SERMANN ESZTER – TAMÁS DÓRA
SZTE Idegennyelvi Kommunikációs Intézet, Szeged
Országos Fordító és Fordításhitelesítő Iroda Zrt., Budapest
tamas.dora@t-online.hu; sermann@lingo.u-szeged.hu

Elektronikus szótár vagy terminológiai adatbázis?

Kulcsszók: *online szótár, terminológiai adatbázis, szóközpontú megjelenítés, fogalomközpontú megjelenítés*

Az online szótárak és a terminológiai adatbázisok a célnyelvi szövegek terminológiai előkészítésében a fordítók olyan hatékony segédeszközei, amelyek azonos és eltérő jegyekkel egyaránt rendelkeznek. Az általános jellemzők közé sorolható közös vonásként, hogy az online elektronikus szótárak és terminológiai adatbázisok elektronikusan tárolt egy-, két- vagy többnyelvű elektronikus segédeszközök. Az elektronikus tárolás előnyei közé tartozik az adatok egyszerű frissítésének lehetősége, az online, számos esetben ingyenes elérhetőség, a gyorsabb keresési lehetőség. Az eltérések között tarthatjuk számon, hogy amíg az online elektronikus szótár központi eleme a szótári szó és általában szemasziológiai módszerek alapján dolgozzák ki a szócikkeket, addig a terminológiai adatbázis készítésének alapja az onomasziológiai módszer és központi eleme a terminus által jelölt fogalom.

De mennyire élesek ezek a határok? Az online elektronikus szótár nyelvi adataihoz nem alkalmazhatók ugyanúgy a terminológiai módszerek? Vajon a terminológiai adatbázis megjelenítési felülete nem hasonlíthat az elektronikus szótáréra? Brekke (2001) szerint a két szemlélet eleve kiegészíti egymást, és alkalmazásuk mikéntje a kidolgozandó segédeszközök felhasználási céljától függ. Tapasztalataink alapján az online szótárak és terminológiai adatbázisok elkülönítése a valóságban valóban nem egyszerű, egyetértünk Fórissal (2012) abban, hogy a különböző típusú fordítói segédeszközök készítése és használata számos osztályozási kérdést vet fel. A kutatás célja a releváns szakirodalom segítségével egy olyan lehetséges szempontrendszer felvázolása, amely elkülöníti egymástól az online szótárakat és a terminológiai adatbázisokat, és amelynek alapján meghatározhatjuk, hogy egy-egy online segédeszközt mely jellemzői alapján vélhetjük egyik vagy másik kategóriába tartozónak. A vizsgálat során a következő fő kérdésekre kerestük tehát a választ: Az elemzéshez milyen szempontrendszer alkalmazható? Melyek az online elektronikus szótár és a terminológiai adatbázis fő jellemzői? Mennyire elmosódottak köztük a határok?

A vizsgálathoz a következő vizsgálati módszereket alkalmaztuk: a releváns szakirodalom áttekintése és konkrét példák elemzése. Kiindulásként definíciót kerestünk a szótárra: „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 definíció szerint az online szótárak is ebbe a kategóriába sorolhatók. A szótárak elemzéséhez Fóris és Rihmer (2007) egy lehetséges szempontrendszert dolgozott ki, melynek fő elemei a következők: a szótárral kapcsolatos háttér-információk (szűkebb és tágabb háttér), a szótár tartalmáról szóló információk (a szótár fizikai leírása, mega-, makro-, mikro- és mezostrukturális jellemzői), a szótár tartalma (lexikológiai alapok, fonológiai/grafémikai, grammatikai, szemantikai és egyéb releváns nyelvészeti információk) és a szótár használatáról szóló információk vizsgálata.

