erdélyi, 1810 felvidéki, 991 vajdasági, 747 kárpátaljai, 869 horvátországi, 678 muravidéki és 569 örvidéki vonatkozású szócikket tartalmazott. A szójelentések számát és a régiók közötti összefüggést tekintve 2017-ben 5530 jelentés kapcsolódott Felvidékhez, 5396 Erdélyhez, 2446 Kárpátaljához, 2400 Vajdasághoz, 1962 Muravidékhez, 185 Horvátországhoz és 1649 Örvidékhez (M. Pintér, 2017, 162.).

2.4. Az adatbázis elektronikus formában való megjelenése a hagyományos papíralapú szótárhoz képest számos előnnyel jár. Egyrészt a szócikk maximális terjedelmét nem szükséges meghatározni, elvileg bármilyen mennyiségű adattal, kiegészítéssel tovább bővíthető, s nincs korlátozva a példamondatok terjedelme és mennyisége sem: egy-egy szó jelentéseit, a szó kontextuális sajátosságait szinte

⁵ A külső felhasználók számára a szó alaki felépítésére, kontaktológiai jellemzésére és a bibliográfiai adatokra vonatkozó információk nem láthatók.

korlátlan számú példamondattal lehet szemléltetni. Ezek nemcsak egy mondatra korlátozódhatnak, hanem bekezdésnyi terjedelműek is lehetnek, hogy a szó stílusértékét és szerkezeti beágyazódását nagyobb szövegösszefüggésben lehessen tanulmányozni. Elektronikus szótárként adva van a lehetőség az adatbázis összekapcsolására más magyar nyelvű és kétnyelvű internetes szótárakkal, valamint lexikonokkal, ez pedig lehetővé teszi a szavak és szójelentések komplex vonatkozásainak feltárását.

Újabban a szótár **képanyagot** is tartalmaz. Többjelentésű szavak esetében a képek az egyes jelentésekhez kapcsolódnak, nem pedig a szócikk egészéhez. Így egyetlen szócikkben több kép is megjelenhet. A képek elsősorban a tárgyakra vonatkozó főnevek, kultúraspecifikus jelentések (például étel-, ital- vagy ruhafajtákra vonatkozó utalások) pontosabb megértését szolgálják. Jelenleg 805 szócikkben található kép. 2018-tól egy újabb funkció erősíti az adatbázis **multimediális** jellegét: a példamondatokhoz **hanganyag** társítható, ezáltal pedig a későbbiekben a szótározott szavak ejtési sajátosságainak regionális különbségeit is tanulmányozni lehet, bár ez a lehetőség az adatok szintjén a későbbiekben fog megvalósulni, mivel további kutatásokat igényel.

A szótár azt is lehetővé teszi, hogy az „állandó megjegyzés” mezőben részletezzük a szó kulturális, pragmatikai vonatkozásait. Így például az *Er radina* szócikkből megtudhatjuk, hogy ez a megnevezés két jelentésben használatos: 1. gyermekágyas asszony meglátogatása, ez alkalommal vitt ünnepi eledel; 2. keresztelő, keresztelői lakoma. Az állandó megjegyzés mezőből tudomást szerezhetünk arról, hogy a *Magyar katolikus lexikon* szerint a *radina* gyakorlásában az ünnepi eledelről a komasszonyoknak szokás gondoskodni, és hogy keresztelői eseményként két formája is van: egy szűkebb, családi jellegű *kis radina*, és a tágabb körű rokonság részvételével megvalósuló keresztelői lakoma (*nagy radina*).

A lexikográfiai adattár **interaktív, felhasználóbarát** jellegű, ami abban mutatkozik meg, hogy változatos szó- és jelentéskeresési lehetőségeket biztosít; a külső felhasználók továbbá megjegyzéseket fűzhetnek a szócikkekhez, és akár új címszavakat is javasolhatnak. Ugyanakkor a szerkesztők ideiglenes megjegyzéseket tehetnek a korábbi és újabb adatokhoz, folyamatosan látják egymás munkáját, a szócikkek módosításai visszakereshetőek, az új szócikkek létrejöttéről a szerkesztők értesítést kapnak, ami könnyebbé teszi a régióközi összefüggések felismerését.

