

SERMANN ESZTER

SZTE Idegennyelvi Kommunikációs Intézet, Szeged
sermann@lingo.u-szeged.hu

Spanyolországi terminológiai adatbázisok tartalmi és formai jellemzőinek vizsgálata

Kulcsszók: terminológiai adatbázis, terminusautonómia elve, Cercaterm, Neoloteca, UBTerm, UPCTERM, Euskalterm

A fordítói munka során rendkívül fontos a célnyelvi szöveg megfelelő terminológiai előkészítése, ehhez nyújtanak hatékony segítséget az online elérhető terminológiai adatbázisok. A terminológiai adatbázis elektronikusán tárolt terminológiai adatok gyűjteménye, amely lehet szó- vagy fogalomközpontú, egy vagy több tárgykör terminusait és a hozzájuk tartozó definíciókat tartalmazza egy vagy több nyelven, esetenként grammatikai, adminisztratív információkkal kiegészítve, és az adatok forrásának megjelölésével.

Spanyolországban számos terminológiai adatbázis létezik, amelyeket egyetemek, terminológiai szervezetek hoztak létre. Jelen esettanulmány célja hat spanyolországi (négy katalán, egy baszk és egy spanyol nyelvű) terminológiai adatbázis szerkezeti, tartalmi, formai jellemzőinek összehasonlító elemzése: 1. a *Cercaterm*, 2. a *Neoloteca*, a TERMCAT katalán terminológiai központ adatbázisai, 3. az *UBTerm*, a Barcelonai Egyetem terminológiai adatbázisa, 4. az *UPCTerm*, a Katalóniai Műszaki Egyetem adatbázisa, 5. az *Euskalterm*, a baszk nemzeti terminológiai adatbázis és 6. az *Oncoterm* a Granadai, a Valladolídi, a Málagai Egyetemnek és az Hospital Virgen de las Nieves kórháznak az adatbázisa. Az esettanulmány elkészítésének első lépéseként tanulmányoztam a terminológiai adatbázisokról szóló szakirodalmat (részletesen lásd Sermann, Tamás 2010, 2012), majd a Tamás (2009, 2010) által a Trieszti Tudományegyetem Szakfordító- és Tolmácsképző Intézet TERMit adatbázisának vizsgálatához és részletes elemzéséhez használt szempontrendszer néhány elemét alkalmaztam a hat spanyolországi adatbázisra. Tamás (ibid.) szempontjait figyelembe véve az alábbi szempontokat vizsgáltam: 1. *Általános jellemzők:* a célcsoport meghatározása (Kiknek készült az adatbázis?), az adatbázis nyelvei (Mi a fő nyelve? Milyen nyelvű ekvivalensek találhatóak?); 2. *Tartalmi jellemzők:* a terminusok/cédulák/fogalmak száma, tárgykörök, az adatok forrása, 3. *Formai jellemzők:* a terminográfiai cédula felépítése, a keresés módjai (például egyszerű, összetett). Az adatbázisok elemzése a felhasználói felületük vizsgálatát foglalta magában, informatikai aspektusai vizsgálatára nem terjedt ki.

Az esettanulmány második részének célja annak a kutatási kérdésnek a megválaszolása volt, hogy melyek hagyományos adatbázisok, melyek csak elnevezésükben azok, valójában inkább online szótár jellegűek. Az online szótárak és a terminológiai adatbázisok elkülönítésére – Fóris (2012) egyetértve abban, hogy a különböző típusú fordítói segédeszközök készítése és használata számos osztályozási kérdést vet fel – megkíséreltünk kidolgozni egy lehetséges szempontrendszert, amelyet *Online szótár vagy terminológiai adatbázis?* (Sermann, Tamás 2012) című előadásunkban ismertettünk. A hivatkozott adatbázisok elemzéséhez Sermann, Tamás (ibid.), az ISO szabványokban az adatbázisok szerkesztésére vonatkozó ajánlások (ISO 12620:2009) és Tamás (2012) alapján annak vizsgálatára, hogy az adatbázisok közül melyek hagyományos terminológiai adatbázisok, melyek inkább csak elnevezésükben azok, valójában inkább online szótár jellegűek, az alábbi szempontokat jelöltem ki:

- 1) Fogalom- vagy szóközpontú megjelenítés alapján rendezik az adatokat?
- 2) Érvényesül-e a terminusautonómia elve?
- 3) Szerepel-e definíció minden terminográfiai cédulán?
- 4) Meg van-e jelölve az adatok forrása minden cédulán?
- 5) Szerepel-e az adatok felvételének dátuma minden cédulán?

