

# Legyünk kreatívak: milyen is az igazi terminológiai adatbázis?

Tamás Dóra

E-mail: [tamas.dora@offi.hu](mailto:tamas.dora@offi.hu)

**Kivonat:** A terminológiai adatbázisokat napjainkban a hatékony fordítói segédeszközök között tartjuk számon. De vajon milyen mélységig ismerjük ezeket a segédeszközöket, mennyire hasznosítjuk a bennük rejlő lehetőségeket? A tanulmány szerzője konkrét példák elemzésével kísérletet tesz a terminológiai adatbázisok hármass felosztására. A megjelenítési felület alapján megvizsgálja felépítésüket, az adatmezők típusait, az általuk tartalmazott adatmennyiséget és az egyéb kapcsolódó elemeket, majd ismerteti az egyszerű, a hagyományos és az összetett terminológiai adatbázisok alapvető jellemzőit. A terminológiai adatbázisok szerkesztése egyfelől megkívánja, hogy ismerjük a terminológia elveit és módszereit, másfelől szükséges elrugaszkodni a hagyományos, nyomtatott segédeszközökre jellemző szemlélettől és élni a szoftverek kínálta lehetőségekkel. A tanulmány kitér a terminológus összetett feladataira is, amelyeket a modern szemléletet és igényeket követve mutat be.

**Kulcsszavak:** terminológiai adatbázis, adatbank, tudásbázis, adatmezők, fogalomközpontúság, terminusautonómia

## 1. Bevezetés<sup>1</sup>

A szakfordítók egyre gyakrabban élnek munkájuk során az elektronikus fordítástámogató segédeszközök alkalmazásának lehetőségével. Ezek a szoftverek általában két alapvető funkció ellátására törekszenek, azonban előfordul, hogy a szakfordítók csak az egyiket, az ún. fordítási memóriát (Translation Memory, rövid.: TM) hasznosítják, miközben a terminológiai adatbázisról (Termbase, rövid.: TB) akár meg is feledkeznek. Nem véletlen tehát, hogy tanulmányom fő tárgyát ez utóbbi képezi.

Azonban e „feledékenység” nem róható fel kizárólagosan a fordítónak, hiszen számos oka lehet. A fordító általában szűk időkerettel dolgozik, a terminusok ekvivalenseinek megkeresése gyakran időigényes tevékenység, sőt, bizonyos szövegtípusoknál még az is előfordulhat, hogy a terminológiai keresés akár a fordítási idő 40%-át is kiteszi. Eközben nemcsak nincs idő minden nyelvi adat ellenőrzésére és rögzítésére, hanem sokszor hiányzik az a szakember is, aki a nyelvi adatokat szakmai szempontból ellenőrizné. Hiányozhat a megfelelő szoftverkörnyezet is, vagy a fordító nem használja ki az adott fordítástámogató szoft-

---

<sup>1</sup> Jelen tanulmány *A magyar fordítók és tolmácsok napja* című, az ELTE BTK FTT és az OFFI Zrt. által közösen rendezett konferencián elhangzott előadásom írott változata.

veren belüli lehetőségeket sem (például idejéből legfeljebb két- vagy háromszlopos glosszárrium készítésére futja, amelybe elsősorban címszavak és néhány megjegyzés, pontosítás kerül). További gyakori nehézség, hogy a fordító sokszor nem rendelkezik megfelelő terminológiai ismeretekkel, nem ismeri a modern terminológiát mint diszciplínát, sem annak modern elveit és módszereit.

Éppen a felsoroltak miatt az adatbázisok leginkább csapatmunkában valósíthatók meg. Nézzük, hogy melyek a terminológiai adatbázisok készítésének legfontosabb hozzávalói, hiszen nélkülözhetetlen a megfelelő eszköz és koordináló személy kiválasztása, ám ne feledkezzünk meg a lehetőségek optimális kihasználása érdekében a kreativitásról sem.

## 2. A terminológiai adatbázis mint fordítói segédeszköz

A terminológiai adatbázisok olyan, elektronikusan tárolt terminológiai adatok gyűjteményei, amelyek az onomasziológiai szemléletet követve készülnek, és egy vagy több nyelven egy vagy több szakterület terminusait tartalmazzák a hozzájuk tartozó definíciókkal együtt (Sermann és Tamás 2010). Az elektronikus tárolás előnyei közé tartozik az adatok egyszerű és célzott frissítésének a lehetősége, az online, számos esetben ingyenes elérhetőség, a megbízhatóság. A terminológiai munka az utóbbi két évtizedben alapvető változásokon ment keresztül: „a terminológiai adatbázisok megjelenése, és a mindenki számára lehetséges nyilvános elérhetőségük alapjaiban változtatták meg a terminológiát” (Fóris 2005: 22).

A terminológiai adatbázisok nem tévesztendőek össze az elektronikus szótárakkal, amelyek központi eleme a szó, és általában szemasziológiai módszerek alapján készülnek. Ezzel ellentétben a terminológiai adatbázis készítésének alapja az onomasziológiai módszer, és központi eleme a terminus által jelölt fogalom (Cabré 1998). Ettől függetlenül a határvonalak nem mindig húzhatók meg élesen, és előfordulhatnak hibrid változatok is (Sermann és Tamás 2012). Az úgynevezett számítógépes ontológiák felhasználása a terminológiai munkában a tudásbázisok létrejöttéhez vezetett.

A nagy mennyiségű információ között való eligazodást, és az információk minél hatékonyabb felhasználását elősegíti, ha egy adatbázisban nemcsak a fogalom meghatározását találja meg a felhasználó, hanem az arra vonatkozó adatok összességét is, hogy a fogalom milyen kapcsolatrendszerben fordul elő, és e kapcsolatrendszeret mi jellemzi. (Fóris 2005: 22)

Speciális terminológiai adatbázisok készülnek fordítók számára, ezek az úgynevezett *fordítói terminológiai adatbázisok*; ezek közül néhánynak – amelyek terminológiai kutatóintézeteknél, egyetemeken készültek – ismertetem a jellemzőit a 2. pontban.

