

FÓRIS ÁGOTA

## A TERMINOLÓGIAI SZÓTÁRAKTÓL A TERMINOLÓGIAI ADATBÁZISOKIG<sup>1</sup>

### 1. Bevezetés

A tanulmány célja a *szakszótárak, terminológiai szótárak és terminológiai adatbázisok* témakörén belül a legfontosabb magyar nyelvű vonatkozások áttekintése és összefoglalása. A szakszótárak és terminológiai adatbázisok témája rendkívül aktuális, és régóta foglalkoztatja a nyelvészeket, lexikográfusokat és terminológusokat. Ha történeti szempontból nézzük, a szakszótárak egy nagyobb vagy kisebb szakterület szókincsét gyűjtik össze egy-, két- vagy többnyelvű szótárakban, és a szótár típusától függően adnak meg információt (elsősorban értelmezést és/vagy ekvivalenseket) a címszavakról. A szakszótárak a 18–19. századtól kezdve jelentős szerepet játszottak a szakszókincs összegyűjtésében, közzétételében és terjesztésében, egyúttal normatív funkciót is betöltöttek: szerkesztőik a szakterületek elismert szakembereiként azt a szókincset (korabeli terminussal: „műszavakat”) gyűjtötték össze, amelyet az adott szakma használt. Mivel számos szakma és tudományterület magyar nyelvű szókincse nem létezett, vagy nem volt összegyűjtve, általános volt, hogy a szakszótárak szolgálták az újonnan létrehozott terminusok közzétételi helyeként. Az egyes tudományágak önálló nomenklatúrájának és terminológiájának kialakítása a 18–19. századi nyelvújítás során kezdődött meg, és a sikeres megoldáshoz vezető, tudatosan irányított munkálatok eredményeinek nagy részét az új orvosi, matematikai, fizikai, növénytani stb. terminusok képezték a 20. században (Fóris 2005). Az egységes szakszókincs terjesztését a szakterület szakembereinek szervezett együttműködésében készült szakszótárak kiadásával oldották meg. Tolnai (1929: 21) alapján a szakszótárakat nagyobb mennyiségben a magyar nyelv egyes jogait törvénybe iktató 1805. évi 4. tc. után kezdték kiadni, mivel létre kellett hozni a közigazgatás, a törvénykezés és a hivatali eljárás magyar szakszókincsét, amelyet a „Tiszti Szótárak”-nak nevezett szakszótárakban adtak ki. A Magyar Tudományos Akadémia is

---

<sup>1</sup> A tanulmány az MTA TMNP2023-1/2023 projekt támogatásával készült.

szakszótárak szerkesztésével és kiadásával kezdte a működését. Tolnai (1929) szerint a szakszótárakat jellemzően nem egy ember, hanem egy bizottság készíti, többnyire normatív céllal. A nyelvújítás idején létrehozott szakszókincsel és az azt közreadó szótárakkal, a téma elméleti és gyakorlati kérdéseivel több összefoglaló kötet is foglalkozik (pl. Tolnai 1929, Pais 1955), az 1922-ig kiadott szakszótárak adatai pedig megjelentek a Sági-féle bibliográfiában (Sági 1922). A szakszótárak népszerűsége és jelentősége egészen a 20. század végéig megmaradt, amikor a lexikográfiában az 1990-es években lezajlott paradigmaváltás következtében (lásd Fóris 2002a: 12–16) az online szótárak egyre népszerűbbé váltak a nyomtatott szótárakhoz képest; majd újabb jelentős változást hozott a 2010-es évekre a szemantikus web technológiák<sup>2</sup> megjelenése és elterjedése (Tóth 2010), melynek következtében az online szótárak, adatbázisok és ontológiák készítése a korábbiaknál sokkal egyszerűbbé vált, ezzel pedig a szakszókincs közzététele is könnyebb lett (részletesen lásd Fóris 2012). Az utóbbi években Muráth tett erőfeszítéseket a magyar szaklexikográfia eredményeinek összegyűjtésére, rendszerezésére, melyet tanulmánykötetben jelentettek meg (Muráth 2020b), továbbá folyamatban van a magyar nyelvű szakszótárak és -lexikonok bibliográfiájának készítése (Arató–Muráth 2019, lásd Muráth 2020a).

A nyelvészet és a természettudományok is megkülönböztetik egymástól a terminológiát és a nomenklatúrát (nevezéktant). Az egyes szaktudományok saját, önálló rendszert hoztak létre a szakterületek nomenklatúrájának kialakítására (pl. kémia, biológia, anatómia). A nomenklatúráknak vannak ugyan bizonyos egymással megegyező szabályai, de tudományterületenként alapvetően különböző szabályokat alkalmaznak. Nyelvtörténeti szempontból az egyes tudományterületek magyar nyelvű terminológiájának és nomenklatúrájának első létrehozása a 18–19. század során lezajló nyelvújítás idejére esik, de az ekkor kialakított kifejezések nagy része elszigetelt és időszakos használatú megjelölés maradt. Nemzetközi szinten a nomenklatúra és a terminológia koordinálása és szabványosítása a 20. században valósult meg, és ehhez a nemzetközi szabványosítási tevékenységhez kapcsolódtak a Magyar Tudományos Akadémián keresztül a magyar kutatók is a magyar nyelvű nevezéktanok (szabályzatok) és szótárak létrehozásával. E tevékenység összefoglaló bemutatását lásd *A névtudomány, a nevezéktan és a terminológia viszonya, érintkezési pontjai* című tanulmányban (Bölcskei–Fóris 2022).

<sup>2</sup> A háromtagú összetétel írása a szerző kérésére a terminológiai szakirodalom írásmódját követi. (A szerkesztők)

## 2. Szótárak, szakszótárak

A pályám kezdetétől foglalkoztatott a szótár, mint referenciamű fogalmának szűkebb és tágabb értelmű meghatározása, a szakszótárak, terminológiai szótárak és terminológiai adatbázisok közötti tartalmi és szerkezeti hasonlóságok és különbségek összegyűjtése és leírása (pl. Fóris 2002a: 12–28, 2002b, 2007, 2012, 2013). Ezek eredményeként megállapítottam, hogy a magyar nyelvben a *szótár* szónak van egy szűkebb és egy tágabb jelentése: szűkebb értelemben a nyelvi szótárakra vonatkozik; tágabb értelemben pedig a *lexikográfiai referenciamű* fogalmával azonosítható, amely nemcsak a nyelvi szótárakat, hanem a szakszótárakat, lexikonokat és enciklopédiákat is magába foglalja. „A szótár olyan egy-, két- vagy többnyelvű gyűjteményes adatbázis, amely – hordozótól függetlenül – nyelvi egységek jellemzőit, és ezeknek adott szempont szerint, kulcsszavakkal történő osztályozását tartalmazza, egy meghatározott korpusz alapján” (Fóris 2007: 301).

A *Dictionary of Lexicography* (DoL 1998) és Hartmann (2001) meghatározásai alapján a *lexikográfia* olyan szakmai tevékenység és kutatási terület, amely a lexikográfiai típusú művekkel foglalkozik: szótárakkal és más referenciaművekkel, vagyis a lexikográfusok nem csak nyelvi szótárakkal foglalkoznak, hanem más típusú lexikográfiai művekkel is (például enciklopédiákkal és tezaurusokkal). A *szótár* a legelterjedtebb típusú referenciamű („dictionary – the most common type of reference work”); *referenciamű*nek nevezünk minden olyan terméket, akár könyv, akár számítógépes szoftver, amely lehetővé teszi az embereknek, hogy információt viszonylag könnyen és gyorsan tároljanak és előhívjanak.

A *szakszótár* elnevezés a magyar nyelvben arra utal, hogy olyan szótárakról van szó, amelyek címszavai a szakszókincset tartalmazzák, és arra vonatkozóan adnak információkat. A szakszótárak lehetnek egy nyelvűek, kétnyelvűek vagy többnyelvűek; listaszerűek, továbbá tartalmazhatnak nyelvi vagy enciklopédikus információkat. Vannak csak egyféle információt megadó szakszótárak, más szakszótárak egyszerre többféle információt is megadnak: idegen nyelvi ekvivalenst, értelmezést és példamondatokat. A szakszótárak egy része alfabetikus, más része tematikus elrendezésű. Hordozójuk lehet nyomtatott, elektronikus vagy online. Álláspontom szerint az, hogy egy adott művet szakszótárnak tartunk-e vagy más kézikönyvnek, nemcsak attól függ, hogy a címében szerepel-e a szótár szó, hanem attól is, hogy tartalmában és szerkezetében is szótári jellemzőket mutat-e (lásd Fóris 2013).

A nyelvi szótárak és enciklopédikus szótárak közötti különbség az információ megadásában van: az egyik a címszavakról ad információt, a másik a címszavak denotációjáról. A nyelvi szótárak (a szakszótárak is) jelalapúak (a kiindulás a lexikai egységből történik, a lexikai egységről mint jelölőről adnak információt), míg a terminológiai szótárak fogalomalapúak (a kiindulás a fogalomból történik, a denotációról adnak információt). A szakszótárak lemmaállományát a köznyelvi szókincshez viszonyítva gyűjtik, vagyis a szakszótárakban a leggyakoribb és legfontosabb szakszavakról adnak információkat (nyelvészeti és enciklopédikus információ is szerepelhet bennük). A terminológiai szótárak lemmaállománya egy domént fed le (vagyis nem a gyakorisági és fontossági szempontokat veszik figyelembe a címszavak kiválasztásakor, hanem az adott domén minél teljesebb körű leírására törekednek).

### 3. A terminológiamenedzsment és a terminológiai források

A terminológiai adatbázisok elődei a terminológiai szótárak és szójegyzékek. Elektronikus terminológiai adatbázisokat már az 1960-as években készítettek, ezeket kezdetben szerverekről lehetett elérni modemkapcsolaton keresztül, majd az 1980-as évektől CD-ROM-ra kerültek át az adatok. Csak az 1990-es, 2000-es évektől kezdve jelentek meg a sokkal több adatot tartalmazó és kezelni tudó és a weben keresztül elérhető online terminológiai adatbázisok. A terminológiai adatok elektronikus tárolási lehetősége a kezdetektől fogva hasznosnak látszott, több ok miatt is: a tárolandó terminológiai adatok mennyisége hatalmas, és biztosítani kell a gyors visszakereshetőségüket; a szakmai ismeretek gyorsan elavulnak, ezért gyakran kell frissíteni őket; a terminológiai leírások fogalomalapúak (ellentétben a jelalapú általános nyelvi szótárakkal), ezért rendszerezettebben rögzítik őket; továbbá a terminológiai adatok rögzítésével kapcsolatos kezdeti aggodalmak szerencsésen egybeestek a strukturált adatok tárolására szolgáló technológiák elérhetőségével (L'Homme 2014: 1480).

A terminológiai adatokat többféle kiadványban és módon lehetséges közzétenni. Az ISO 1087:2019 szabvány szerint *a terminológiai források* (terminology resources) nyomtatott és elektronikus formában egyaránt léteznek, és többféle fajtájuk van. Ide tartozik

- a terminológiai adatbázis (terminology database, termbase);
- a terminusbank (term bank, terminological data bank), amely terminológiai adatbázisok gyűjteménye;

- a terminológiai szótár (terminological dictionary, special-language dictionary, LSP dictionary), amely olyan terminológiai forrás, amelyet azzal a céllal terveztek, hogy referenciaműként legyen használható;
  - a szójegyzék (vocabulary), amely egy-, két-, vagy többnyelvű terminológiai szótár;
  - a glosszárium (glossary), amely többnyire egynyelvű terminuslista;
  - és a nómenklatúra (nomenclature), amely nevezéktani szabályok alapján összeállított terminológiai forrás.
- (Ahogy ebből a felsorolásból kitűnik, a szabvány egy kategóriába sorolja a terminológiai szótárakat és a szakszótárakat.)

A terminológia gyakorlati területét terminológiai munkának, illetve terminológiamenedzsmentnek hívják. A terminológiai munkára vonatkozó ISO 1087:2019 szabvány szerint a *terminológiai munka (terminológiamenedzsment)* a fogalmak és jelölők szisztematikus gyűjtésének, leírásának, feldolgozásának és közzétételének összefoglaló neve, a *terminográfia* pedig a terminológiai munkának az a része, amely a terminológiai források létrehozásával és fenntartásával foglalkozik. A terminológiamenedzsment tehát tágabb fogalom, mint a terminográfia. Különösen fontos felhasználási területe a fordításorientált terminográfia (lásd Fóris 2020).

A terminológiai munka jellemző tevékenységei közé az alábbiakat sorolják a szakirodalomban: a terminusok azonosítása, gyűjtése, elemzése és leírása; a terminusok osztályozása és ábrázolása; az esetlegesen hiányzó terminusok képzése és pótlása; a terminológiai rendszer terminusainak szabványosítása és érvényesítése; végül azok kezelése, rögzítése és közzététele (Pavel 2004). A terminológiai munka során ma már elengedhetetlen az informatikai eszközök, legfőképp a terminológiai munkát segítő szoftverek használata. A terminológiai munka során végzett tevékenységek eredményét elsősorban terminológiai adatbázisokban és tudásbázisokban jelenítik meg (Galinski–Budín 1993). „A terminológiai adatbázis olyan elektronikusan tárolt terminológiai adatok összessége, amely onomasziológiai szemléletet követve készült, megalkotásának előfeltétele az adott szakterület fogalmi rendszerének felvázolása. Egy vagy több szakterület terminusait és a hozzájuk tartozó definíciókat tartalmazza egy vagy több nyelven” (Sermann–Tamás 2010: 113).

Gyakori kérdés, hogy egy adott forrásmű a gyűjteményes adatbázis, a terminológiai adatbázis vagy tudásbázis kritériumainak felel-e meg.

A 2. pontban adott meghatározás szerint ezek mind referenciaműveknek tekinthetők, amelyek a legfőbb adatok mellett sok más információt is tartalmazhatnak. Az adatbázis egy adott doménra vonatkozó adatokat tárol, hasonló jellemzőjű strukturált adatok gyűjteménye, amely alkalmas az adatok gyors lekérdezésére; általában elektronikus formátumú. A terminológiai adatbázis egy- vagy többnyelvű, terminológiai adatokat tartalmazó adatbázis, nemcsak címszavak szerepelnek benne, hanem úgynevezett fogalomazonosító szám, nemcsak nyelvi, hanem fogalmi és pragmatikai információkat is megadnak a készítői. A tudásbázis enciklopédikus és ontológiai ismereteket tárol, és nemcsak adatokat, hanem ismeretelemeket és szabályrendszereket is tartalmaz.

A terminológiai adatbázisokban az adatoknak általában két, három vagy négy szintje van, pl. fogalmi szint, nyelvi szint, a terminus szintje, szó szint. A terminológiai adatbázisok adatmezői a terminusokra, fogalmakra, definíciókra vonatkozó adatokon kívül más információkat is tartalmaznak, pl. a forrás adata, a készítő neve vagy a létrehozás/módosítás időpontja. A legtöbb terminológiai adatbázisban a fogalmakat és terminusokat egy úgynevezett *fogalomazonosító* segítségével kötik össze, amely egy azonosító kódszám (Prószéky–Kis 1999, Tamás 2014). Gyakorlati szempontból Demeczky (2008) ismerteti az IBM terminológiai szoftverének kétféle – fogalomközpontú és szóközpontú – megjelenítési felületét. A szóközpontú megjelenítés a jelölő adatait, alakváltozatait, grammatikai jellemzőit, szöveggörnyezetét és definícióját adja meg. A fogalomközpontú megjelenítés egyetlen fogalmat mutat, amelynek részletesen megadja az adatait: a definícióját, a fogalomhoz kapcsolódó szóalakokat, a domént és a kapcsolódó fogalmakat.

A gyakorlatban a terminológiai adatbázisok sokféle formában készülhetnek, sokféle adatot tartalmazhatnak, és sokféle szerkezetük (struktúrájuk) lehet. Sager (1990) a terminológiai adatbázisoknak négy típusát különbözteti meg: a) hierarchikus adatbázisok, b) adatbázis-hálózatok, c) egyszerű adatbázisok, d) relációs adatbázisok. Tamás (2014: 111) háromféle adatbázistípust különített el: egyszerű, hagyományos és összetett adatbázisokat; a terminológiai adatbázisok legfontosabb jellemzőiként a fogalomközpontúságot és a terminusautonómiát jelölte meg; egy táblázatban foglalta össze a terminológiai adatbázisok típusait és főbb jellemzőiket (1. táblázat).