A tanulmány további részében bemutatom az adatbázisok tartalmi és formai jellemzőit, majd összegzem a kutatás második részének eredményeit, és javaslatot teszek egy ajánlható modellre, amely a vizsgált adatbázisok közül a legalkalmasabb a szabványosított terminológiai adatok elhelyezésére.

A *Cercaterm* a TERMCAT által szerkesztett adatbázis különböző tárgykörök szakemberei, kiadványszerkesztők, valamint fordítók és tolmácsok számára készült. Az adatbázis nyelve katalán, a fő terminus katalán nyelvű, ekvivalens formák – nem minden esetben – spanyol, francia angol és olasz nyelven is szerepelnek. Tartalmi jellemzőit tekintve 230.000 terminográfiai cédulát és 31 tárgykört tartalmaz a mezőgazdaságtól az elektronikáig, a földrajztól a nyelvészet és irodalom területéig számos tudományos és műszaki tárgykör terminusai és definíciói között tájékozódhatunk. Kéthetente frissítik az alábbi források alapján: 1. a *Neoloteca*, a TERMCAT adatbázisa, amely a szervezet Felügyelő Tanácsa (*Consejo Supervisor*) által szabványosított terminusokat tartalmazza, 2. a TERMCAT által kidolgozott terminológiai termékek, 3. a TERMCAT kutatási projektjeinek eredményei, 4. egyéb szervezetek adatbázisaiból átvett terminográfiai cédulák. Minden cédula jobb felső sarkában a *Font* ikonra kattintva megtaláljuk a terminus és a definíció forrásának feltüntetését.

A terminográfiai cédula minden esetben tartalmazza a fogalom definícióját és a katalán terminust, a hozzá tartozó nyelvtani kategóriával, a tárgykört vagy tárgyköröket, amelyekhez a terminus tartozik és változó számú idegen nyelvű ekvivalenst ISO-kóddal (nyelvi azonosítóval) ellátva. A *Cercaterm* esetében kétféle keresési módot találunk: egyszerű és összetett.

A *Neoloteca* szintén a TERMCAT adatbázisa, amely a szervezet Felügyelő Bizottsága által szabványosított összes katalán (műszaki és tudományos tárgykörökbe tartozó) terminust tartalmazza a hozzájuk kapcsolódó katalán nyelvű definíciókkal és egyéb nyelvű ekvivalensekkel együtt. Elsősorban olyan szakemberek számára készült, akiknek katalán nyelven kell kiadványokat szerkeszteniük, kiadóknak, valamint fordítóknak, tolmácsoknak.

Főként katalán nyelvű terminológiai adatokról nyújt információt, de számos cédulán megtalálhatóak egyéb nyelvű referenciamunkákban (forrásokban) található spanyol, angol, német, francia, olasz ekvivalens formák. Tartalmi jellemzőit tekintve a *Neoloteca*-ban katalán, spanyol, angol, német, olasz és francia alfabetikus tárgymutató útján lehet hozzáférni a TERMCAT Felügyelő Tanácsa által szabványosított terminusokhoz, valamint latin nyelvű némenklatúrákhoz. 4500 katalán terminust tartalmaz, 29 tárgykör terminusait és definícióit találjuk meg a legkülönbözőbb tárgykörökben (pl. egészség tudomány, társadalomtudományok, sport, fizika). A cédulákon nem tüntetik fel egyenként a terminológiai adatok forrását, mivel a cédulákat a Felügyelő Tanács döntései alapján dolgozzák ki. Ha a formai jellemzőket vizsgáljuk meg, akkor azt látjuk, hogy a *Neoloteca* adatbázis felhasználói felülete és a cédulák felépítése a *Cercaterm*-ével azonos.