Maguk a fordítóirodák, fordítással foglalkozó cégek is készíthetnek fordítói terminológiai adatbázist, akár nyilvánosan megvásárolható szoftverek segítsé-

gével – de e tanulmány tárgyát nem ezek a célcsoport igényei szerint rugalmasan alakítható szoftverek képezik<sup>2</sup>. Ezek ugyan fontos eszközök, azonban nem mindegy, milyen adatokkal töltjük fel őket.

Bármilyen adat rögzítése előtt fontos meghatározni, hogyan strukturáljuk az adatokat, és a bevitt adatok mennyire megbízhatók. A terminológiai adatbázisoknak három alapvető típusát különböztetem meg struktúrájuk és a bevitt adatok mennyisége alapján:

- (1) egyszerű terminológiai adatbázis, illetve adatbank;
- (2) hagyományos (vagy klasszikus) terminológiai adatbázis, illetve adatbank;
- (3) összetett terminológiai adatbázis (fajtái: tudásbázis, terminológiai információs rendszer).

E három csoportba sorolom be az elemzett nyilvános fordítói terminológiai adatbázisokat, ebben a keretben elemzem részletesen ezeket. Célom, hogy a teljesség igénye nélkül, rövid körképet adjak e témáról, és az elemzett adatbázisokat – az egyszerűtől az egyre összetettebb változatok felé haladva – egyfajta skálán elhelyezve mutassam be.

## 2.1. A terminológiai adatbázis egyszerű változatai

### 2.1.1. A *Termin*

A *Termin* ([external.kim.gov.hu](http://external.kim.gov.hu)) a Közigazgatási és Igazságügyi Minisztérium terminológiai adatbázisa. A *Termin* létrehozásának eredeti célja az volt, hogy az európai uniós joganyag magyarra fordításának támogatásával elősegítse Magyarországnak az Európai Unióhoz való csatlakozását. 2004. május 1-jét követően ez a feladat lezárult, és a terminológiakezelés feladata átkerült az Európai Unió intézményeihez.

A *Termin* négy (magyar, angol, francia, német) nyelven tartalmaz bejegyzéseket, és többféle keresési módot kínál: rákereshetünk a teljes kifejezésre, a kifejezés elejére, valamilyen tetszőleges részre a kifejezésen belül, rövidítésre, szakterületre vagy hasonló terminusok listája alapján válogathatunk. A munka nem zárult le teljesen, az újonnan rögzített bejegyzések elsődleges jóváhagyója az Európai Uniós Jogi Főosztály. Előnyei közé tartozik, hogy folyamatosan bővül a magyar minisztériumok és állami szervek által jóváhagyott kifejezésekkel. Az új adatfelvitelek közé tartoznak a beosztások, tisztségek és intézménynevek terminusai is. Ezek valóban speciális területet képeznek, mert az országspecifikus intézménynevek, a foglalkozásokat, szolgáltatásokat és okiratokat jelölő terminusok esetében ritka a terminusok 100%-os ekvivalenciája, emiatt külön figyelmet érdemelnek (Castagnoli 2003; 2006, Bullo 2009).

<sup>2</sup> SDL Trados MultiTerm, qTerm, Déjà Vu X2, Webstar, Cross Term, MultiTrans, TermStar<sup>NXT</sup>, LookUP, flashterm®, termXplorer, quickTerm, Tipp-Term®

Az adatbázis számos szakterületet ölel fel: belügy, rendészet, biztonságpolitika, biztosítás, egészségügy, földrajzi nevek, energia, hírközlés, ipar, jog, kereskedelem, pénzügy, mezőgazdaság. Ezen felül az adatbázisból elérhető a minisztérium magyar államigazgatási és a Szellemi Tulajdon Nemzeti Hivatalának szellemi tulajdoni kifejezéstára.

Ha például rákeresünk a *nemzetgazdasági miniszter* terminusra, azt láthatjuk, hogy a felhasználói felület megjelenítési módja egyszerű, táblázatos forma, amelyben a terminusokon kívül a nyelvi kódok, a szakterület és a „kötet” elnevezés alatt a szakterület megjelölése szerepel. Az ún. uniós terminusok esetében a „kapcsolódó uniós aktus” elnevezésű adatmező alatt kódszám is szerepel, amely hiperlink révén átirányít az EUR-LEX webes felületére (<http://eur-lex.europa.eu>), összekötve ezzel a két adatbázis adatait.

### 2.1.2. *A Terminum Plus*

A *Terminum Plus* (<http://www.terminumplus.gc.ca>) a kanadai kormány hivatalos terminológiai adatbankja (*The Government of Canada's terminology and linguistic databank*), amelyet megrendelésre a kanadai kormány – az Országos Fordító és Fordításhitelesítő Iroda Zrt. tevékenységéhez hasonló feladatokat ellátó – központi fordítóirodája (*Translation Bureau*) készít. A *Terminum Plus* négy – angol, francia, portugál és spanyol – nyelven körülbelül 4,2 millió terminust tartalmaz, és a honlapon leírtak szerint havonta megközelítőleg 4 000 frissítést végeznek el benne. Többféle keresési módot kínál: a kulcsszavas, a tárgykör szerinti keresést, és – akárcsak a *Termin* – felkínálja a hasonló szavak listáját.

A párhuzamos oszlopos elrendezést mutató, ún. terminográfiai cédulák (más néven: adatlapok) legfontosabb, általában rövidítéssel jelölt elemei a következők: tárgykör (Subject Field), terminus, definíció (DEF = Definition), megjegyzés (OBS = Observation), valamint szerepel kontextus (CONT = Context) és kollokáció (PHR = Phraseologism) is. A megjegyzés adatmezőben például azt jelölik, hogy a terminus használata milyen szövegtípusra jellemző (például: „OBS – Terminology used by Passport Canada”). Ezen felül találunk még nyelvtani jellemzőket (többes számú használat, nyelvtani nem), a használat szerinti ország megjelölését (például *Great Britain*), és feltüntetik a bejegyzés dátumát. Különböző címkék is segítik az eligazodást, hogy csak néhányat említsünk: elfogadottság (ezen belül: kerülendő, szabványosított, a Terminológiai Bizottság által hivatalosan elfogadott) vagy időbeli címke, amely az archaikus szóhasználatot is jelölheti.