Gaál (2010a, 2010b) és Dringó-Horváth (2011) e szempontrendszer alkalmazhatóságát vizsgálja az online szótárakra vonatkoztatva. Gaál (2010a) írásában a különbségek minősítési kritériumokban való megfogalmazására törekszik. A hivatkozott szempontrendszert alkalmazza két online szótárra, az MTA-SZTAKI és a *The Longman Dictionary of Contemporary English Online* szótárakra, és megállapítja, hogy a szempontok nagy része az online szótárak elemzésére is megfelelő. Gaál (ibid.) a következőképpen összegzi az online szótárak előnyeit a papíralapú szótárakkal szemben: nagyobb információmennyiség, folyamatos frissíthetőség, az információk szemléltethetőségének lehetősége audiovizuális eszközökkel, külső, belső hivatkozások, gyorsabb, egyszerűbb használat a beépített kereső opciónak köszönhetően, valamint az ingyenesség. Egyedüli hátrányukként az jelöli meg, hogy kérdéses a megbízhatóságuk. Dringó-Horváth (2011) szerint új szempontokra van szükség, például elérhető infor-

máció, kiegészítő anyag a vásárlás előtt, üzembe helyezés és programkezelés, súgó és használati utasítás, licence-lehetőségek és ár, keresési funkciók, technikai és tartalmi flexibilitás, multimedialitás, hipermedialitás.

Pastor és Alcina (2010) szerint az online szótáraknak számos előnye van a hagyományos, papíralapú szótárakkal szemben, de a felhasználók gyakran nem jutnak hozzá a bennük található információkhoz általában két okból: 1) a szótár nem teszi lehetővé megfelelő módon az információhoz való hozzáférést, 2) a felhasználók nem tudják használni őket. Az elektronikus szótárak szerkesztői gyakran nem használják ki az elektronikus formátum adta lehetőségeket, így szótáraik a papíralapú szótárak egyszerű másolatai. De Schryver (2003) az elektronikus szótárak előnyei közül – egyebek mellett – a következőket említi: nem betűrendes, nem lineáris, nem konstans, az utalások rendszere optimális, nagy adatmennyiség, a lexémákhoz való gyors hozzáférés korpuszban, beépíthetőség más szoftverekbe, dinamikus frissítés, olcsó ár vagy ingyenesség.

Az elektronikus szótárak jellemzőinek leírását követően nézzük, hogy mit érdemes a terminológiai adatbázisokról megemlíteni. Korábbi tanulmányunkban (Sermann, Tamás 2010) a szakirodalom alapján összefoglaltuk a terminológiai adatbázisok jellemzőit, és egyes adatbázisokat részletesen elemeztünk. A terminológiai adatbázist olyan elektronikusan tárolt terminológiai adatok gyűjteményeként határoztuk meg, amely onomasziológiai szemléletet követve készült, és egy vagy több szakterület terminusait tartalmazza egy vagy több nyelven a hozzájuk tartozó definíciókkal együtt. A terminológiai adatbázisok építésére léteznek szabványok (ISO 12620:2009), amelyek például felsorolják az adatmezők típusait. A terminológiai adatbázisok készítésének két alapelvét több szabvány (ISO 704:2009, ISO 12200:1999 és az ISO 16642:2003) is tartalmazza. A terminológiai adatbázisok szerkesztésében a két domináns alapelv a következő: a fogalomközpontúság és terminusautonómia.

A fogalomközpontúság elve szerint az adatbázis központi eleme az azonosító kódszámmal ellátott fogalom, amelyet a terminus jelöl és a definícióban írható le. A terminusautonómia elve szerint a terminusok önálló szerepeltetésére kell törekedni, azaz egy fő terminust (main term) kell kiválasztani és a szinonimákat külön bevitelként kezelni, azaz külön adatmezőben feltüntetni, és lehetőség szerint hivatkozás révén a rá vonatkozó adatokat külön terminográfiai cédulán részletesen kidolgozni.

Arntz et al. (2009) egy sematikus ábrát is mellékel, amely az egyszerű terminológiai adatbázis terminográfiai cédulájának alapvető leíró mezőit (adatmezőit) tartalmazza, mint például: szakterület, nyelv, azonosítószám és egyéb kódok, terminus, nyelvtani jellemzők, definíció és kontextus, szinonima feltüntetése, szerkesztők, dátum, ahol a nyelvi adatok mindegyike mellett szerepel a forrás. A hagyományosnak vagy klasszikusnak nevezhető terminológiai adatbázisok, adatbankok szigorúan onomasziológiai felépítésűek, azaz a *fogalomközpontú megközelítés* alapján rendezik az adatokat. Ugyanakkor Demeczky (2008: 201) tanulmányában az szerepel, hogy az IBM vállalaton belül a fogalomközpontúság szerkesztési elvei szerint kidolgozott terminológiai adatbázis két különböző megjelenítési felülettel rendelkezik. Az egyik *fogalomközpontú*, míg a másik *szóközpontú* felfogású és megjelenítési módú. A szóközpontú megjelenítés megmutatja egy jelölő összes adatát, alakváltozatait, nyelvtani jellegzetességeit, szövegkörnyezetét és a szó különböző jelentéseiként tárolva a nyelvi jel által jelölt fogalom meghatározását vagy meghatározásait, ezáltal formailag elektronikus szótárra emlékeztet. Ezzel szemben a fogalomközpontú megjelenítés csak egyetlen fogalmat mutat, amelynek megadja definícióját, a fogalmat jelölő szavakat (szinonimákat, ekvivalenseket), a tárgykört és a kapcsolódó fogalmakat, vagyis érvényesül a terminusautonómia elve. Azonban hangsúlyoznunk kell, hogy az alapvető szerkesztési mód a megjelenítési móddal nem keverendő össze. Megjegyezzük viszont, hogy megjelenítési mód elemzéskor a szerkesztési módnál egyszerűbben vizsgálható, mivel a szerkesztési módszerek bizonyos szempontjai rejtve maradhatnak.

Tamás (2012) kísérletet tesz a terminológiai adatbázisok hármas felosztására, amely szerint megkülönbözteti az egyszerű terminológiai adatbázist az ún. hagyományos és az összetett terminológiai adatbázistól. A besorolás szerint az egyszerű terminológiai adatbázisban az adatmezők száma a minimális mennyiséget nem éri el (Arntz et al. 2009) vagy a fogalomközpontúság mellett nem érvényesül az terminusautonómia, míg az ún. hagyományos terminológiai adatbázistól, amelyben a minimálisan előírt adatmezők szerepelnek, és központi eleme a definíció, érvényesül a terminusautonómia. Osztályozásában külön csoportot alkotnak az összetett terminológiai adatbázisok, amelyek a hagyományos változaton kívül további ismeretelemeket vagy kiegészítő eszközöket tartalmaznak (például jogszabályok gyűjteménye, terminológiai rendszerek grafikai ábrázolása, terminuskivonatoló és terminuskinyerő segédeszközök).

Lássunk most – a teljesesség igénye nélkül – néhány konkrét példát. Elsőként egy online szótárt elemeztünk. A Wordreference a honlapon olvasható információk szerint ingyenesen hozzáférhető online szótárak gyűjteménye. 1999-ben létesítették, azóta folyamatosan bővül, és az egyik legtöbbet használt internetes online szótár-forrás, amely nyelvi fórumokat is tartalmaz. Elemzésünkhöz konkrét szócikkre kerestünk rá. Kiválasztottuk az angol−olasz nyelvpárt, és a keresési mezőbe begépeltük a *formazione* (it.) szót. A keresés eredményeképpen látható, hogy ebben a nyelvpárban a honlap két szótárt tartalmaz: a *The Concise Oxford Paravia Italian Dictionary* és a *WordReference English−Italian Dictionary*. A szócikk a két szótárból vett információkat külön, egymás után közli. Az oldal tetején a címszó helyezkedik el. Bal oldalon külön sávban a hasonló szavak listája található, míg közvetlenül a címszó alatti első sorban négy hivatkozást találunk: *Definizione, Conjugator, In context* és *Images*, amely mind külön funkciót takar. A *Definizione* (definíció) linkre kattintva egy olasz egynyelvű szótárból másik szócikket találunk. A *Conjugator* egy olasz igeragozási segédlethez vezet át bennünket. Az *In context* a *formazione* szót kontextusba helyezve mutatja be. Az *Images* alatt pedig a Google keresőmotor a *formazione* szóhoz kapcsolódó képekre keres rá. Az első szótár bejegyzései, amely a *Concise Oxford Paravia Italian Dictionary* csak ezt követően olvashatók. A szócikkben a fonetikai átírás mellett a hangszóró ikonja hangfájlt takar, így segítve a felhasználót. Információt kapunk azonban a nyelvtani nemről is (a szófaj megjelölése hiányzik), majd a *formazione* (it.) különböző jelentéseihez kapcsolódó angol nyelvű ekvivalensekről. Ezt követik azok a szócikkek, amelyekben még megtalálható a *formazione*. Az első szótár alatt található a második szótár, amely a *WordReference English−Italian Dictionary*-ból származó információkat sorolja fel. Az egyes angol nyelvi jelentések külön szócikként szerepelnek, a szócikk után a szófaj és a nyelvtani nem megjelölése következik, majd az angol nyelvű ekvivalens mint fordítási javaslat. Végül az utolsó információk linkek útján azokba a fórumokba engednek betekintést, amelyekben szerepel a *formazione* (it.).

A terminológiai adatbázisokban ugyanarra a terminusra kerestünk rá. Az EOHS Term tudásbázis a Bolognai Tudományegyetem Fordító- és Tolmácsképző Intézet műhelyében készült. Az interneten 2003 óta tanulmányozható, egy viszonylag szűk szakterületet, azaz a munkahelyi baleset-megelőzésről és biztonságról szóló jogszabályok anyagának kb. 800 terminusát dolgozza fel négy nyelven (olasz, angol, francia, német). Elnevezése is jelzi, hogy nem egyszerűen adatbázisról vagy adatbankról van szó, hanem tudásbázisról, azaz a terminológiai adatbázison kívül más ismeretelemeket is tartalmaz. Három részből áll: adatbank, a jogszabályokon belüli keresést lehetővé tevő források könyvtára és a terminológiai rendszerek grafikai ábrázolása, amelyek között az átjárás minden esetben biztosított. A forrásoknak köszönhetően tanulmányozhatók az annotált jogszabályok és konkordanciák is lekérhetők. A terminológiai rendszerek gyűjteményének előnye, hogy gyorsan és egyszerűen lehetővé teszi a fordító számára, hogy egy-egy terminus megfelelő fordításához a fogalmi rendszerre, az alá, fölé- és mellérendeltségi viszonyokra rálátást kapjon. A terminográfiai cédulán *formazione* (it.) terminusra való keresés eredményeképpen a következő adatmezők szerepelnek: a terminus alatt elsőként hiperhivatkozással a tárgyköri besorolás látható (rákattintva terminuslistát kapunk), ez alatt találhatók meg a nyelvtani jellemzők, a definíció, annak forrása, a jogszabályi hivatkozás (európai uniós és olaszországi), a kapcsolódó terminusok, a fogalmi háló hiperhivatkozással, végül más nyelveken szereplő megfelelők. Az ekvivalenciaszintet jelen esetben az = jel mutatja, amely a teljes ekvivalenciára utal. A részleges ekvivalencia jele a ±, míg az ekvivalencia hiányát a ~ jelzi. Ezek a jelölések az egyes idegen nyelvű terminusok előtt szerepelnek, és az ekvivalensnek a fő terminushoz való viszonyát mutatják meg. A beazonosítás érdekében az egyes idegen nyelvek ekvivalensei előtt az adott nyelv rövidítése (ISO kódja) szerepel. A terminus felett pedig – szintén hiperhivatkozással – a „Concordance" mezőre kattintva a források könyvtárából konkordancia-találatok kérhetők le (Tamás 2010).

A Termium Plus a kanadai kormány hivatalos terminológiai adatbankja, amelyet a kanadai kormány központi állami fordítóirodája készít. A Termium Plus négy nyelven (angol, francia, portugál és spanyol) körülbelül 4,2 millió terminust tartalmaz, és a honlapon feltüntetett információk szerint havonta körülbelül 4.000 frissítést végeznek el rajta. A terminográfiai cédula bal oldalán a hasonló szavak listáját találjuk. Maga az adatlap párhuzamos oszlopos elrendezést mutat. A terminológiai adatbankban az EOHS Term tudásbázisban ekvivalensként feltüntetett angol nyelvű terminusra kerestünk rá: *training*. Ennek eredményeképpen egy terminográfiai cédulán összesen 17 számozott találatot kapunk, amelyen belül a felsorolt terminusok különböző fogalmakat jelölnek. Az adatlapon a 3-as számú találat felel meg a keresett fogalmat jelölő terminusnak. Az egyes találatok alatt következő adatmezők szerepelnek: tárgykör (Subject Field), definíció (DEF = Definition), megjegyzés (OBS = Observation) a szövegtí-

pusra vonatkozóan, kontextus (CONT = Context) és kollokáció (PHR = Phraseologism). Ezen felül találunk még nyelvtani jellemzőket, országkódot, a bejegyzés dátumát, és különböző címkék segítik az eligazodást például: elfogadottság. A megjelenítési felület megőrzi a szólista jelleget, a terminusautonómia elve nem érvényesül, így formailag inkább elektronikus szótárra emlékeztet (Tamás 2012).

Kutatásunk célja az volt, hogy megállapítsuk, milyen szempontok alapján különíthetők el egymástól az online szótárak és a terminológiai adatbázisok. Elsőként a szótárakról és a terminológiai adatbázisokról szóló szakirodalmat tekintettük át, majd konkrét példákat elemeztünk.

Arra a következtetésre jutottunk, hogy az elektronikus szótár és a terminológiai adatbázis között a határvonalak nem húzhatók meg minden esetben egyértelműen. Például míg az egyszerű adatbázisok esetében ugyan érvényesül a fogalomközpontúság elve, ugyanez nem mondható el a terminusautonómia elvéről (lásd Termium Plus). Gyakran megőrzik az elektronikus szótárakra emlékeztető szólista jelleget, az egyes bevitelek nem különülnek el élesen egymástól, azaz nem szerepelnek külön terminográfiai adatlapon. Szintén a terminusautonómia elvének mond ellent, ha a hagyományos terminológiai adatbázisoknak nemcsak fogalomközpontú, hanem szóközpontú megjelenítési felülete is létezik (Demeczky 2008). Az elektronikus szótárak pedig a nyelvi adatok kidolgozásakor nemcsak lexikográfiai, de terminológia módszerek szerint is készülhetnek, ahogyan a terminológiai adatok kidolgozásakor is előfordulhat, hogy elsődleges a nyelvi jel jelenti a kiindulási alapot és utána érvényesül a fogalomközpontú megközelítés. Vizsgálataink alapján kijelenthetjük, hogy léteznek ún. hibrid változatok.

IRODALOM

Arntz, R., Picht, H., Mayer, F. 2009. *Einführung in die Terminolgiearbeit*. Hildesheim, Zürich, New York: Georg Olms Verlag.

Brekke, M. 2001. LSP Lexicography and Terminography: A Complementary View. In: Mayer, F. (szerk.) 2001. *Language for Special Purposes. Perspectives for the New Millenium. Vol. 1., Linguistic and Cognitive Aspects, Knowledge Representation and Computational Linguistics, Terminology, Lexicography and Didactics.* Tübingen: Gunter Narr. 179–187.

Demeczky J. 2008. Terminológia a szoftveriparban. *Magyar Terminológia* 1. évf. 2. szám. 189–204.

Dringó-Horváth I. 2011. Hogyan válasszunk elektronikus szótárat a nyelvtanuláshoz? *Iskolakultúra* 2011. évf. 6–7. szám. 141–156.

Fóris Á., Rihmer Z. 2007. A szótárak minősítési kritériumairól. *Fordítástudomány* 9. évf. 1. szám. 109–113.

Fóris Á. 2007. A szótár terminus értelmezéséről. In: Magay T. (szerk.) 2007. *Félmúlt és közeljövő. (Lexikográfiai füzetek 3.)*, Budapest: Akadémiai Kiadó. 295–302.

Fóris Á. 2012. *Terminológiai szótárak és adatbázisok tipológiai kérdései.* Elhangzott: Társadalmi változások – nyelvi változások. Alkalmazott nyelvészeti kutatások a Kárpát-medencében. XXII. Magyar Alkalmazott Nyelvészeti Kongresszus. Szegedi Tudományegyetem, Szeged 2012. április 12−14.

Gaál P. 2010a. A hagyományos szótárak vizsgálati szempontjainak alkalmazhatósága az online szótárak esetében In: Boda I. K., Mónós K. (szerk.) 2010. *Az alkalmazott nyelvészet ma: innováció, technológia, tradíció. XX. Magyar Alkalmazott Nyelvészeti Kongresszus.* Budapest–Debrecen: MANYE–Debreceni Egyetem. 432–437.

Gaál P. 2010b. Online szótárak a web 2.0 platformon – A Wikiszótár és a Wictionary. *Magyar Terminológia* 3. évf. 2. szám. 279–296.

Pastor, V., Alcina, A. 2010. Search Techniques in Electronic Dictionaries: a Classification for Translators. *International Journal of Lexicography* 23. évf. 3. szám. 307–354.

de Schryver, G.-M. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16. évf. 2. szám. 143–199.

Sermann E. 2012. Spanyolországi terminológiai adatbázisok formai és tartalmi jellemzőinek vizsgálata. Előadás. Elhangzott: *Társadalmi változások – nyelvi változások. Alkalmazott nyelvészeti kutatások a Kárpát-medencében.* XXII. Magyar Alkalmazott Nyelvészeti Kongresszus. Szegedi Tudományegyetem, Szeged 2012. április 12−14.

Sermann E., Tamás D. 2010. Hogyan definiálhatjuk a fordítói adatbázist? Egy olasz és egy spanyol fordítói terminológiai adatbázis vizsgálata. In: Károly K., Fóris Á. (szerk.) 2010. *Nyelvek találkozása a fordításban. Doktori kutatások Klaudy Kinga tiszteletére.* Budapest: ELTE Eötvös Kiadó. 101–116.

Tamás D. 2010. *A gazdasági szakszövegek fordításának terminológiai kérdéseiről olasz-magyar nyelvpár esetében.* Doktori disszertáció. Kézirat. Budapest: ELTE.

Tamás D. 2012. Legyünk kreatívak: milyen is az igazi TB? Elhangzott: *„A magyar fordítók és tolmácsok napja”*. ELTE BTK FTT–OFFI Zrt., Budapest 2012. március 30.

FORRÁSOK

www.wordreference.com
www.eohsterm.org
www.termiumplus.gc.ca

ISO SZABVÁNYOK

ISO 12620:2009 Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources.
ISO 704:2009 Terminology work – Principles and methods.
ISO 12200:1999 Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange.
ISO 16642:2003 Computer applications in terminology – Terminological markup framework.