Az *UBTerm* a Barcelonai Egyetem terminológiai adatbázisa, az egyetemi közösség és a Nyelvi Szolgálat (*Serveis Lingüístics*) által publikált, már megjelent összes szótár, glosszárrium terminusait tartalmazza, az egyetemhez kapcsolódó tevékenységek során használatos terminusokat gyűjti egybe (oktatás, kutatás). Elsősorban egyetemi oktatók, hallgatók, műszaki és tudományos szövegek fordítói, szerkesztői, valamint szaknyelvet oktató tanárok számára készült. Az adatbázis nyelve katalán, minden cédulán a katalán terminus szerepel fő terminusként, ebből juthatunk el a hozzá kapcsolódó egyéb információkhoz, minden cédula tartalmaz egyéb nyelvű ekvivalens terminusokat is (spanyol, angol, francia, német, olasz), és a cédulákhoz kapcsolódó egyes információkhoz az egyes ekvivalens terminusokra kattintva is hozzáférhetünk. Tartalmát tekintve az adatbázis ismertetője nem szolgáltat információt a terminusok vagy a cédulák számáról. A terminusok besorolása alfabetikus vagy kereshetünk az egyes szótárak tartalmán belül, így témakör szerint összesen 37 műszaki és tudományos tárgykörben (pl. anatómia, biokémia, ökológia, statisztika, fizika stb.). A cédulán legfelül megtaláljuk az adott cédulán szereplő adatok forrását (a Nyelvi Szolgálat egyik kiadványának címét), és a forrásra kattintva annak bibliográfiai adatait. Formai jellemzőit megvizsgálva azt látjuk, hogy a cédulán elsőként szerepel a fő terminus katalán nyelven, majd szinonimák, egyéb nyelvű ekvivalensek, a grammatikai kategória megjelölése, a definíció és egyéb megjegyzések következnek. Egyszerű vagy összetett keresést választhatunk.

Az *UPCTERM* műszaki, technológiai tárgykörű többnyelvű terminológiai adatbázis, amelyet a Katalóniai Műszaki Egyetem Nyelvtechnológiai Szolgálat (*Servicios de Lengua y Tecnología de la Universidad Politécnica de Cataluña*) dolgozott ki. Az egyetem különböző egységeinek munkatársai közösen hozták létre, katalán nyelvű, esetenként spanyol, angol és francia ekvivalensekkel. Tartalmi jegyeit tekintve a terminusok besorolása ábécérendű, de végezhetünk tematikus keresést is, amelynek eredményeképpen a kiválasztott tárgykörhöz tartozó összes terminust megtaláljuk ábécérendben. 92.408 terminust tartalmaz 28.115 terminográfiai cédulán, 14 tárgykörben pl. matematika, statisztika, építészet, optika, gépészet, elektronika, informatika, telekommunikáció, földtan, textilipar, gazdaság, vegyészet. A honlapon található információk szerint a terminológiai adatok forrását főként a

nyelvtchnológiai szolgálat kiadványai, valamint a műszaki szövegek fordítási projektjei során kinyert terminusok adják, viszont az egyes cédulákon a forrásra vonatkozó adatokat nem találunk. A terminográfiai cédulán elsőként a fő terminust találjuk a nyelvi azonosítóval és a tárgykör megjelölésével, majd az egyéb nyelvű ekvivalenseket szintén nyelvi azonosítókkal ellátva. Grammatikai információ (a nyelvtani nem feltüntetése) minden esetben csak a katalán terminushoz tartozik. Esetenként a cédulán megtalálható a katalán nyelvű definíció, a definícióhoz fűzött megjegyzések, illetve katalán szinonimák. Kétféle keresési mód áll a rendelkezésünkre: ábécérendű vagy interaktív.

Az *Euskalterm* a Baszk Terminológiai és Lexikográfiai Központ (UZEI) által létrehozott terminológiai adatbázis. A szervezet több mint 80 terminológiai szótárt szerkesztett sokféle tárgykörben, ezek fúziójából hozták létre 1987-ben az adatbázist, ebbe integrálták a később szerkesztett szótárak terminológiai adatait, 2001-ben pedig az *Euskalterm*et baszk nemzeti terminológiai adatbázissá nyilvánították. Az adatbázis elsősorban baszk anyanyelvű fordítók és tolmácsok számára készült, felülete spanyol, baszk, angol vagy francia nyelven tanulmányozható, a baszk fő terminusok mellett spanyol, francia, angol és latin ekvivalenseket találunk. Tartalmi jellemzőit tekintve a terminusok besorolása kizárólag alfabetikus, témakörök szerint nem tudunk keresni, több mint 100 ezer terminográfiai cédulát tartalmaz 27 tárgykörben (pl. kereskedelem, jogtudomány, sport, informatika stb.). A terminológiai cédulákat rendszeresen frissítik az *Euskaltzaindia* (a Baszk Nyelv Királyi Akadémiája) normái szerint és a terminológiai szabványosító műszaki bizottságok munkája nyomán. A formai jellemzőket megvizsgálva azt látjuk, hogy a terminográfiai cédulák nyelve a baszk, a cédulán legfelül a tárgyköri besorolást találjuk, majd a baszk terminus következik grammatikai információk nélkül, ezt követi a baszk nyelvű definíció, majd az egyéb nyelvű ekvivalensek, legalul pedig az adatok forrásának megjelölését, és a rögzítés évét találjuk. Az adatbázis csak egyszerű keresési módot kínál, beírhatjuk a keresett elemet, valamint kiválaszthatjuk a nyelvet.