A megjelenítési felület formailag megőrzi a szólista jelleget, így inkább elektronikus szótárra emlékeztet, egy-egy bejegyzés alatt viszonylag kevés információt tárol, és sok helyen nincs feltöltve a minimális adatmennyiség sem, mint például a definíció, a kontextus vagy a megjegyzés. Nagyon fontos viszont, hogy szerepel az adatok megbízhatóságát növelő forrás, amely hiperhivatkozással külön adatlapon érhető el.

### 2.1.3. A TERMDAT

A svájci közigazgatási hivatalok a TERMDAT ([www.termdat.ch](http://www.termdat.ch)) elnevezésű projekt alapjait még a messzi 1987-ben, egy, az Európai Közösségek és Svájc között létrejött együttműködési megállapodás alapján fektették le. Ennek az együttműködésnek köszönhetően a TERMDAT eredeti struktúrája az Európai Unió intézményközi terminológiai adatbázisának (IATE) egyik elődje, az EUODICATOM felépítésével egyezett meg, és számos adatcserére is sor került. A közös munka azonban közel hat éve megszűnt.

A TERMDAT felügyeletét a Svájci Szövetségi Kancellária (mint közigazgatási hivatal) Terminológiai Szekciója (Sektion Terminologie der Schweizerischen Bundeskanzlei/Sezione di terminologia della Cancelleria Federale) látja el. A Svájci Államszövetség négy hivatalos nyelvén (német, francia, olasz és rétoromán) és angol nyelven készült adatbázis elsősorban jogi és közigazgatási terminusokat tartalmaz, és eredetileg azzal a céllal hozták létre, hogy a közigazgatásban dolgozók munkáját támogassák vele. A terminológiai információk elsősorban jogszabályokból származnak, és különböző területeket ölelnek fel: oktatás, energia, pénzügy, kereskedelem, informatika, orvostudomány, katonaság, ipar, kultúra, mezőgazdaság, környezetvédelem, közlekedés, gazdaság, politika, műszaki terminusok stb., de a hangsúly a szövetségi jogi és államigazgatási terminusok kezelésén van. Ez utóbbiba a nemzetközi és szövetségi intézmények megnevezései és rövidítései is beletartoznak.

A folyamatosan frissített és bővített adatbázis validált adatai 2009-től az interneten is hozzáférhetők. A honlapon közzétett adatok szerint az adatbázis kb. 1,5 millió bejegyzést tartalmaz, amelyből 400 000 kifejezetten svájci vonatkozású. A TERMDAT internetes és intranetes verziója nem egyezik meg teljesen: az intranetes változat két felületből és adatbázisból áll. A TERMDAT/EDIC a validált nyelvi adatokat, míg a TERMDAT/MODS az erre a célra tartott tréningen részt vett és szerkesztésre jogosult személyek nem validált bejegyzéseit tartalmazza.

A kancellária nyelvi központjában a fordítók a jogászokkal karöltve dolgoznak, majd a lefordított és ellenőrzött szövegek a terminológusokhoz kerülnek. A terminusok kézi kivonatolását követően a terminológusok előkészítik a terminográfiai adatlapokat, majd azokat szakértő (jogász) részvételével, megbízható források felhasználásával terminológiai munkacsoportokban megvitatják. Csak a többszintű validálást követően kerülhetnek be a terminusok a végleges terminológiai adatbázisba. A nyelvi adatok megbízhatóságát kódokkal is jelzik. A validálást követően a megbízhatósági kód az 1-től 5-ig terjedő skálán a 4-es jelölést kapja, mert az 5-ös kódot a szabványosított terminusok számára tartják fenn. Külön érdekesség, hogy terminológiai munka nemcsak a központban, hanem az egyes kantonokban is folyik, ahonnan az adatokat szintén a központi adatbázisba töltik fel. Ezzel egyfelől elismerik a kantonok autonómiáját, illetve az adatbázison belül a bejegyzéseket többek közt a számontartás végett sorolják a különböző kantonokhoz.

A terminográfiai cédula három fő részből áll: bal oldalt a hasonló szavak listája, középen felül egy azonosító adatokat tartalmazó fejléc szerepel, alatta pedig a különböző nyelvek szerint besorolt nyelvi adatok vízszintes sorokba

rendezve található meg. Külön útmutató segíti a számos rövidítés pontos értelmezését; ezek a következőket jelölik, különválasztva a fejlécben és az alatta megjelenő adatokat:

- a fejlécben szerepel az illetékes terminológiai hivatal (BE), a szószedet (TY) és az ahhoz kapcsolódó szószedet (NB), a fogalom azonosítószáma (NI), a szakterület (CM), a szerző (AU) és a megbízhatósági kód (CF);
- alatta található a nyelvi adatok: a terminus és a szinonima (VE), a rövidítés (AB), a kulcsszó (MC), a definíció (DF), a kontextus (PH), a megjegyzés (NT), mint például regionális használat (kanton jelölése), az országkód (PS) és a forrás (RF), amely lehet jogszabályi hivatkozás is.

Bár az adatbázisban számos információ fellelhető, azonban a fogalomközpontúságra épülő terminológiai adatbázisok központi kategóriája, a definíció nem minden esetben, legtöbbször csak egy rövid mondat formájában, tömör meghatározásként jelenik meg, éppen ezért az internetes megjelenítési felület alapján a TERMDAT-ot az egyszerű adatbázisok kategóriájába soroltam.

Érdeemes azonban megemlíteni, hogy a terminológiai munkának Svájcban igen nagy hagyománya van. A közigazgatáson belüli összehangolt együttműködésnek köszönhetően a jogszabályok hatálybalépésekor interneten nemcsak a jogszabályok szövegeinek különböző nyelvi változatai lelhetők fel, hanem abban az időpontban már a TERMDAT adatbázisban a terminusok is elérhetők és tanulmányozhatók.

## 2.2. A terminológiai adatbázisok hagyományos változata

Ezen a ponton jogosan felvetődhet a kérdés, hogy egy terminológiai adatbázisnak vajon milyen adatokat kell minimálisan tartalmaznia. A terminográfiai cédula felépítését pontosan ki kell dolgozni, hogy csak néhány érvet említsünk:

- ha meghatározatlan struktúrájú adathalmazzal rendelkezünk: az lehetetlenné teszi a keresést;
- ha nem rögzítik az adatbevitel dátumát: nincs információ a naprakészségről;
- ha hiányoznak a források és a jóváhagyás: kétes információkkal rendelkezünk, amelyek megbízhatósága nem ellenőrizhető.

Érdeemes ezen felül elrugaszkodni a hagyományos, nyomtatott segédeszközökre jellemző gondolkodásmódtól, és a webes felület adta lehetőségeket maximálisan hasznosítani.

A terminológiai adatbázisok építéséhez léteznek szabványok (12620:1999; 12620:2009), illetve kutatók is foglalkoztak a kérdéssel (Arntz et al. 2009: 223, Schmitz 2011: 238). Arntz et al. (ibid.) egy sematikus ábrát is mellékel (1. ábra), amely az egyszerű terminológiai adatbázis terminográfiai cédulájának alapvető leíró mezőit (adatmezőit) tartalmazza, mint például: szakterület, nyelv, azonosító szám és egyéb kódok, terminus, nyelvtani jellemzők, definíció

és kontextus, szinonima, szerkesztők, dátum; kiemelendő, hogy a nyelvi adatok mindegyike mellett szerepelnie kell a forrásnak.

### 1. ábra

*Az egyszerű terminológiai adatbázis terminográfiai cédulájának sematikus rajza*  
(Arntz et al. 2009: 223)

<b>Fachgebiet(e)</b>			Sprache
Teilbestand	Projektcode	Notation	Klassifikations- schlüssel
<b>Benennung</b>			<b>Quelle</b>
(Kurzformen, Abkürzungen, orthographische Varianten)			Grammatische Angaben
<b>Definition(en)</b>			<b>Quelle(n)</b>
<b>Kontext(e)</b>			<b>Quelle(n)</b>
Bemerkungen			
Synonyme (falls nicht als separater Eintrag, dann mit Angabe der (Quelle))			Quelle(n)
Erfasser – Datum	Bearbeiter – Datum		Eintragsklasse

A hagyományosnak vagy klasszikusnak nevezhető terminológiai adatbázisok, adatbankok szigorúan onomasziológiai felépítésűek, azaz a *fogalomközpontú megközelítés* alapján rendezik az adatokat. Ez azt jelenti, hogy az adatbázis központi eleme az azonosító kódszámmal ellátott fogalom, amelynek megjelenési formája a definíció; egy fogalomhoz pedig egy terminus és egy terminográfiai cédula rendelendő hozzá. A szoftverek általában lehetővé teszik az adatok különböző szinteken történő tárolását (*entry level, index level, term level*). A fogalomközpontúság megvalósítása érdekében a gyakorlatban a terminológiai adatbázisok adatainak bevitele legtöbbször három szinten történik:

- (1) a fogalmi szinten, amely a fogalomazonosító kódszámot és egy vagy több nyelven a fogalomazonosító kódszámmal ellátott definíciót tartalmazza;
- (2) az index szinten, amely informatikai szempontból fontos a használt anyanyelv vagy idegen nyelv megadása, ún. indexálása miatt;
- (3) a jelölő szintjén, amely a fogalomhoz hozzárendelt jelölő és egyéb adatok megadására szolgál.

A fogalomközpontúság mellett a terminológiai adatbázisok másik alapelve a *terminusautonómia*, azaz a terminus önálló szerepeltetése. Ezt azt jelenti, hogy az adatbázison belül a szerkesztők egy fő terminus használatára törekszenek, tehát egy fogalomhoz egy fő terminust (*main term*) rendelnek hozzá, a szinonimákat pedig külön bevitelként kezelik. Ezt a két alapelvet több szabvány (ISO 704:2009, ISO 12200:1999 és az ISO 16642:2003) is tartalmazza (Schmitz 2010; 2011: 238).

Ezzel kapcsolatban érdemes megemlíteni, hogy Demeczky (2008: 201) tanulmányában leírtak szerint az IBM vállalaton belül a terminológiai adatbázis két különböző megjelenítési felülettel is rendelkezik. Az egyik a *fogalomközpontú*, a másik a *szóközpontú* felfogást és megjelenítési módot érvényesíti. 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. A korábban említett felépítési elv azonban nem tévesztendő össze a megjelenítési móddal. A megváltoztatott felület oka az, hogy meg kívánták könnyíteni a munkavégzést a hagyományos szótári struktúrához szokott felhasználók számára.

A fentiek alapján a hagyományos vagy klasszikus adatbázisok kategóriájába olyan típusokat soroltam, amelyeknél érvényesül a fogalomközpontúság és a terminusautonómia alapelve, és egy terminushoz – az elektronikus szótári jellegtől elrugaszkodva – egy terminográfiai adatlapot rendelnek hozzá.

### **2.2.1. Az Európai Unió intézményközi terminológiai adatbázisa (IATE)**

Az Európai Unió 23 hivatalos, és latin nyelven – 2010. évi adat szerint – 8,9 millió terminust tartalmazó intézményközi terminológiai adatbázisa (InterActive Terminology for Europe – IATE) (<http://iate.europa.eu>) 2007 óta érhető el a nagyközönség számára is, és ez nevezhető a fordítók körében a legismertebb és legtöbbet használt terminológiai adatbázisnak. Felépítéséről részletes elemzést olvashatunk Rádai-Kovács (2009), Fischer (2008), Lesznyák (2010a, 2010b) és Tamás (2010) írásaiban, ezért jelen tanulmányban csak legfontosabb jellemzőit emelem ki.

Az egyes terminusokra vonatkozó adatok külön terminográfiai cédulán szerepelnek, az alapvető adatmezők: a témakör, a definíció és forrása, a meg-

jegyzés, a kifejezés (azaz a terminus), esetleg referenciával (ekkor hiperlink feltüntetésével szerepel), a megbízhatóság és a bejegyzés kelte. A hiperlinkeknel előfordulhat, hogy a külső hivatkozások néhány esetben elavultak, azaz nem érhető el a keresett weboldal. Mivel ez sajnos általában igaz, éppen ezért egyre gyakrabban szokás a kiegészítő információkat a kontextus vagy megjegyzés adatmező alatt feltüntetni.

Ami az adatbázis erőssége, az egyben a gyenge pontja is: rendkívül sok adatot tartalmaz. Megszületésekor több uniós intézmény adatait szűrés nélkül összesítették, és a szükséges ellenőrzést utólag igyekeznek elvégezni (Lesznyák 2010a), ami már önmagában is hatalmas munkát jelent, miközben az adatbázis folyamatosan bővül is, és ez természetesen megbízhatósági problémákhoz vezet. További gond, hogy a terminográfiai cédulák nem minden esetben tartalmaznak definíciót, előfordul, hogy a kontextus helyettesíti azt. Már kidolgozott források híján ilyenkor ideális esetben a definícióalkotás a terminológus feladata. A nyelvi adatok megbízhatóságát az adatbázisban csillaggal jelzik, ami külön értelmezésre szorul. Kiemelendő azonban, hogy – a korábban említett Termin adatbázison kívül – a tanulmányban elemzett rendszerek közül ez az adatbázis tartalmaz magyar nyelven is terminusokat.

### 2.2.2. A TERMit

A TERMit terminológiai adatbank a Trieszti Tudományegyetem Fordító- és Tolmácsképző Intézetében a hallgatók aktív bevonásával készül, jelenleg csak intraneten érhető el (<http://www.messaggiami.net/sslmit/termit.htm>). Olasz, angol, német, francia, orosz, spanyol, portugál, horvát és kínai nyelven kb. 50 000 terminust tartalmaz. A TERMit célcsoportját elsősorban szakfordító- és tolmács hallgatók, valamint más karok diákjai képezik, másodsorban azonban gyakorlati szakembereknek, szakfordítóknak, tolmácsoknak, illetve a terminológia iránt érdeklődőknek is ajánlják. A terminológia területén jelentős múlttal rendelkező intézet berkein belül eleinte hagyományos módszerrel rögzítették az adatokat, majd a fokozatos fejlesztéseket követően 2002-ben született meg az elektronikus terminológiai adatbank, amelyet azóta is folyamatosan fejlesztenek.

Az adatok két cédulán szerepelnek: a terminográfiai és a bibliográfiai cédulán. Algoritmusa aprólékosan kidolgozott (Magris 1996, magyarul Tamás 2009; 2010), a hallgatók szakdolgozatuk részeként a vizsgált tárgykör (domén) terminusait már eleve a TERMit legapróbb részletekig kidolgozott szerkesztési útmutatója alapján készítik el.

A terminográfiai cédulán nemcsak a definíció szerepel minden esetben, hanem az ekvivalencia mező is. A hagyományos felépítésnek megfelelően a definíció a terminológiai adatbázisban központi szerepet játszik, pontosságára és a megbízható forrásra nagy hangsúlyt fektetnek. Az ekvivalencia mezőben információt találunk arról, milyen szintű ekvivalencia áll fenn a terminusok között: teljes, részleges, vagy nem áll fenn ekvivalencia. A részleges ekvivalencia és az ekvivalencia hiánya esetében felsorolást kapunk a fogalmak jellemzőinek eltéréseiről, sőt az adatbázis fordítási ekvivalenst is megad úgy, hogy jelzi: terminusjavaslatról, nem a szakma által elfogadott és használt terminusról van szó.

Az adatok megbízhatóságát az adja, hogy minden esetben szakember ellenőrzi őket. Részletes kidolgozottsága és megbízhatósága miatt ez az adatbázis a hagyományos változatok közül is prototípusnak tekinthető.

### 2.3. A terminológiai adatbázisok összetett változatai

Az e csoportba sorolt adatbázisok alapvető jellemzői megegyeznek a hagyományos adatbázisok jellemzőivel, ezért csak a különbségeket és a kreatív elemeket ismertetem.

#### 2.3.1. Az EOHS Term

Az EOHS Term tudásbázis ([www.eohsterm.org](http://www.eohsterm.org)) szintén egyetemi műhely terméke, a Bolognai Tudományegyetem Fordító- és Tolmácsképző Intézet műhelyében készült, ahol ma már megrendelésre is készítenek hasonló termékeket. Az interneten 2003 óta tanulmányozható, egy viszonylag szűk szakterületet, a munkahelyi baleset-megelőzésről és biztonságról szóló jogszabályok anyagának kb. 800 terminusát dolgozza fel négy (olasz, angol, francia, német) nyelven.

A terminológiai adatbázison belül ötféle kiindulási keresési lehetőség áll rendelkezésre: az összes adat megjelenítése alfabetikus sorrendben az adott nyelven belül; az adott doménen belül az aldomének szerint csoportosított terminusok listája; többnyelvű keresés táblázatos formában az ekvivalensek jelölésével (többnyelvű glosszárium); egyszerű keresés és részletes keresés. A terminográfiai cédulákon szintén jelölik a terminusok közti ekvivalenciaszintet, amelyhez különböző írásjeleket használnak: a teljes ekvivalencia jelöléséhez az egyenlőségjelet (=) választották, a részleges ekvivalenciát jele:  $\pm$ , az ekvivalencia hiányát pedig a  $\sim$  mutatja. Ezek a jelek az egyes idegen nyelvű ekvivalensek vagy kapcsolódó terminusok előtt szerepelnek, és a fő terminushoz való viszonyukat mutatják.

A tudásbázis elnevezés is jelzi, hogy nem egyszerűen adatbázisról vagy adatbankról van szó, hanem egy olyan segédeszközzel, amely a terminológiai adatbázison kívül más ismeretelemeket is tartalmaz: a jogszabályok gyűjteményét és a terminológiai rendszerek grafikus ábrázolását. A jogszabályokon belüli keresést a források könyvtára négy nyelven teszi lehetővé. Ezen belül két funkció is létezik: tanulmányozhatók az annotált jogszabályok, és konkordanciák is lekérhetők. A másik újdonság a terminológiai rendszerek gyűjteménye, amely lehetővé teszi az elemzett domén terminológiai hálójának grafikai megjelenítését. A szakfordító ennek segítségével egy-egy terminus megfelelő fordításához rálátást kap a terminológiai rendszerre, azaz a terminusok közti alá-, fölé- és mellérendeltségi viszonyokra. A három elem, azaz az adatbank, a források könyvtára és a terminológiai rendszerek grafikai ábrázolása között az átjárás minden esetben biztosított, és ezek ennek köszönhetően együttesen komplex rendszert alkotnak.

#### 2.3.2. A BISTRO

A Bolzanói Európai Akadémia Szaknyelvi és Többnyelvűségi Intézetének (*Europäische Akademie Bozen/Accademia Europea di Bolzano, Institut für*

*Fachkommunikation und Mehrsprachigkeit/Istituto di comunicazione specialistica e plurilinguismo*) jogi terminológiai projektje, a BISTRO (*Informationssystem für Rechtsterminologie/Sistema informativo per la terminologia giuridica*) (<http://eurac.edu/bistro>), amely azért jött létre, hogy a dél-tiroli kisebbség, azaz a német és ladin ajkú népesség számára megkönnyítse az igazságügyi és közigazgatási feladatok ellátását. Használatára a hivatalos honlapokon is felhívják a figyelmet.

Az 1992-ben alapított Bolzanói Európai Akadémia innovatív kutatóintézet, elsődlegesen a többnyelvűségre és a környezetvédelmi kérdésekre koncentrál, ezen belül tevékenysége öt kutatási területet ölel fel: alkalmazott nyelvészet, kisebbség és autonómia, fenntartható fejlődés, kulturális menedzsment és természettudományok. Az intézetben gazdasági, jogi, természettudományi szakemberek és nyelvészek végeznek kutató tevékenységet.

Maga a BISTRO elnevezés valóban a gyorsítkezésre utal, a kutatók a név választást annak etimológiájával indokolták: amikor Napóleon vereségét követően a szövetséges erők orosz katonái Párizsba értek, akkor a pincéreket a „*Bystro!*” ‘gyorsan’ szóval biztatták a minél fürgébb kiszolgálásra. Állítólag innen származik a vendéglátóiparban magyar nyelven is használatos *bisztró* elnevezés. A terminológiai információs rendszer célja pedig valóban az, hogy gyors kereséssel és (a hagyományos szótárakhoz képest) többletinformációk nyújtásával a felhasználó számára hasznos és megbízható információkat nyújtson.

A BISTRO a következő szövegtörzseket tartalmazza (Streiter et al. 2004):

- (1) egy kétnyelvű szövegtörzset: jogszabályokat olasz és német nyelven, illetve
- (2) egy háromnyelvű szövegtörzset: közigazgatási szövegeket olasz, német és ladin nyelven.

Az információs rendszer a következő elemekből épül fel:

- (1) A *term search* funkció alatt: egy olasz, német és ladin nyelven 50 000 terminust tartalmazó adatbank, amelynek adatmezői a következők: terminus, definíció, kontextus, ekvivalens, forrás, országkód.
- (2) A *corpus search* funkció alatt például:
  - CATEX: párhuzamos szövegrészek elemzésére nyílik lehetőség;
  - KWIC: a rendszer egynyelvű annotált szövegszegmenseket kínál fel, amelyektől azonban a párhuzamos szövegekhez (CATEX) hiperlink révén vissza lehet térni.
- (3) A *term tools* funkciót az értelmezni kívánt szöveg bevitelével a következőkre hasznosíthatjuk:
  - terminusfelismerés
  - terminuskivonatolás.

Az adatbankban a keresés leszűkíthető kizárólag a szabványosított terminusokra. Ezen felül a BISTRO a felsőoktatási terminusokra egy külön kétnyelvű adatbázist is tartalmaz.

A BISTRO nem lezárt projekt, folyamatosan bővül és fejleszthető. A német nyelv területén segítséget nyújt azon nehézségek megoldásában, amelyek abból erednek, hogy az olasz jogi fogalmak és terminusok német nyelvre történő fordításakor a többi, német hivatalos nyelvet használó jogrendszerhez próbálnak igazodni: a németországi, az osztrák és a svájci jogrendszerhez, amelyek közül – Dél-Tirolról lévén szó – természetesen az osztrák előnyt élvez. A BISTRO használatában érdekeltek a helyi köztisztviselők, jogászok, fordítók, tolmácsok és lakosok.

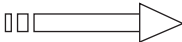
## 2.4. A terminológiai adatbázisok típusainak áttekintése

Az áttekintő táblázatban – az egyszerűbb eszközöktől az egyre összetettebb felé haladva – az interneten elérhető megjelenítési felületet, a felépítésen belül az adatmezőket és a kiegészítő elemeket figyelembe véve kísérletet tettem az adatbázisok kategóriákba történő besorolására. Az elemzett adatbázisokat az alábbi fő kategóriák szerint csoportosítottam:

- (1) egyszerű adatbázis: Termin, Termium Plus, TERMDAT;
- (2) hagyományos (klasszikus) adatbázis: IATE, TERMit;
- (3) összetett adatbázisok: EOHS Term, BISTRO.

### 2. ábra

*Terminológiai adatbázisok besorolásának áttekintő táblázata*

Alapváltozat:	Egyszerű:	Klasszikus:		Összetett:	
Glosszárium (pl. Excel-táblázat)	Termin Termium Plus TERMDAT	IATE	TERMit	EOHS Term	BISTRO
-					+
<b>KREATIVITÁS</b>					
					

Az egyszerű két- vagy többhasábos, esetleg excel-táblázat formájában készült glosszárium és az összetett tudásbázisok és terminológiai információs rendszerek között igen nagy a különbség. Véleményem szerint azonban minden eszköznek megvan a maga létjogosultsága, mert a glosszárium akár egyéni, akár csoportos fordításnál konkrét szövegeknél használt ad hoc megoldások rögzítésére – az egységes használat érdekében – jó szolgálatot tehet, de ennek az eszköznek a megbízhatóságával érdemes tisztában lenni. Ellenben a fordítóirodáknak a minőség biztosítása érdekében ennél jóval összetettebb eszközre is szükségük van: aprólékosan kidolgozott, egységes terminushasználatot biztosító adatbázisra, adatbankra, tudásbázisra vagy terminológiai információs rendszerre, amely tartalmazza a problémás vagy egyszerűen csak egységesítésre szoruló

nyelvi adatokat, esetleg párhuzamos szövegeket, jogszabályokat, terminológiai rendszereket bemutató gráfokat (ontológiai rendszereket) és egyéb kreatív elemeket. Mivel ezek az eszközök rugalmasan alakíthatók, ezért igény szerint további elemekkel bővíthetők.

### 3. A terminológusról és feladatairól

Az eszközök áttekintését követően érdemes a terminológiai adatbázisok létrehozását koordináló személyre is röviden kitérni és megvizsgálni, vajon mi a terminológus (vagy terminológia-menedzser) szerepe: melyek a feladatai, mi az, ami munkájában adott, és mennyire lehet kreatív. Ahogyan a fordító feladata sem az, hogy „kitaláljon” terminusokat, vagy kizárólag a megérzéseire hagyatkozzon, ma a terminológus sem „mindentudó”, hanem vállalaton vagy intézményen belüli terminológia-menedzser, aki kidolgozza a vállalati stratégiát, és meghatározott keretek között végzi feladatát. Felméri az igényeket, kezeli a nyelvi adatokat, gondoskodik azok rögzítéséről és közzétételéről. A szoftverháttér kiválasztásában egyeztet, felelős a terminológiai adatbázis felépítéséért, szerkesztéséért, a terminológiai elvek és módszerek érvényesítéséért, a megbízhatóságot biztosító kritériumok érvényesüléséért. Gondoskodik az adatok karbantartásáról, a célzott frissítésekről, a csapatmunka koordinációjáról. Ennek érdekében konzultál, egyeztet, workshopokat szervez, és karöltve dolgozik:

- az informatikussal;
- a nyelvi szakemberrel (fordító, lektor);
- az adott domén/tárgykör szakemberével (jogász, közgazdász, mérnök, orvos stb.);
- sőt, az ügyfelek visszajelzéseit, esetleges reklamációit sem hagyja figyelmen kívül.

Ez nem azt jelenti, hogy a terminológus ne rendelkezzen nyelvi vagy tárgyi ismerettel. Az adatok rögzítésében és közzétételében is van mozgástér, munkája során megoldásokat kell keresnie például az alábbi kérdésekre:

- Milyen felépítésű adatbázis felel meg leginkább a felhasználói igényeknek: szükség van-e jogszabályi hivatkozásokra, ábrára vagy fotóra (például műszaki terminus esetében), külső hivatkozásra (hamar elavulhat, ezért szerencsésebb a kontextus beillesztése), terminológiai rendszer ábrázolására stb.?
- Ha nincs a terminusra definíció, akkor a terminológus a fogalmi jellemzők alapján alkothat-e definíciót?
- Melyik definíció típus felel meg leginkább a felhasználók igényeinek?
- Hogyan jelölje a terminusok közti ekvivalencia-szintet? Milyen kategóriákat használjon?
- Melyik a legmegbízhatóbb forrás?
- Milyen szakemberhez forduljon?

Tehát a terminológia-menedzser fő feladatai közé tartozik, hogy háttérrel nyújtson a csapatmunkához, megbízható forrásokat gyűjtsön, konzultáljon a megfelelő szakértővel (nyelvi aspektusból és a tárgyismeret szempontjából egyaránt), és felügyelje a terminológiai elvek és módszerek érvényesülését.

Türelmet és precizitást igénylő, aprólékos munkáról van szó, amely véleményem szerint kreativitást is kíván, de azt meghatározott keretek között teszi lehetővé. Igaz ez például a terminológiai adatbázis felépítésének megtervezésére, hiszen a nyelvészeti algoritmus kidolgozásával el kell döntenetni, hogy az adott terminográfiai cédulán milyen adattípusokat, illetve adatokat (tartalmat) érdemes előnyben részesíteni. A végső cél az, hogy az adatbázis könnyen hozzáférhető, felhasználóbarát környezetben, egy helyen, megbízható adatok nyújtásával úgy kínáljon a fordító számára megoldásokat, hogy az egyben túlmutasson az egyszerű behelyettesítésen. A terminológiai adatbázis a többletinformációknak köszönhetően ezt lehetővé teszi, mert mérlegelési lehetőséget is biztosít, az információk széles körű közzétételével pedig mód van a terminusok használatára és minőségi munka végzésére.

#### 4. Összegzés

Tanulmányomban kísérletet tettem a különböző típusú terminológiai adatbázisok leírására és hármas csoportosításukra. Az egyszerű terminológiai adatbázisokra jellemző, hogy nincsenek elég aprólékosan kidolgozva, az adatmezők száma általában igen kevés, illetve hogy a fogalomközpontúság mellett nem érvényesül az terminusautonómia elve. Előfordul, hogy megőrzik a szólista (glosszárium) jelleget, emiatt a megjelenítési felület inkább elektronikus szótár-ra emlékeztet.

A hagyományos terminográfiai adatbázisokat ellenben már aprólékosan kidolgozták, az adatmezők száma megfelel a minimálisan előírt adatmennyiségnek vagy meg is haladja azt, központi elemük a definíció, és érvényesül a terminusautonómia elve.

Az összetett terminológiai adatbázisok, adatbankok a hagyományos változaton túl további ismeretelemeket vagy kiegészítő eszközöket tartalmaznak. Megtalálhatjuk bennük jogszabályok gyűjteményét, terminológiai rendszerek grafikai ábrázolását vagy akár terminuskivonatoló és terminuskinyerő segéd-eszközöket is. Mivel általában rugalmasan alakíthatók, ezért további elemekkel bővíthetők, tehát a lehetőségeknek az igények és a kreativitás szabnak határt.

### Irodalom

- Arntz, R., Picht, H., Mayer, F. 2009. *Einführung in die Terminologearbeit*. Hildesheim, Zürich, New York: Georg Olms Verlag. (6. Aufl.)
- Bullo, F. 2009. Introduzione alla terminologia del diritto. In: Cavagnoli S., Ioratti Ferrari E. (eds.) *Tradurre il diritto. Nozioni di diritto e linguistica giuridica*. Milano: Cedam. 223–258.

- Cabré, M. T. 1998. *Terminology. Theory, Methods and Applications.* (Terminology and Lexicography Research and Practice 1) Amsterdam/Philadelphia: John Benjamins.
- Castagnoli, S. 2003. *Creazione di una knowledge base multilingue in materia di sicurezza sul lavoro: database terminologico quadrilingue, codifica e modalità di interrogazione del corpus di riferimento.* Dissertazione. Alma Mater Studiorum Università di Bologna. Scuola Superiore di Lingue Moderne per Interpreti e Traduttori. Kézirat.
- Demeczky J. 2008. Terminológia a szoftveriparban. *Magyar Terminológia* 1. évf. 2. szám. 189–204.
- Fischer M. 2008. Az európai uniós fordítás és terminusalkotás magyar vonatkozásai. *Magyar Nyelvőr* 132. évf. 4. szám. 385–402.
- Fóris Á. 2005. *Hat terminológia lecke.* Pécs: Lexikográfia Kiadó.
- Lesznyák Á. 2010a. Kulisszatitkok az uniós terminológiáról. Elhangzott: *Fordítók és Tolmacsok Őszi Konferenciája.* Budapest, Budapesti Corvinus Egyetem, 2010. október 1.
- Lesznyák Á. 2010b. Az európai uniós intézmények terminológiai adatbázisa: A IATE. *Magyar Terminológia* 3. évf. 2. szám. 161–181.
- Magris, M. 1996. Costituzione di una banca dati terminologica presso la S.S.L.M.I.T. di Trieste. *Ritt = Rivista Internazionale di tecnica della traduzione*, n. 2, Trieste: Università degli Studi di Trieste. 138–159.
- Rádai-Kovács É. 2009. *Az euroterminus, avagy az európai uniós terminológia jellemzői, különös tekintettel az újlatin nyelvekben megjelenő sajátosságokra.* Doktori disszertáció. ELTE BTK Nyelvtudományi Doktori Iskola. Budapest. Kézirat.
- 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.) *Nyelvek találkozása a fordításban. Doktori kutatások Klaudy Kinga tiszteletére.* Budapest: ELTE Eötvös Kiadó. 101–116.
- Sermann E., Tamás D. 2012. Elektronikus szótár vagy terminológiai adatbázis? *Társadalmi változások – nyelvi változások.* A XXII. MANYE Kongresszus. Elhangzott: Szeged, 2012. április 12–14. (Megjelenés alatt a konferenciakötetben.)
- Schmitz, K-D. 2010. A Closer Look at Terminology Management Systems. Elhangzott: *International Terminology Summer School (TermNet).* University of Vienna, Center of Translation Studies, 24–28 May 2010.
- Schmitz, K-D. 2011. Managing terms in terminology management. *Magyar Terminológia* 4. évf. 2. szám. 238–245.
- Tamás D. 2009. A trieszti TERMit adatbázis vizsgálata. *Magyar Terminológia* 2. évf. 2. szám. 213–235.
- Tamás D. 2010. *A gazdasági szakszövegek fordításának terminológiai kérdéseiről az olasz–magyar nyelvpár esetében.* Doktori disszertáció. ELTE BTK Nyelvtudományi Doktori Iskola, Budapest. Kézirat.

## Internetes hivatkozások

- Castagnoli, S. 2006. *Turning a termbase into an integrated knowledge base: the Environmental and Occupational Health & Safety terminology website.* Proceedings of „La Formazione in Terminologia”, *mediAzioni* 3, On-line Journal of Interdisciplinary Studies in Languages and Cultures. [http://www.mediazionion-line.it/monografici/castagnoli\\_ita.htm](http://www.mediazionion-line.it/monografici/castagnoli_ita.htm) (a letöltés ideje: 2009. július 23.)

Streiter O., Ralli N., Ties I., Voltmer L. 2004. BISTRO. The Online Platform for Terminology Management: Structuring Terminology without Entry Structures. <http://www.eurac.edu/en/research/institutes/multilingualism/Documents/bistro/lans.pdf> (letöltés ideje: 2011. március 12.)

## **Források**

Szabványok

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.

ISO 704:2009 Terminology work – Principles and methods.

## **Elektronikus források**

[external.kim.gov.hu](http://external.kim.gov.hu) (a letöltés ideje: 2012. március 14.)

<http://www.termiumplus.gc.ca> (a letöltés ideje: 2012. március 14.)

[www.termdat.ch](http://www.termdat.ch) (a letöltés ideje: 2012. március 12.)

<http://iate.europa.eu> (a letöltés ideje: 2012. március 14.)

<http://www.messaggiami.net/sslmit/termit.htm> (a letöltés ideje: 2010. május 26.)

[www.eohsterm.org](http://www.eohsterm.org) (a letöltés ideje: 2012. március 14.)

<http://dev.eurac.edu> (a letöltés ideje: 2012. március 12.)

<http://eur-lex.europa.eu> (a letöltés ideje: 2012. március 14.)