3. ADATBÁZIS-ÉPÍTÉSI, KUTATÁSSZERVEZÉSI LEHETŐSÉGEK

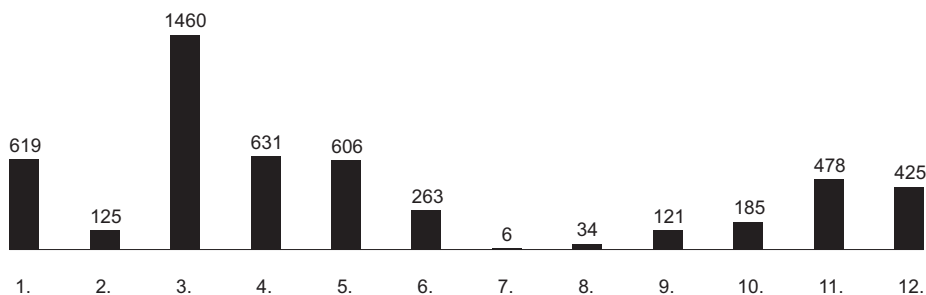
A szótárszerkesztési munkálatokat a különböző földrajzi régiókban tevékenykedő szerkesztő munkatársak már tizenegy éve végzik. A térben és időben ennyire kiterjedt munkálatok összehangolására online felületet alakítottunk ki, amely

szervezi és irányítja a tevékenységeket. A szócikkszerkesztési webes űrlap meghatározza, és így egységesíti a szócikkek szerkezetét és tartalmát. A szótári anyag áttekintését, feldolgozását és a hibák javítását számos online funkció és lekérdezés támogatja.

3.1. A szótárszerkesztési munkálatok támogatása, illetve a szótári anyag feldolgozása céljából a szerkesztők huszonhárom különböző típusú alkorpuszt generálhatnak három fő kategórián belül. Ezek a következők: kutatási, fejlesztési és dokumentációs listák. A kutatási listák különböző összetett tartalmi szempontok szerinti szójegyzékek, például adott fogalomkörü minősítést tartalmazó szócikkek, stílusminősítés és régió szerinti szócikkek, két választható régióban előforduló szócikkek stb. A **fejlesztési cikkek a szótárszerkesztési feladatokat hivatottak segíteni például az adathiányos szócikkek listázásával. Ilyen szójegyzékek az ún. rejtett (azaz a külső felhasználó számára egyelőre nem látható) szócikkek régió szerint, példamondat nélküli jelentéseket tartalmazó szócikkek stb. A **dokumentációs listák** a korábbi munkát hivatottak dokumentálni adminisztratív célból, vagyis nyomon követhetővé és kimutathatóvá teszik az egyes munkatársak munkáját, hozzájárulását az adatbázis építéséhez.**

A szócikkek adattartalma is beállítható: például meghatározhatjuk, hány példamondat jelenjen meg a listán, láthatóak legyenek-e a fényképek, a külső vagy belső megjegyzések, az utolsó módosítás adatai. A szócikkek sorrendje is választható betűrend, időpont vagy szerkesztő szerint.

A **statisztikai kimutatások** számszerűsített adatokat szolgáltatnak a szótári korpuszra, a szerkesztési munkálatokra és a szótár felhasználására vonatkozóan. A látogatók tevékenységét, a szótár internetes jelenlétét három kimutatás dokumentálja. Ha például megnézzük a keresések számának időbeli eloszlását az elmúlt tíz évre vonatkozóan, kiderül, hogy az jellegzetesen kötődik a tanév és a vakációk váltakozásához. Ezt jól példázza az alábbi, 2018-ra vonatkozó kimutatás (lásd az *1. ábrát*). A keresések időbeli eloszlásához kapcsolódó mutató a



1. ábra. Keresések időbeli eloszlása 2018-ban

felhasználók regisztrációjának időbeli eloszlása. Amint az ábra jelzi, júliusban és augusztusban a legalacsonyabb a keresések száma.

A statisztikai kimutatások a szótárszerkesztők munkájának időbeli eloszlását is dokumentálják. Megnézhető, hány szócikket módosítottak a munkatársak egy bizonyos időszakban. A szótári korpusz elemzését három kimutatás segíti: a rendszerből lekérdezhető a példamondatok száma példamondattípusok szerint, a különböző szófajú címszavak száma, a példamondatok eloszlása régiók szerint, a különböző felségjelzésű jelentések száma.

3.2. Külön szójegyzéket is lehet generálni a belső megjegyzésekkel ellátott szócikkekről a gyors munkavégzés érdekében. Ez a gyakorlat szintén a térbeli és időbeli távolság áthidalását szolgálja. Újabban a közösségi média kommentkultusza igazolja vissza ezt az együttműködési szerkesztési gyakorlatot.

Kezdetben a Termini-szótárba csak a közvetlen kölcsönszavakat gyűjtöttük, később más lexikai elemeket, például tükörfefejezéseket is felvettünk. Az ilyen és más szerkesztési munkák kísérleti jellege miatt vezettük be a szócikkek elrejtésének gyakorlatát. A **rejtett jelölésű szócikkek** csak annyiban különböznek más társaiktól, hogy azokat csak a szerkesztők láthatják. Így sikerült kialakítani egy kétlépcsős rendszert, ahol az új szócikkek először csak belső használatra készülnek, s csak bizonyos fokú kidolgozottság elérése után kerülnek nyilvánosságra.

A szótárszerkesztők viszonylag nagy száma és térbeli távolsága miatt a munkálatok során **szerkesztési hibák** szinte elkerülhetetlenül létrejönnek. Gyakori hiba a szócikkek adatainak hiányos bevitele. Ezek kijavítását több automatikus elemzés, illetve ezek alapján generált szójegyzék segíti. Szójegyzékeket lehet generálni hiányzó címszó, szófaji minősítés, felségjelzés, dialektus, regiszter, stílusváltozat, valamint példamondatok régióbeli hovatartozása és típusa szerint. A módosítások mentése előtt a program is figyelmezteti a szerkesztőket a hiányzó adatokra, ugyanakkor lehetőséget ad a mentés végrehajtására.

3.3. Az adattár lehetővé teszi az adatok további kódolását, valamint kimutatások, statisztikák készítését, szólisták generálását. Egy szócikk harminckét különböző típusú adatmezőből épül fel. Egyes adatok a szócikkben csak egyszer fordulnak elő. Ilyenek a szócikkfej adatai, például a címszó, a címszó kiejtése, írásváltozata, szófaji minősítése stb. A szócikktörzs adatai többszörözhetők, például egy szócikknek több jelentése lehet, a jelentésekhez különböző, régióként eltérő minősítéseket lehet megadni, illetve korlátlan számú példamondat kapcsolható hozzájuk. Az összes adattípus közül tizenkét mező zárt értéktartományú. Ezek értékeit egy legördülő menüből kell kiválasztani. Ilyenek például a felségjelzés, dialektus, stílusváltozat, példamondat típusa. Ezzel szemben húsz szabadszöveges mező tetszőleges értéket tárolhat. Az adatokat strukturált adatbázisban helyezzük el,

és így gyakorlatilag tetszőleges statisztikai elemzés hajtható végre rajtuk. A huszonhárom eltérő típusú szójegyzék és a hat statisztikai kimutatás jól példázza a kutatási-feldolgozási lehetőségeket. Emellett a nagyszámú példamondat olyan strukturált élőnyelvi korpuszt képez, amely számos további kutatási lehetőséget kínál a digitális szövegfeldolgozás számára.

4. A SZÓTÁRI ANYAG SAJÁTOSSÁGAI

Az adatbázis építése során a külső régiók sajátos szókincsének változatossága komoly lexikográfiai problémákat vetett föl. Az ezek megoldására való törekvés motiválta többek között a stílusminősítés kérdéseivel való behatóbb foglalkozást. Bár kiindulópontunk a külső régiók szókinccse volt, a fölmerülő problémákat a közmagyar szókészlet egészére való tekintettel tanulmányoztuk, így az a többdimenziós minősítési rendszer, amelyet kialakítottunk, nemcsak a külső régiók szókinccsét tartalmazó szótárakban alkalmazható, hanem bármilyen egynyelvű magyar szótárban, még akkor is, ha a rendszert mindig a konkrét szótár kívánalmaihoz kell igazítani.

4.1. Az adatbázisunkban alkalmazott új rendszerezés kidolgozásában az értelmező kéziszótár minősítéseiből indultunk ki, arra törekedve, hogy azoktól indokolatlanul ne térjünk el. A változtatások többfélék voltak: néhány kéziszótárbeli stílusminősítést átértelmeztünk, az egyiket átneveztük; a kéziszótárban is meglévő minősítések egy részét árnyaltuk; a kéziszótár „vulgáris” minősítését teljesen elvetettük; új minősítéseket vezettünk be. A változtatásokkal a fő célunk az volt, hogy a létrejövő új minősítési rendszer lehetővé tegye a stílusérték különböző típusú forrásainak egymástól való elkülönítését, s ezzel a minősítések többértelműségének felszámolását.

Az új minősítések bevezetésével ezenkívül tehermentesíteni kívántuk a meglévő, hagyományos jelöléseket úgy, hogy ezek használatát leszűkítettük a legjellemzőbb területre. Például a „választékost” csak stílusváltozati minősítésként alkalmazzuk, mégpedig a jellemző módon a „gyönyörködtető szándékú” szövegekben előforduló lexémák jelölésére; az iskolázott beszélőkre jellemző lexémákat nem „választékos”, hanem „iskolázott” minősítéssel látjuk el, az olyan expresszív szavakat pedig, amelyek más kategóriába – például „pejoratív”, „durva”, „kedveskedő” stb. – nem sorolhatók be, „expresszív” vagy „átvitt” minősítéssel.

Az új, többdimenziós minősítési rendszer fő hozadéka az, hogy alkalmazásával sikerült megszüntetni a stílusminősítések jelentős részének többértelműségét. Ez nagy előrelépés a nyelvi anyag feldolgozása szempontjából, mert jelentősen megkönnyíti a szótárírók minősítő munkáját: egy logikus rendszer birtokában könnyebb eldönteni egy-egy szóról, milyen minősítéssel szükséges

ellátni, mint akkor, ha a minősítéseket nem tudjuk a maguk rendszerbeli összefüggéseiben szemlélni. Ezenkívül az anyag sokkal könnyebben hasznosítható a stíluskutatásban.

4.2. Adatbázisunk internetes megjelenési formájából következik, hogy sokkal nagyobb mennyiségben tartalmazhat **példamondatokat**, mint a nyomtatott szótárak. Példamondataink nemcsak a jelentések illusztrálására szolgálnak, hanem a címszó tényleges használatának minél több példán való dokumentálására is. A példamondatok segítségével történik időről időre új jelentések, jelentésárnyalatok azonosítása.

A nagyon gyakori szókészleti egységekre vonatkozó példamondatok összegyűjtésében nem törekedhetünk teljességre; viszont minél ritkább használatú egy-egy szó vagy szókapcsolat, annál inkább igyekszünk létezését és használatát minél több hiteles példával dokumentálni. Azokban az esetekben, amikor nem állnak rendelkezésünkre hiteles írott vagy beszélt nyelvi adatok, saját magunk is alkotunk a szó jelentését szemléltető mondatokat, szövegrészleteket, továbbá felhasználjuk más szójegyzékek, szótárak példáit. A példamondatok konstruált mi-voltát külön jellel jelöljük. A jelölésből az is kiderül, hogy az ilyen példamondatot laikus (azaz nem nyelvész) beszélő vagy nyelvész (rendszerint a szótár valamelyik fejlesztője) alkotta-e.

Fontos azonban megjegyezni, hogy az adatbázisban található példáknak csupán csekély hányada konstruált, a legtöbb példa hiteles, azaz meg tudjuk határozni a lelőhelyét (az URL-cím vagy bibliográfiai adatok megadásával). A példamondat után feltüntetjük a *locus*ra utaló adatot, így a példamondatok visszakereshetők. A locus után közöljük az esetleges egyéb tudnivalókat is, például a példamondat szerzőjére vonatkozó információkat, ha azok valami miatt fontosak.

A hiteles példamondatokban külön jel utal egyfelől arra, hogy írott vagy beszélt nyelvek-e, másfelől arra, hogy szokványos szóhasználatról van-e szó, vagy metanyelvi használatról. Mindezek a kutatás szempontjából alapvetően fontos információk, melyek nélkül az érintett szóelőfordulás sokkal nehezebben értelmezhető. A beszélt nyelvi példákat átiratban közöljük, de amint fentebb jeleztük, adatbázisunk legújabbban lehetővé teszi a hangzó példamondatok beépítését is.

4.3. A külső régiók anyagának egyetlen adatbázisban való tárolása lehetővé teszi, hogy az egyes régiók adatai könnyen összehasonlíthatók legyenek egymással. Az egyetlen adatbázis használata ugyanis mintegy kikényszeríti az azonos minősítési szempontok és formai megoldások alkalmazását. A kutatásvezető egyik legfontosabb feladata, hogy őrködjön az adatbázis egységessége fölött, vagyis figyelmeztesse a kutatókat azokra az esetekre, amikor például az adatok alapján gyanítható, hogy valamely stílusminősítést az egyes régiókban nem teljesen egyformán értelmeznek, vagy mondjuk, az etimológiák közlés módjában kisebb kü-

lönbségek vannak kialakulóban. Magyarországi munkatársunk – amellet, hogy az internetről az összes régió számára példamondatokat gyűjt – figyelmeztet a különféle következetlenségekre, illetve ezek bizonyos típusait a saját hatáskörében korrigálja is.

Terjedelmi korlátok nem lévén, megengedhetjük magunknak, hogy a szócikkekben az egyes szójelentéseknél minden régióra vonatkozóan külön-külön stílusminősítések szerepeljenek, s így könnyen kitűnjenek mind a régióközi hasonlóságok, mind a különbségek. Az egyes jelentések melletti felségjelzésekből nagyon könnyű áttekinteni, mely jelentések vannak meg csak egy, melyek több régióban. Mivel – amint fentebb jeleztük – az etimológiákban nemcsak a közvetlen átadó nyelvet jelöljük meg (az legtöbbször az adott régió többségi nyelve), hanem a távolabbi átadó nyelveket is, az etimológiai sorok összehasonlításával arra is fény derül, milyen úton-módon került be egy-egy nemzetközi vagy más idegen eredetű szó az átadó nyelvbe.

5. ÖSSZEGZÉS HELYETT

A Termini magyar–magyar szótár és adatbázis a hét külső régióban zajló lexikológiai és lexikográfiai kutatások eredményeinek folyamatos beépítésével a Magyarországon kívül használt magyar nyelvváltozatok tanulmányozásához szolgáltat nagyszámú adatot. A szócikkek a mai magyar nyelvi jelenségek lexikológiai, szociolingvisztikai, kontaktológiai, etimológiai, leíró nyelvészeti, stilisztikai, pragmatikai kérdéseinek vizsgálatát teszik lehetővé, ahogy ezt az adattár alapján készült tanulmányok sokszínűsége is jelzi (Benő, 2011, 2015; Eöry, 2007; Kolláth, 2007; Lanstyák, 2006a, 2006b, 2006c, 2008; Tolcsvai Nagy, 2007, 2018; Szoatak, 2017). A nyelvészeti kutatásokon kívül az internetes szótár régióspecifikus adatokkal segítheti a néprajzi, kultúrtörténeti, antropológiai, szociológiai és szociálpszichológiai vizsgálatokat. A Termini-szótárt és adatbázist fentebb említett jegyei egyedülállóvá teszik az internetes szótárak körében.

Befejezésül arról szeretnénk szólni, mi a jelentősége magának a határtalanításnak, melynek a Termini-szótár és -adatbázis a legfontosabb eszköze. A magyar szótárak és kézikönyvek egyetemes léptékűvé válása először is **szimbolikus** okokból fontos, hiszen arra utal, hogy a kizárólag vagy nagyjából a határon túli régiókban használatos szavak, kifejezések is a magyar nyelv legitim elemei. A szimbolikus okokon túl a külső régióbeli szóanyag bekerülésének **gyakorlati jelentősége** is van: fontos szerepet tölt be a magyar–magyar kommunikáció „akadálymentesítésében”. Végül a szimbolikus megjelenítés és a gyakorlati hatás mellett a külső régiók szóincsének szótári rögzítése **elméleti jelentőségű**: serkentője lehet a magyar szókészlet sokszínűségével, differenciálódásával kapcsolatos vizsgálatoknak, ezeket egy újabb dimenzióval gazdagíthatja.

IRODALOM

- Benő A. (2011): Köznyelvi lexikai egységek szemantikai sajátosságai a kisebbségi magyar nyelv-változatokban. In: Benő A. – Péntek J. (szerk.): *A Termini Magyar Nyelvi Kutatóhálózat tíz éve*. Dunaszerdahely–Kolozsvár: Gramma Nyelvi Iroda–Szabó T. Attila Nyelvi Intézet, 116–126.
- Benő A. (2013): Tükörszók, tükörszerkezetek és jelentéskölcsönzések a kisebbségi magyar nyelv-változatokban. In: Szoták Sz. – Vargha F. S. (szerk.): *Változó nyelv, nyelvváltozatok, területiség*. A VII. Nemzetközi Hungarológiai Kongresszus szekcióelőadásai. Kolozsvár: Egyetemi Műhely Kiadó, 15–26. <http://mek.oszk.hu/13200/13242/13242.pdf>
- Benő A. (2014): *Kontaktsujelenségek az erdélyi magyar nyelvváltozatokban*. Kolozsvár: Erdélyi Múzeum-Egyesület, https://www.academia.edu/37577944/Kontaktsujelens%C3%A9gek_az_erd%C3%A9lyi_magyar_nyelvv%C3%A1ltozatokban
- Benő A. (2015): Többirányú lexikai átvételek a mai magyar nyelvváltozatokban. In: Benő A. – Fazakas E. – Zsemlyei B. (szerk.): *Többnyelvűség és kommunikáció Kelet-Közép-Európában*. Kolozsvár: Erdélyi Múzeum-Egyesület, 248–256. <https://eda.eme.ro/handle/10598/31292?show=full>
- Benő A. (2017): Termini – az internetes magyar–magyar szótár és adatbázis. In: Fábán Zs. (szerk.): *Szótárírás a Kárpát-medencében*. Budapest: Tinta Könyvkiadó, 209–227.
- Eöry V. (2007): Határtalanítás a magyarországi szótárakban: Maticsák S. et al. (szerk.): *Nyelv, nemzet, identitás*. I. Debrecen–Budapest: Nemzetközi Magyarságtudományi Társaság, 27–32. <https://mek.oszk.hu/05100/05146/pdf/>
- Kolláth A. (2007): Akkor hogyan is beszélünk?... (A ht adatbázis muravidéki elemei és használati gyakoriságuk) *Kisebbségkutatás*, 16, 4, 767–783. <https://epa.oszk.hu/00400/00462/00036/1512.htm>
- Lanstyák I. (2006a): A nyelvi változatosság mint szótártani probléma. *Nyelvből nyelvbe*. Pozsony: Kalligram Könyvkiadó, 57–104., 245–257.
- Lanstyák I. (2006b): Nyelvi változatosság a határon túli magyar szókinésben. *Kisebbségkutatás*, 15, 1, 60–73. <https://epa.oszk.hu/00400/00462/00029/cikk8825.html>
- Lanstyák I. (2006c): Határtalanítás. In: Mártonfi A. – Papp K. – Slíz M. (szerk.): *101 írás Pusztai Ferenc tiszteletére*. Budapest: Argumentum Kiadó, 179–286.
- Lanstyák I. (2008): A magyar szókészlet szétfejlődése 1918 után. In: Fedinec Cs. (szerk.): *Értékek, dimenziók a magyarságkutatásban*. Budapest: MTA Magyar Tudományosság Külföldön Előnlöki Bizottság, 118–135. https://mta.hu/data/dokumentumok/magyar_tudomanyosság_kulfoldon/Ertekek-dimenziok.pdf
- Lanstyák I. – Benő A. – Juhász T. (2010): A Termini magyar–magyar szótár és adatbázis. *Régio*, 4, 37–58. http://epa.oszk.hu/00000/00036/00079/pdf/Regio_2010_03_037-058.pdf
- M. Pintér T. (2017): Adatbányászat a kétnyelvűség mögött. In: Benő A. – Gúti E. – Juhász D. et al. (szerk.): *Tudományköziség és magyarságtudomány a nyelvi dimenziók tükrében*. Törökbálint: Termini Egyesület, <http://mek.oszk.hu/18600/18698/18698.pdf>
- Szoták Sz. (2017): A Termini magyar–magyar szótár és adatbázis burgerlandi gyakorlati példái. In: Fábán Zs. (szerk.): *Szótárírás a Kárpát-medencében*. Budapest: Tinta Könyvkiadó, 228–240.
- Tolcsvai Nagy G. (2007): A határtalanítás egy újabb megvalósulásáról. In: Benő A. – Fazakas E. – Szilágyi N. S. (szerk.): *Nyelvek és nyelvváltozatok II*. Kolozsvár: Anyanyelvápolók Erdélyi Szövetsége, 399–405.
- Tolcsvai Nagy G. (2018): Esettanulmány a szlovákiai magyar nyelvváltozat lexikai szemantikai változatairól. *Magyar Nyelvőr*, 142, 401–414. <http://real.mtak.hu/93150/>