Az *Oncoterm* adatbázis nyelvészek és egészségügyi dolgozók együttműködéséből született, angol–spanyol nyelvű, és nemcsak fordítóknak és szakszövegíróknak szánták, hanem egészségügyi dolgozóknak, betegeknek, illetve családtagjaiknak is. A terminusok besorolása ábécérendű, az adatbázis 1896 onkológiával, onkoterápiával kapcsolatos fogalmat és 4033 terminust tartalmaz: betegségeket, gyógyszereket, kezelési módokat, és az ezekhez kapcsolódó fogalmakat, amelyek hiperlinkek útján is összekapcsolódnak egymással, egységes fogalmi rendszert alkotnak. Az *Oncoterm* egy kb. 32 millió angol és spanyol szövegszóból álló korpusz alapján született, amelyet internetről származó dokumentumokból, CD-ROM formátumú orvosi enciklopédiákból, kézikönyvekből, rákbetegséggel foglalkozó nemzetközi szervezetek honlapjairól származó általános közönségnek szánt informatív anyagokból állították össze. Ha belépünk az *Oncoterm* honlapjára, bal oldalon találjuk az angol nyelvű terminusok listáját ábécérendben, és az egyes terminusokra kattintva a következőket látjuk: a terminográfiai cédula címe a fő terminus angol nyelven, ezt követi a tárgykör (*subject field*), amelyhez a terminussal jelölt fogalom tartozik, az adminisztratív adatok, majd egyes esetekben olyan honlapok elérhetőségét is feltüntetik, amelyek segítik a fogalom megértését, esetenként képeket is találunk. Majd a fogalmi struktúrák (*conceptual structures*) elnevezésű táblázat következik, ahol megtalálhatjuk a fogalmi címkét (*is a*), a hozzá kapcsolódó alkategóriákat, illetve azokat a fogalmakat, amelyekhez a terminus kapcsolódik (*ancestors, descendants*). Ezután következik a spanyol, illetve angol nyelvű definíció, majd a fogalom jelölőire vonatkozó terminológiai adatok (a terminus típusa, nyelvtani kategóriája, valamint megbízhatósági kódja). A 2002-ben készült terminológiai adatbázist elkészülése óta nem frissítették, csak egyszerű keresés alapján böngészhetünk.

Vizsgálatom célja – az adatbázisok tartalmi és formai vizsgálatán túl – annak a kutatási kérdésnek a megválaszolása volt, hogy a hat adatbázis közül melyek hagyományosak, melyek csak elnevezésükben azok, valójában inkább online szótár jellegűek. Ennek megállapítására öt szempontot jelöltem ki, melyek közül az első a fogalomközpontú megjelenítés. A terminológiai adatbázisok vizsgálatának alapján azt állapítom meg, hogy fogalomközpontú megjelenítésről csak az *Oncoterm* esetében beszélhetünk, ahol a terminográfiai cédula címe az angol nyelvű fő terminus, a fogalmi struktúrák is fel vannak tüntetve, a terminográfiai cédula központi eleme a fogalom, amelyet angolul és spanyolul definiáltak, és hozzárendelték a terminusokat két nyelven. A többi adatbázis felülete inkább online szótár jellegű, ahol a központi elem a címszó, ehhez rendelték hozzá a többi adatot. A második szempont a terminusautonómia elve volt, amely szerint a terminusok önálló szerepeltetésére kell törekedni, azaz egy fő terminust kell kiválasztani, és a szinonimákat külön bevételként kell kezelni. Annak megítéléséhez, hogy a terminusautonómia elve valóban érvényesül-e, az adatbázis szinte összes terminográfiai céduláját meg kellene

vizsgálni. Jelen kutatás nem terjed ki az adatbázisok minden cédulájának vizsgálatára, így azt, hogy a terminusautonómia elve érvényesül-e, nem minden esetben lehet teljes bizonyossággal megállapítani. A harmadik szempont a definíció jelenléte volt minden cédulán, amely álláspontom szerint szoros összefüggésben áll a terminusautonómia elvével: azokban az adatbázisokban, ahol nem szerepel definíció minden cédulán, nem érvényesülhet a terminusautonómia elve, mivel maga a fogalom sem jelenik meg, holott a fogalomalapú adatbázisokban a fogalmat a definíció jeleníti meg, a definíció hiánya pedig a fogalom pontos körülhatárolásának hiányáról árulkodik. A vizsgált adatbázisok esetében a terminusautonómia elve a *Neoloteca* és az *Oncoterm* esetében valósul meg, ahol minden fogalom definiálva van, és egy fogalomhoz egy terminus tartozik. A többi adatbázis esetében csak részben valósul meg, mivel egyes esetekben megjelenik a fogalom definíció formájában, és egyetlen terminus tartozik hozzá, más esetekben pedig a definíció hiányából fakadóan a terminusautonómia elve nem valósul meg.

Az ISO 12620-as szabványa, amely a terminológiai adatbázisok adatkategóriáit határozza meg, az egyes terminográfiai cédulákra vonatkozó minimális követelményként a fő terminus szerepeltetése mellett a forrás megjelölését és az adatok felvételének dátumát adja meg, így a vizsgálathoz negyedik és ötödik szempontként ezeket jelöltem ki. A *Cercaterm*, az *UBTerm* és az *Euskalterm* esetében minden cédulán szerepel a forrás, amely az adatbázist létrehozó szervezet egyik kiadványa, szótára. A *Neolotecában* külön nem szerepel a forrás, mivel a TERMCAT Felügyelő Tanácsa által konszenzus útján szabványosított terminusokat tartalmazza. Az *Oncoterm*ben legtöbb esetben feltüntetik a forrást, hiperlinkek útján juthatunk el a terminológiai adatok forrásáig, de találunk olyan cédulákat is, ahol nem szerepel. Az adatok felvételének pontos dátuma csak az *Oncoterm* esetében van feltüntetve minden cédulán.

Összegzésképpen elmondható, hogy a vizsgált adatbázisok közül 4 (*Cercaterm*, *UBTerm*, *UPC Term*, *Euskalterm*) cédulái szóközpontú megjelenítésűek, ezek inkább online szótár jelleget tükröznek, és nem rendelkeznek a klasszikus, fogalomalapú terminológiai adatbázisok jellemzőivel. Az *Oncoterm* fogalomközpontúnak mondható a megjelenítés szintjén: az adatbázis központi eleme a fogalom, amelynek megjelenési formája a definíció, egy fogalomhoz pedig egy terminus és egy terminográfiai cédula rendelhető hozzá. A *Neolotecában* a cédula megjelenítése szóközpontú, a cédula fő eleme a címszó, ehhez rendelik hozzá az egyéb információkat. Viszont az adatbázis leírásából az derül ki, hogy a fogalmakat szisztematikusan definiálták, majd hozzájuk rendeltek a preferált terminusokat, amely a fogalomalapú adatbázisok szerkesztésére jellemző. Hiányoznak azonban a fogalmi struktúrák, amelyek hiányában azt feltételezhetjük, hogy a fogalmi rendszerek pontos felvázolása nem történt meg. A *Neoloteca* tehát részben rendelkezik a klasszikus terminológiai adatbázisok jellemzőivel. Ajánlható modellként az *Oncoterm*et emelem ki, mivel az ilyen típusú adatbázisok a legalkalmasabbak az szabványosított terminológiai adatok elhelyezésére: a cédula tartalmazza az idézett ISO szabvány által minimálisan javasolt adatkategóriákat, valamint a cédula alapján pontosan meghatározhatjuk a terminus helyét a fogalmi rendszeren belül.

IRODALOM

- Fóris Á. 2012. Terminológiai szótárak és adatbázisok. 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.
- Sermann E., Tamás D. 2012. Online szótár vagy terminológiai adatbázis? Elhangzott: *Társadalmi változások – nyelvi változások. Alkalmazott nyelvészeti kutatások a Kárpát-medencében*. XXII. Magyar Alkalmazott Nyelvészeti Kongresszus. Szegedi Tudományegyetem, Szeged 2012. április 12–14.
- Tamás D. 2009. A trieszti TERMit adatbázis vizsgálata. *Magyar Terminológia* 2. évf. 2. szám. 213–236.
- Tamás D. 2010. *A gazdasági szakszövegek fordításának terminológiai kérdéseiről olasz-magyar nyelvpár esetében*. Doktori disszertáció. 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

Cercaterm [<http://www.termcat.cat/> – 2012.02.20.]
Neoloteca [<http://www.termcat.cat/> – 2012.02.20.]
UBTerm [http://www.ub.edu/slc/ubterm/td_Arrencada.html – 2012.02.20.]
UPCTerm [<http://www.upc.edu/sit/upcterm/> – 2012.02.20.]
Euskalterm [<http://www1.euskadi.net/euskalterm/> – 2012.02.20.]
Oncoterm [<http://www.ugr.es/~oncoterm/> – 2012.02.20.]

ISO SZABVÁNY

ISO 12620:2009 Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources.