

TÖBBNYELVŰ TEZÁURUSZOK

KÉSZÍTETTE:

Dula Marina

informatikus könyvtáros

TÉMAVEZETŐ:

Barátné dr. habil Hajdu Ágnes

egyetemi docens

Budapest • 2015

A dokumentum lezárásának dátuma: 2015. január 7.

TARTALOMJEGYZÉK

1. ELŐSZÓ	4
2. BEVEZETÉS	5
3. A TÖBBNYELVŰ TEZÁURUSZOK TÖRTÉNETI NÉZŐPONTBÓL.....	6
3.1 A KIALAKULÁS ELŐZMÉNYEI.....	6
3.1.1 Az alfabetikus tezauruszok korai dominanciája	7
3.1.2 A facettás tezauruszok megjelenése	9
3.2 A TÖBBNYELVŰ TEZÁURUSZOKRA VONATKOZÓ SZABVÁNYOK, IRÁNYELVEK	10
3.2.1 Kelet-Európa tezaurusz szabványai	10
3.2.2 Nyugat-Európa tezaurusz szabványai és irányelvei	11
4. A TÖBBNYELVŰ TEZÁURUSZOK FOGALMI MEGHATÁROZÁSAI ÉS SZERLEZETI ELEMEI.....	13
4.1 A TÖBBNYELVŰ TEZÁURUSZOK FOGALMA A SZAKIRODALOMBAN	13
4.2 A TÖBBNYELVŰ TEZÁURUSZOK STRUKTÚRÁJA ÉS SZERKEZETI ELEMEI	16
4.2.1 A deszkriptorok és szerepük	16
4.2.2 A nemdeszkriptorok és szerepük	17
4.2.3 Egyenlőségi, hierarchikus, és asszociatív kapcsolatok és a jelölők.....	18
4.3 A TÖBBNYELVŰ TEZÁURUSZOK SZERKEZETI MÉLYSTRUKTÚRÁI	19
4.3.1 A fordítókód.....	21
4.3.2 A természetes nyelv, mint közvetítő nyelv.....	22
5. TÖBBNYELVŰ TEZÁURUSZOK KIALAKÍTÁSÁNAK LEHETŐSÉGEI	23
5.1 A NEMSZIMMETRIKUS ÉS A SZIMMETRIKUS TEZÁURUSZ JELLEMZŐI.....	23
5.2 A TÖBBNYELVŰ TEZÁURUSZ KIALAKÍTÁSA ALULRÓL-FELFELÉ ÉPÍTKEZVE	24
5.2.1 A tulajdonnevek megjelenítése	25
5.2.2 A homonimák kérdése	26
5.2.3 Az összetett kifejezések kérdése	27
5.2.4 Az ekvivalenciák	29
5.3 MÁR KIALAKÍTOTT TEZÁURUSZOK ÖSSZEOLVASZTÁSA, ÖSSZEKAPCSOLÁSA	34
5.3.1 Az ekvivalenciák:	34
5.4 A TEZÁURUSZOK FORDÍTÁSA.....	36
6. A TÖBBNYELVŰ TEZÁURUSZOK MAI MEGJELENÉSI FORMÁI ÉS A JÖVŐBELI ELKÉPZELÉSEK.....	39
6.1 A TÖBBNYELVŰ TEZÁURUSZOK FORMAI ELRENDEZÉSE	39
6.1.1 Az UNESCO többnyelvű tezauruszának megjelenítési lehetőségei.....	41
6.1.2 Az Eurovoc megjelenítési lehetőségei	42
6.2 A TÖBBNYELVŰ TEZÁURUSZOK PROBLEMATIKUS TERÜLETEI.....	43
6.2.1 Az EXPO 2000 és az InfoDEFT újszerű megoldásai	45
6.2.2 Az interoperabilitás, a jövő tezauruszaiban	46
7. ÖSSZEGZÉS.....	48
8. BIBLIOGRÁFIA	51

1. ELŐSZÓ

A 21. században a többnyelvű eszközök egyre nagyobb teret nyernek és egyre szélesebb körű felhasználói csoporthoz jutnak el, olyanokhoz is, akik eltérő kulturális és nyelvi háttérrel rendelkeznek és az ismeretek hasonlóan változatos területei iránt érdeklődnek. Éppen ezért olyan rendszerekre van szükség, amelyek támogatják a többnyelvű információhoz való hozzáférést, az egyre inkább globálissá váló információs társadalomban. Tény, hogy a teaurusz alapú rendszerek rendelkeznek a legjobb pontossági aránnyal. Ez a párhuzam indokolja, hogy foglalkoznunk kell a többnyelvű teauruszokkal.

Ezen kívül, azért tartom fontosnak a terület értékelő feldolgozását, mivel nincs a többnyelvű teauruszokra vonatkozó magyar szabványunk, illetve a szakirodalmunk is bár alapvető és értékes feldolgozói munkákat mutat, mára ezek meglehetősen elavultak a témában. Dolgozatommal, jelen szerény keretek között, de nagy elhivatottsággal kívánom betölteni a fennálló űrt.

2. BEVEZETÉS

Dolgozatomban két kérdésre keresem a választ, az első: „Miről lehet felismerni és milyen tulajdonságok jellemzik a jó többnyelvű tezaurszt?” Illetve: „Hogyan lehet kialakítani egy ilyen rendszer, hatékony és felhasználóbarát verzióját?”

A fentiek megválaszolását a többnyelvű tezaurszok kialakulásának áttekintésével kezdem, bemutatva, hogy a Roget-féle 19. századi tezaursz óta, mennyit változott a fogalom értelmezése és azt is, hogy adott évtizedekben milyen újítások hatottak és ezek hogyan formálták az ilyen típusú rendszereket. Illetve áttekintem Kelet- és Nyugat-Európa tezaurszokra vonatkozó szabványait és irányelveit.

A következő részt a többnyelvű tezaurszok fogalmi meghatározásainak bemutatásával kezdem, majd az egyes szerkezeti elemek jellemzőinek leírásával folytatom. Ezen belül két szinten dolgozom fel a témát, először klasszikus értelmezésben a felszínen is megjelenő összetevők, így a deszkriptorok, nemdeszkriptorok és relációk szintjén, ezt követően pedig a mélystruktúrák elemeivel foglalkozom.

A következő fejezet, adja majd ki a dolgozat úgynevezett magját, amikor is a többnyelvű tezaurszok kialakításának vizsgálatára kerül sor. A lehetséges eljárásokhoz (alulról-felfelé építkezés, összeolvasztás, összekapcsolás és fordítás) összegyűjtöm a szabványok és a szakirodalom által rögzített tanácsokat és az egyes alternatívákat. Ezekkel párhuzamban pedig kiemelten jellemzem az ekvivalenciák kérdéskörét.

Később a már alakot öltött ellenőrzött terminológiai szótárak megjelenítési lehetőségeit mutatom be, amihez az Eurovoc és az UNESCO tezaurszának elektronikus felületei nyújtanak segítséget. Majd a többnyelvű tezaurszok problematikus kérdéseivel foglalkozom, melyek a kulturális és nyelvi különbözőségekből fakadnak és ismertetem az ezekre megoldást jelenthető, az EXPO 2000 és az InfoDEFT újításait. Végül pedig bemutatom az interoperabilitás fogalmát, amely a jövőben a többnyelvű tezaurszokat meghatározó központi szempont lehet.

3. A TÖBBNYELVŰ TEZAUROSZOK TÖRTÉNETI NÉZŐPONTBÓL

3.1 A kialakulás előzményei

A „tezausz” görög eredetű szó, a kifejezés angol nyelvterületen először 1736-ban jelent meg, az Oxford Szótár rövidített változatában. Jelentése akkoriban a tudás kincsesládát vagy az ismeretek-enciklopédiáját jelölte. A leghíresebb – ebben az értelemben vett tezausz – *Peter Mark Roget* gondozásában készült el 1852-ben „*Thesaurus of English Words and Phrases*” címen, amelynek felépítése alfabetikus volt, de szisztematikus témakörök szerinti elrendezés is szerepelt benne. Egy évszázaddal később a tezausz kifejezést már az információ visszakeresésre használt tárgyszólisták megnevezésére alkalmazták, azonban eddig a pontig többféle megközelítésben volt használatos a fogalom (*Gilchrist*, 1971).

Kezdetben a tezausz többeknél, így *Batten* (1947) és *Mooers* (1963) korai elméleteiben is manuális posztkoordinált rendszerként jelent meg. Ebben a struktúrában az indexelő kifejezéseket anélkül rendelik hozzá a dokumentumhoz, hogy figyelembe vennék a szintaxist, ellentétben a prekoordinált rendszerekkel, ahol az indexelő által összekapcsolják a kifejezéseket egy állandó karaktersorozatba. A posztkoordinált rendszerek sajátossága, hogy a kereső, a kifejezések bármilyen kombinációjával megtalálhatja a releváns elemeket.

A továbbiakban a tezauszt a posztkoordináltság helyett az információ visszakeresés szempontjából határozták meg. Egyetértés van abban, hogy az információ visszakereséssel összefüggésben *Peter Luhn* (1957) az IBM munkatársa használta először a tezausz kifejezést. Ez az aspektus már jól dokumentált, nemcsak *Gilchrist*, hanem *Vickery* (1960), *Lancaster* (1986) és *Roberts* (1984) is ebben a szemléletben készítette el munkáit.

Az első tezauszt valójában egy információ visszakereső rendszer szókincsének ellenőrzésére használták, amelyet a *Du Pont Szervezet* fejlesztett ki 1959-ben. Azonban az első már széles körben is hozzáférhető ellenőrzött szótár az *ASTIA* (1960) volt, amelyet az *Amerikai Hadászati Műszaki Információs Központ* tett közzé. Ennek szorosan a nyomában, másodikként jelent meg a *Vegyésszmérnökök Amerikai Intézetének* mikrotezausza a „*Chemical Engineering Thesaurus*” (1961).

A tezauszok az 1950-es években többféle formában alakultak ki, amelyben a legnagyobb szerepet Mortimer *Taube*, *Uniterm* rendszere játszotta. A *Taube* által kifejlesztett struktúra nem ellenőrzött, a dokumentum szövegéből kiválasztott egy tagú, uniterm kifejezések

jellemezték a tartalmat. Ezekkel viszont problémák merültek fel a szinonimák, a homonimák és az azonos alakú szavak összefüggésében. Az „unitermek” gyakran eltérő jelentéssel bírtak a szöveggörnyezettől, az alkalmazástól, vagy a nézőponttól függően, a rendszerből azonban hiányoztak a minősítők, így a kifejezések közötti speciális különbségek nem voltak kellően kifejezve. Így például az „electrical insulation” – magyarul elektromos szigetelés - kifejezés különbözik a „thermal insulation” (hőszigetelés kifejezéstől) de ezek közös része, az egytagú „insulation” alapján nem lehet a jelentésbeli eltérés árnyalatát érzékelni. Az unitermek tehát alkalmatlanok voltak az egyedi indexelés és a velük való keresés biztosítására, így a rendszer ebben a formájában nem maradt fenn sokáig, annak ellenére, hogy elméleti szerepe és jelentősége máig ható.

A Taube által kivitelezett struktúrát végül felváltották azok a szótárak, amelyek már jelentős számú összetett kifejezést is tartalmaznak. Ezekben a korai fogalomtárakban már elkezdtek ellenőrizni a szinonimákat, a homonimákat és megjeleníteni a kifejezések közti hierarchikus és asszociatív kapcsolatokat is (Aitchinson és Clarke, 2004).

A 60-as években váltak szembeűnővé a különbségek a hagyományos tárgyszólisták és a fejlődő tezauszok között. Bár a tárgyszólistákban is jellemzőek voltak a "Lásd" és a „Lásd még” hivatkozások, de fontos eltérés volt, hogy a tárgyszólistákkal kapcsolatban hiányosak voltak a pontos szabályok a szókincs ellenőrzésére és a fogalmak közötti kapcsolatok kifejezésére. Ebben az időszakban minden tezausz kifejezéseket sorolt fel alfabetikus sorrendben. A szabványos forma 1967-ben alakult ki, amikor a *Thesaurus of Engineering and Scientific Terms (TEST)* megjelent a *Mérnökök Közös Tanácsának* gondozásában és helyébe lépett *Thesaurus of Engineering Terms of the Engineers Joint Council (EJC)* korábbi verziójának (1964). Clarke szerint (2001) a *TEST* alfabetikus tezauszban már jelen voltak azok a funkciók, ami a következő 30 év és a ma tezauszait egyaránt jellemzik, ide értve az ekvivalenciákat és a hierarchikus és az asszociatív kapcsolatokat is.

3.1.1 Az alfabetikus tezauszok korai dominanciája

A korai tezauszok domináns eleme volt a deskriptorok és a nemdeskriptorok alfabetikus megjelenítése. A téma áttekintése, illetve a szisztematikus megjelenítés, ha egyáltalán létezett, csak másodlagos fontosságú volt ebben az időszakban. Egy részletes, minősített elrendezést, mint amilyen az enumeratív osztályozási rendszerekben van jelen, túl bonyolultnak és elavultnak tartottak ahhoz, hogy itt szerepet kapjon információ

visszakeresésben. Azonban néhány tezauszoknak voltak kezdetleges szisztematikus szakaszai, amelyek tág témakategóriákból álltak.

A *TEST*-ben a kifejezések betűrendes felsorolásával – meglehetősen jelentős terjedelemben – 690 oldalon keresztül foglalkoztak, illetve külön listán szerepeltették a csúcsdeszkriptorokat, ahol feltűntették a szűkebb értelem minden szintjét. A *TEST*-nek is volt egy szisztematikus fejezete, amely 22 tárgyi kategóriát különböztetett meg, amelyeken belül egy hierarchia szinttel lejjebb alcsoportok helyezkedtek el. Azonban a tezausz összes deszkriptora csak az alfabetikus sorrendben jelent meg.

Sok tezauszt készítettek el nemzetközi szervezetek, a legismertebb ilyen makrotezausz előállítója *Jean Viet* volt, aki művében már sokkal kifinomultabb tárgyi csoportosításokat szerepeltett. Az *ENSZ Makrotezaurusza* először 1972-ben jelentet meg, jelenleg már az ötödik kiadását használják (1998). Egy példa az 1985-ös kiadásból (*Aitchinson és Clarke*, 2004):

14 DEMOGRAPHY. POPULATIONS
14.01 POPULATION DYNAMICS
14.01.01 CIVIL REGISTRATION DEMOGRAPHIC STATISTICS <i>POPULATION DATA</i> <i>USE: DEMOGRAPHIC STATISTICS</i>

3.1. táblázat: Az *ENSZ* tezauszának hierarchikus felépítése (*Aitchinson és Clarke*, 2004)

A deszkriptor csoportok megjelenítése főcsoportokat tartalmaz, amelyeket alcsoportokra bontanak, ezeken belül pedig az alcsoportok deszkriptorai további klaszterekre tagolódnak (3.1. táblázat). A tezauszok ezen formájában a szisztematikus rész jellemzően alárendelt az alfabetikus szekciónak, ahelyett, hogy ez képezné az ellenőrzött szótár velejét. A főlé-, alárendelt és kapcsolódó kifejezések csak a betűrendes megjelenítésben vannak feltűntetve.

A betűrend dominanciája alól kivételt képezett a *Medical Subject Headings* - gyakran csak *MESH*-ként emlegetjük - amely az *Orvostudomány Nemzetközi Könyvtárának* gondozásában jelent meg az 1960-as években. Az orvosi tezauszban a betűrendes rész nem felelt meg a ma szokásos tezausz elrendezésnek, mivel nem mutatta a főlérendelt és az alárendelt kifejezéseket nem tartalmazott magyarázatokat, szinonimákat és kapcsolódó kifejezéseket. Az alfabetikus rész elsősorban a gráfok mutatójaként szolgált.

Az *MESH*-ben nagy hangsúlyt kaptak a gráfok, vagyis kiterjedt hierarchiába rendezve helyezkedtek el a deskriptorok, egy témakörön belül általában tíz, illetve néhány helyen még ennél is több szintre bontva a kifejezések reprezentálását. Az ábrázolás során alkalmazott gráfok pontosan kifejezték a hierarchia szinteket és lehetővé tették, hogy a kereső úgynevezett „robbantott” keresést végezzen, vagyis hogy letöltéskor az adott deskriptorral indexelt tételek és a kifejezés alatt elhelyezkedő szűkebb értelmű változatok megtalálását is. Éppen ezért ezt a típusú grafikus megjelenítést a mai napig alkalmazzák a teauruszok szótári részében (*Aitchison és Clarke, 2004*).

3.1.2 A facettás teauruszok megjelenése

Később a facetta analízis alapvetően befolyásolta a teauruszok fejlődését. Körülbelül egy évtizeddel a teauruszok kialakulása előtt *S. R. Ranganathan* kifejlesztett egy innovatív osztályozási technikát. Gondolatai elterjedtek az 1940-es évek és a korai 50-es években az Egyesült Királyságban is. Az osztályozás nagyjai közül kikerült ezen új technika kedvelői hozták létre Londonban a *Classification Research Group*-ot (*CRG*), hogy megosszák egymással ötleteiket és különböző facettás osztályozási rendszereket dolgozzanak ki számos speciális területen. Az egyik ilyen speciális a *CRG* által kivitelezett munka az angol *Elektrotechnikai Társaság Könyvtárának* készített rendszer volt, amelyet később alfabetikus teaurusszal támogattak az élvonalbeli információ visszakeresés érdekében, nem pedig indexszel. Az így kapott teaurusz 1969-ben jelent meg, ez volt az első olyan ellenőrzött szótár, amely facetta analízist is használt. A *Thesaurofacet*-et egy sor hasonló elven alapuló teaurusz követte, ezek közé tartozott az *UNESCO Teaurusz* első kiadása (1977), a *BSI ROOT Teaurusz* (1985) és a *Nemzetközi Menekültügyi Teaurusz* (1996). A teauruszok ezen változatában facetta analízist használtak a kifejezések közötti kapcsolatok megállapítására és egyidejűleg építették ki a hierarchikus és asszociatív kapcsolatokat. *Bliss Bibliográfiai Besorolásának* második kiadása (1986) nagy hatással volt a fogalmak sorrendjére ebben a teaurusz stílusban.

A facettás teauruszok egy másik megjelenési formája az, amikor az ismeretek területe a teauruszban fel van osztva a *Ranganathan* által használt facettákra és ezek nem a tárgy mezőben jelennek meg. Ide sorolható a *Construction Industry Thesaurus* (1976) és az *Art & Architecture Thesaurus* (1994).

3.2 A többnyelvű teauruszokra vonatkozó szabványok, irányelvek

Krooks és Lancaster (1993) azt igazolják, hogy a teauruszok kialakításának már a korai „evolúciós” időszakában is voltak szabványaik és irányelveik. Nagy jelentőséget tulajdonítanak *Eugene Wall* tevékenységének, aki számos megoldást eszközölt az alfabetikus teauruszok kialakításának problémáin (1967). Éppen ezért a TEST számos szabályt és konvenciót köszönhet *Wall* befolyásának, amely egyfajta alapot jelentett a többnyelvű terminológiai szótárak kialakításának kérdéskörében.

Az első teauruszok megjelenése után az *UNESCO* és az *NTMIR* (*Nemzetközi Tudományos és Műszaki Információs Rendszer*) vezérlete alatt jelentek meg a többnyelvűséget célzó fő iránymutatások. Majd nemzetközi és nemzeti szinten is megkezdődtek a terminológiai szabványosítás céljára készülő szótárak fordítására, többnyelvűsítésére tett kísérletek. Így pl. az 1972-ben adták ki az NSZK-ban a *Beling, Schuck, és Wersig* által elkészített módszertani útmutatót az idegen nyelvű teauruszok német nyelvre fordítására.

3.2.1 Kelet-Európa teaurusz szabványai

Az *NTMIR* „*Többnyelvű információkereső teaurusz*” c. szabványa és ennek második kiadása is, igazából korábban készült el, mint az *UNESCO* ugyanezen témában megjelent irányelve vagy az erre vonatkozó *ISO* szabvány. 1974-ben az *NDK*-ban a még előkészületben lévő *NTP MCNTI 15-76* előírással összhangban már elkészültek az orosz-német/német-orosz nyelvű teauruszok kidolgozásának irányelvei, majd ennek példájára folyamatosan jelentek meg a kétnyelvű mikroteauruszok, építésügyi, kémiai és mérésügyi témákban. 1978-ban Magyarországon megjelent *Ungváry Rudolf* „*Teauruszfordítási irányelvek*” c. műve, amely az akkori *Kohó- és Gépipari Tudományos Informatikai és Ipargazdasági Központnál* készülő hat *NTMIR*-teaurusz orosz nyelvű bázisteauruszának fordítási irányelveit tartalmazta. A teaurusz-szerkesztéssel foglalkozó dokumentumok magyar fordításait az *OKDK* az *NTMIR* normatív-műszaki dokumentumaiból összeállított köteteteiben (*NTP MCNTI 15–76*, 1978) is kiadta. A Szovjetunió megszűnésével azonban az *NTMIR* is befejezte szakmai tevékenységét, így 1985-ben ugyan még sor került egynyelvű szabványuk magyarországi honosítására (*KGST SZT 174*), azonban a többnyelvű irányelvek átvétele már nem – és azóta sem – történt meg.

Így az *MSZ 3418* (1987) – bár nemzetközi társainál jóval részletesebben írja le a relációk fajtáit - első és második kiadása is csak az egynyelvű teauruszokat szabályozza, tehát nincs többnyelvű teauruszokkal kapcsolatos magyar szabványunk, ami azért is kár, mivel a *KGST*

államok többnyelvű teauruszszabványai megegyeznek az *UNESCO* irányelvekkel és az *ISO szabvánnyal*, mivel tartalmukat illetően független szerkesztés eredményei. Az *NTMIR* szabványai tehát részben tovább öröklődtek hazai szabványunkban, valamint az orosz tudományos és műszaki információs intézet, a *VINITI*-ben keretei között (*Sonnenvend*, 2013).

3.2.2 Nyugat-Európa teaurusz szabványai és irányelvei

Derek Austin és *Dale* nagy hatással voltak a nyugati szabványokra, ők az *UNESCO* egynyelvű teauruszaira vonatkozó irányelv első (1974) és második kiadásának (1981) szerkesztésében is részt vettek, majd az általuk kidolgozott elméleteket később beépítették az egynyelvű teauruszokra vonatkozó nemzetközi kiadásba, az *ISO 2788*-ba is (1986). Emellett *Austin* a *PRECIS* indexelő rendszerrel is foglalkozott, mely egy szövegen belüli szerepet megőrző osztályozási rendszer. Az itt szerzett tapasztalatai alapján a nemzetközi szabványba is átültette az összetett kifejezésekkel kapcsolatos pontosabb szabályait és bevezette a szisztematikus teaurusz és a facetta analízis fogalmát.

Az egynyelvű teauruszokra vonatkozó *ISO 2788* szabványt először 1974-ben publikálták, majd 1986-ban jelent meg ennek második kiadása. Számunkra ez a kiadvány azért érdekes, mert a többnyelvű teauruszokra vonatkozó *ISO 5964*-ben már nem ismételték meg a korábban már az 1974-es szabványban rögzített a többnyelvű rendszerekre is jellemző gondolatokat.

A többnyelvű teauruszokhoz kapcsolódó *UNESCO* irányelvek kialakítását 1974-ben az *International Scientific Symposium*-on kezdeményezték, az így megjelent dokumentumot (1976) tartjuk hivatalosan a szakterület szabványosítási kiindulópontjának. 1980-ban megjelent az *Austin* által szerkesztett javított változat, amely 1985-ben a többnyelvű teauruszok szerkezetére vonatkozó *ISO 5964* szabványt eredményezte. Ezeket a nemzetközi szabványokat számos ország fogadta el sajátjának, köztük Franciaország, ahol már ugyanebben az évben elkészült *A relációk jelölése egy és többnyelvű teauruszokban*” c. szabvány. Később pedig Németország és az Egyesült Királyság is felzárkózott, ezért az Egyesült Királyság szabványai a *BS 5723* (1987) és a *BS 6723* (1985) megegyeznek az *ISO 2788* és az *ISO 5964* szabványokkal.

Az *ISO* két teauruszokról szóló szabványát 2008-ban felülvizsgálták, kiterjesztették és egyetlen új *ISO 25964* szabvánnyá alakították, ami *„Teauruszok és interoperabilitás más szótárakkal”* címen jelent meg. A szabvány első része *„Teauruszok az információlekérdezéshez”* egynyelvű és többnyelvű teauruszra vonatkozó nemzetközi

szabványtervezet 2011-ben jelent meg. Míg a szabvány második része, melynek témája az „*Interoperabilitás más szótárakkal*” jelenleg is fejlesztés alatt áll.

Egy másik nemzetközi szervezet, az *IFLA (International Federation of Library Associations and Institutions)*, a tezauruszok kialakításával kapcsolatos irányelveik megalkotásával képviseltette magát a területen. 1992-ben az *IFLA* egyik munkabizottsága elkészítette „*A tartalmi feltárás lexikai egységeire vonatkozó szabványtervezet*” amiben részletesen szabályozták, hogy milyen szerkezetű tételt alkossanak az állományba felvett tárgyszavak és deszkriptorok (*IFLA Guidelines*, 1992).

2005-ben az *IFLA* megjelentette többnyelvű tezauruszokra vonatkozó iránymutatását, majd 2009-ben kiadták ennek átdolgozott változatát. Jelen dolgozat is ezen *IFLA* irányelv alapján ismerteti majd a többnyelvű tezauruszok szerkezeti elemit és kialakításuk lehetőségeit.

4. A TÖBBNYELVŰ TEZAUROSZOK FOGALMI MEGHATÁROZÁSAI ÉS SZERLEZETI ELEMEI

4.1 A többnyelvű tezauruszok fogalma a szakirodalomban

Az 1974-es UNESCO szabvány tervezetben a többnyelvű tezauruszok meghatározása elsősorban a gyakorlati munka szempontjait tartotta szem előtt: „*A többnyelvű tezaurusz segítségével az osztályozott dokumentumok tartalma legalább két természetes nyelven azonos módon képviselhető. Segítségével az osztályozó az általa ismert nyelveken osztályozhatja és keresheti a dokumentumokat.*” (Ungváry, 1981. 10. o.)

Ugyanitt rögzítik, hogy a többnyelvűség első feltétele az, hogy a bevezetést, meghatározásokat, magyarázatokat stb. egynél több természetes nyelven kell szerepeltetni a tezauruszban. A második feltétel, hogy úgy kell a deszkriptorokat meghatározni és a dokumentumokban szereplő különböző nyelven kifejezett azonos tartalmak kifejezésére használni, hogy biztosítva legyen a tartalom egységes jelölése.

Soergel definíciója tiltott kifejezések meglétét említi, ezek az osztályozás szempontjából érdektelen kifejezések, amelyeket más néven negatív deszkriptoroknak neveznek. A meghatározásban a Dahlberg (1975) által normatív tevékenységként definiált szempontok érvényesülnek:

„*Az információkereső tezaurusz kifejezések és/vagy egyéb jelölések/vagy szimbólumok/olyan jegyzéke, amelyben kifejezik az ezen elemek közötti kapcsolatokat is, ha az alábbi feltételek teljesülnek: a jegyzék jelentős mértékben tartalmaz tiltott kifejezéseket, amelyek deszkriptorként nem használhatóak; igyekezzenek biztosítani a terminológiai ellenőrzést.*” (Soergel, 1974. 294. o.).

A Rolling (1976) által használt meghatározás így szól: „*Azonos vagy hasonló fogalmakat képviselő deszkriptorokat két vagy több nyelven tartalmazó tezaurusz.*”

Ebben a megfogalmazásban Soergel-ével ellentétben nincs szó terminológiai ellenőrzésről, valamint a meghatározás a nemdeszkriptorok fogalmát sem rögzíti. Ennek az eltérésnek a hátterében, a korábbiakkal ellentétben fontos kiindulási különbség rejtőzik, mivel, ha válogatott kifejezésekről beszélünk, akkor abból az következik, hogy nem minden kifejezés lehet a tezaurusz lexikai egysége, ami alapvetően ellenkezik a természetes nyelvű rendszerekkel. Azonban meg kell jegyeznünk, hogy a tezauruszoknak nem céljuk az összes lexikai egység gyűjtése. Mikrotezauruszok esetében egy bizonyos szakterületen belül, míg makrotezauruszoknál a cél az, hogy széles spektrumban tartalmazzák a szakterminológia

kifejezéseit, tehát a releváns fogalmakat. Ezért véleményem szerint a válogatás helyett az ellenőrzésen, vagyis a dezskriptorok kitüntetett szerephez juttatásán kell, hogy legyen a hangsúly.

Az Austin által szerkesztett (1977) UNESCO irányelvben is ez a felfogás érződik: „A többnyelvű teaurusz meghatározható célja vagy szerkezete szempontjából. Célját tekintve terminológiailag ellenőrző eszköz, melyet a dokumentumok, az osztályozók, vagy a felhasználók természetes nyelve és egy szabályozottabb „rendszer nyelv” szótára többnyire a teljes teaurusz részhalmaza. Az egynyelvű teaurusztól abban különbözik, hogy segítségével a fordítás egynél több nyelvről végezhető és fordítva. Szerkezetét tekintve az ismeretek meghatározott területére vonatkozó, szemantikailag kapcsolódó kifejezések ellenőrzött és válogatható szótára. A teauruszban a fogalmakat általában egynél több nyelv kifejezései képviselik. A többnyelvű teauruszt vagy a többnyelvű osztályozásra és keresésre, vagy csak e feladatok egyikéhez készítik.”

A „szabályozott rendszer nyelv” kifejezés azt jelzi, hogy az egyes kifejezések komplett struktúrában kapnak helyet, de ezzel nem zárnak ki bizonyos terminológiai egységeket a teauruszból, így az értelmezés hű marad a természetes nyelven alapuló jelzőhöz, ami a teauruszok sajátja. A „szemantikailag kapcsolódó kifejezések” elem ennek gyakorlati megvalósulására, a relációk használatára utal, de csak utal és a meghatározásban ez nincs pontosan kifejtve. Mivel a szerzők a definíció megalkotásakor az ISO 2788-ból (1974 indultak ki így az egy- és többnyelvű teauruszra egyformán vonatkozó jellemzőket itt nem ismételték meg.

Az 1980-as – szintén Austin által szerkesztett – UNESCO irányelvben ismét használják a „válogatott kifejezések” összetételt, szemben az ISO/DIS 5126 tervezettel. Pozitívum, hogy a 80-as iránymutatásban jóval nagyobb hangsúlyt kap a fogalmak közötti *kapcsolatok* sokféleségének kifejezése: „...egynél több természetes nyelvből válogatott kifejezéseket tartalmazó teaurusz. Nemcsak a kifejezések közötti [egy nyelven belüli] kapcsolatokat, hanem az összes érintett nyelv [közötti] ekvivalens kifejezés[é]t is feltüntetik benne.”

A KGST államok nemzetközi tájékoztatási rendszerében elfogadott meghatározás lényegesen általánosabb, nem beszél se válogatásról, sem dezskriptorokról és nemdezskriptorokról, csak lexikai egységekről és ellenőrzésről. Ez a definíció – az UNESCO irányelvekkel ellentétben – a többnyelvű és egynyelvű teauruszok közös elemeit csak az egynyelvű teaurusz meghatározásában fogalmazza meg és utal a teauruszok közötti *kölcsönös kapcsolatra*:

„A többnyelvű információkereső teaurusz lexikai egységek olyan ellenőrzött és változtatható szótára /egymással kapcsolatban álló szótárak összessége/, amely két vagy több természetes nyelv szókincsén alapul, tükrözi a lexikai egységek közötti szemantikai kapcsolatokat, és információk feldolgozására és keresésére szolgál.” (Ungváry, 1981. 13. o.)

Klücklich is a teauruszban megjelenő kapcsolatokra építi definícióját:

„Többnyelvű teauruszok esetében olyan egynyelvű teauruszokról van szó, melyeknek azonos jelentésű szavait a nyelvi akadályok áthidalása érdekében egységesen fordítókódokkal kapcsolják össze, figyelembe véve a szinonim, hierarchikus és asszociatív kapcsolatokban fennálló különbségeket.” (Klücklich, 1974, 74. o.)

Klücklichnél megjelenik a „fordítókód” kifejezés, de nemcsak ő emeli ki ezt a fogalmat a saját meghatározásában, hiszen Rosenbaum (1974) is ennek a közvetítő eszköznek a szerepét hangsúlyozza munkájában, ahol a KGST-tagállamok többnyelvű teauruszának kidolgozási módszereivel foglalkozik. A továbbiakban a gépesítési feladatok erősen hatottak a megfogalmazásokra, Wolff-Terroine definíciójában a többnyelvű teaurusz helyett már a többnyelvű rendszer kifejezést használja:

„A többnyelvű rendszer adott számú munkanyelvből áll, melyek közül bármelyiket választva forrásnyelvül /X/ a dokumentumok tartalmának kifejezésére, ezt a tartalmat egy másik munkanyelven /Y/ is ki lehet fejezni.” (Wolff-Terroine, 1976, 152. o.)

Vagyis a szerző, az általa meghatározott többnyelvű rendszert elemeire bontott összetevőkből „csillagszerűen” felépítve, a csillag közepébe egy közvetítő nyelvet/teauruszt állítva készíti el. Az értelmezések tehát egyre jobban közelítenek egyfajta kapcsolatközpontú rendszerszemlélethez és visszaszorul a kisebb egységekre bontott meghatározások jelenléte. De mi a helyzet a legújabb irányelvekben?

A 2009-es IFLA irányelvben ez található: *„A többnyelvű információkereső teaurusz a kifejezéseket és a köztük lévő összefüggéseket is egyaránt megjeleníti több, mint egy nyelven.” (IFLA Guidelines, 2009, 2. o.)*

Ez a megfogalmazás szintén elég általános, de az iránymutatás jelentősége nem is ebben, hanem a benne rögzített szimmetrikus és nem-szimmetrikus teaurusz szerkezeti kialakítását részletező, példákkal gazdagon ellátott ismertetésében rejlik, ami a fejezet további részében kap helyet.

4.2 A többnyelvű teauruszok struktúrája és szerkezeti elemei

Egy igazi többnyelvű teaurusz teljes fogalmi és terminológiai készletet kínál, minden képviselt nyelven. A legfontosabb az, hogy egy teljesen kidolgozott teaurusz struktúrát mutat be – vagyis szemantikai kapcsolatot az egyenértékű kifejezések között illetve hierarchiát és rokonságot – a teaurusz minden egyes nyelvén, így a felhasználó utánanézhet, hogy melyik nyelvi változat a legmegfelelőbb számára, és optimális esetben az értékes szemantikai információk azonos mennyiségét kapja. A teaurusz befogad egy forrásnyelvet és biztosítja a deszkriptorok ekvivalenseit más nyelveken, de a teljes nyelvi struktúrát a nyelvi egyenlőség szempontjából nem garantálja (*Hudon, 1997*).

A többnyelvű teaurusz tehát egy-egy szakterület több, mint egy természetes nyelven kifejezett, teljes fogalmi gyűjteménye, ahol a fogalmak közötti kapcsolatokat is feltüntetik. A teauruszban ezek a relációk és fogalmak azonos értékűek. A rendszer alapját képező fogalmi struktúra nyelvi, kulturális és tudományos behatások miatt folyamatosan változik, ezért fontos a teauruszok bővíthetőségének biztosítása. A teauruszban a fogalmak lexikai egységek formájában vannak jelen, amelyek lehetnek deszkriptorok és nemdeszkriptorok (*Barátné Hajdu és Babiczky, 1998*).

4.2.1 A deszkriptorok és szerepük

Az *IFLA irányelvében* (2009) a deszkriptort azzal a kifejezéssel azonosítják, amit az indexelésnél arra használunk, hogy kifejezzünk vele egy fogalmat. A deszkriptor az információk leírására és visszakeresésére közvetlenül alkalmazható kifejezés. A deszkriptorok általában egyes számú főnevek, kivéve, ha a kifejezés gyűjtőfogalom, vagy csak többes számban használatos, vagy ha az egyes számú és a többes számú alak jelentése között különbség van. Ez az előnyben részesített kifejezés lehet köznév, tulajdonnév, számnév és betűszó is, amennyiben a rövid alakot választjuk, zárójelben közölni kell a teljes alakot és utalni kell rá. Esetükben az egyes nyelveken megszokott jellemző formát kell alkalmazni, így például az angolban a gerund alakot, a hollandban az infinitívet, a románban a főnevekből származó igéket. A deszkriptor lehet egytagú és többtagú kifejezés is. Alapesetben a többtagú kifejezés akkor használatos, ha a jelentés nem következik az összetevők jelentéséből, vagy ezek túl általánosak így nem jellemzik kellően specifikusan a tartalmat (*Barátné Hajdu és Babiczky, 1998*).

A deszkriptor fogalmát *Calvin Mooers* vezette be, 1956-os munkájában. Az indexeléshez használt kifejezéseket nem pusztán szavaknak tekintette, hanem, akárcsak az osztályozási jelzeteket fogalmak reprezentánsainak. A természetes nyelv indexelésnél használt szavai, ha deszkriptorként funkcionálnak, egyrészt szavak (a fogalmak „címkei”) másrészt jelentésegységek, és a jelentésüket meghatározás írja le (a jelentésük rendszerről rendszerre változhat). Ezt gondolta tovább *Charles R. Bernier* és *Paul Needham*, akik a deszkriptornyelvek szülőatyjainak tekinthetők, amikor a természetes nyelv, ellenőrzött, szabványosított változatával (deszkriptornyelv) az eredeti, természetes nyelvnél jobb fogalomreprezentációkat valósítottak meg (Ungváry, 2001).

A tezauruszok lexikai egységeinek ideális számáról nem egyeznek a vélemények. *Rudolf* például 5000-ben adja meg a maximális deszkriptorszámot, az ezt meghaladó rendszerekben – megállapításai szerint – rohamosan csökken a keresés minősége. Ugyanezt az értékelést *Stephan* 2000-3000 deszkriptorban adja meg (Ungváry, 1982).

Ehhez képest az *Eurovoc 4.3. kiadása* az új bulgáriai és romániai régiókkal együtt összesen 6797 lexikai egységet tartalmaz. Nyilvánvaló hogy ez az optimum függ, az adott szakterület jellegétől és nagyságától, a határterületek jellegétől, fontosságától és az alkalmazandó információkereső rendszer jellegétől és logikájától. Így tehát véleményem szerint kijelenthető, hogy a deszkriptorok számának nincs meghatározott optimuma.

4.2.2 A nemdeszkriptorok és szerepük

A nemdeszkriptor a deszkriptor szinonim vagy szinonimnak tekintett kifejezése, ami az információk leírására és visszakeresésre önmagában nem, csak a vele összekapcsolt deszkriptor figyelembevételével használható (Barátné Hajdu és Babiczky, 1998). Az *IFLA irányelvében* (2009) a nemdeszkriptor olyan kifejezésként szerepel, amelyet az indexelés során ugyan nem használunk arra, hogy kifejezzük vele az adott fogalmat, de úgy működik, mint egy hivatkozási pont, ami arra utal, helyette melyik kifejezést kell alkalmazni.

Léteznek olyan többnyelvű tezauruszok is, amelyek nem tartalmaznak nemdeszkriptorokat, azonban ezek a rendszerek éppen ezen tulajdonságukkal szembemennek a természetes nyelvűség alapjaival. A nemdeszkriptorok nélküli rendszer, megnehezíti a feldolgozó és/vagy a kereső dolgát, hiszen tevékenységük végzése közben egy az egyben tisztában kell lenni a rendszer által használt kifejezésekkel, ezzel gyakran el kell szakadni a szerző szóhasználatától és lényegesen nagyobb szakterületi tájékozottságot igényel.

Az UNESCO irányelvek (1976) szerint a szláv nyelvek esetében kifejezetten gyakori, hogy mellékneveket kell felvenni a tezauruszba, azonban ebben az esetben a kifejezést nemdeszkriptorként kell szerepeltetni. Azonban az is kikötés, hogy ebben a struktúrában a melléknév az egyik nyelven se legyen ekvivalens egy másik nyelv főnévével.

4.2.3 Egyenlőségi, hierarchikus, és asszociatív kapcsolatok és a jelölők

1964-ben a dániai *Elsinore*-ban Mills felismerte a relációk rendszeralkotó fontosságát. Az egynyelvű tezaurusz lexikai egységei között megkülönböztetünk szinonima (L-H) relációt amelyben az L rámutat a deszkriptorra a H pedig nemdeszkriptort helyettesít. Továbbá különbséget teszünk generikus vagy más néven fölé-és alárendeltségi reláció (F-A), egész-rész (T-P) reláció, rezultáns és előfeltétel (R-E) reláció, valamint rokonsági (X) reláció között (Barátné Hajdu és Babiczky, 1998).

A többnyelvű tezauruszokban ezekből három típust alkalmaznak, az egyenlőségi a hierarchikus és az asszociatív kapcsolatot. A hierarchikus relációk összekapcsolják a fölérendelt kifejezéseket és az alárendelt kifejezéseket, ez a típus kiterjed a polichierarchikus kapcsolatokra is, vagyis azokra az esetekre, amikor egy kifejezésnek több, mint egy fölérendeltje van. A fölérendelt vagy tágabb kifejezés (BT), egy specifikus és egy általános deszkriptor közötti kapcsolatra utal. Ha nincs fölérendelt kifejezés, akkor ún. csúcsdeszkriptorról vagyis a hierarchiában legmagasabb szintjén elhelyezkedő kifejezésről van szó. A szűkebb kifejezés vagy alárendelt (NT), egy általános és egy specifikus deszkriptor közötti kapcsolatot jelez. A deszkriptorok közötti asszociatív kapcsolatot az (RT), azaz kapcsolódó kifejezés rövidítés vezeti be. A kapcsolat többféle lehet, amelyek közeli, de nem hierarchikus kapcsolatot tükröznek, ilyen lehet például az ok vagy okozat, közeg vagy eszköz, hierarchia, idő vagy térbeli egymásutániség, alkotórész, jellemző tulajdonság, elhelyezkedés stb. összefüggés. A deszkriptorok és nemdeszkriptorok közötti egyenlőségi kapcsolatokat az (UF) és a (USE) rövidítések jelzik, mindkét jelölő egy deszkriptor és egy nemdeszkriptor között áll. Az első rövidítés az előnyben nem részesített lexikai egység – nemdeszkriptor - előtt áll, míg a második eset azt a kifejezést mutatja meg, amelyet helyette használunk, vagyis a deszkriptort. Ezek az egyenlőségi kapcsolatok többféle viszonyt kifejezhetnek, jelezhetnek közeli szinonimákat vagy hasonló jelentésű szavakat, rokon értelmű vagy azonos jelentésű szavakat, hasonló jelentést, antonímia vagy ellentétes jelentés (Koreny, 2005).

A többnyelvű tezauruszok készítésénél sokan kimondták, hogy a generikus kapcsolatok összehangolására szükség van, de a többi kapcsolat összehangolására (nyelvenként) eleve nincs

szükség, de lehetőség sem. Ungváry (1981) szerint a valóságban fordított a helyzet: éppen a nem-faj kapcsolatok közelíthetők a legkevésbé egymáshoz, ha pontatlan ekvivalenciáról van szó.

A tezauszban a lexikai egységekhez fűzött jól ismert jelölők, jelzik a kifejezések természetét és a köztük lévő kapcsolatot. Mivel nincs a többnyelvű tezauszokra vonatkozó magyar nyelvű szabványunk a továbbiakban én is a nemzetközi rövidítéseket (BT, NT, stb.) fogom alkalmazni (4.1. táblázat). Természetesen ezeken kívül léteznek további az egyes tezauszokra jellemző speciális jelölők is, például az *EuroVocban*, a definíciós megjegyzések előtt (DEF), míg a történeti megjegyzések előtt (HN) rövidítés áll.

angol	német	francia	jelentés
USE: Use term ...instead	BS: Benutze	EM: Employer (Voir)	Lásd
UF: Use for..	BF: Benutzt für	EP: Employé pour	Lásd honnan
BT: Broader term	OB: Oberbegriff	TG: Terme générique	Tágabb kifejezés
NT: Narrower term	UB: Unterbegriff	TS: Terme spécifique	Alárendelt kifejezés
RT: Related term	VB:Verwandtnr Begriff	TA: Terme associée	Kapcsolódó kifejezés
SN: Scope Note	D: Definition	NE: Note explicative NA: Note d'application	Megjegyzés

4.1. táblázat: *A lexikai egységek között használt általános jelölők és rövidítésük (s.k.)*

4.3 A többnyelvű tezauszok szerkezeti mélystruktúrái

A többnyelvű rendszerben a nyelvi kompatibilitást csak egy közvetítő rendszer biztosíthatja, még akkor is, ha két természetes nyelvű rendszerből indulunk ki. A közvetítő rendszereknek tehát két típusuk van, a közvetítő és az átkapcsoló nyelvek. A gyakorlatban ezt a megkülönböztetést azonban mégsem használják ennyire egyértelműen. Közvetítő nyelv csak az lehet, amelyre információvesztés nélkül lefordítunk bármely információkereső nyelven kifejezett fogalmakat. A közvetítő nyelv inkább elméleti fogódzó, hiszen a szemantikai összetevők elemzésén, vagyis a kifejezések képviselte fogalmak vizsgálatán alapszik. Az eljárás során a fogalmakat lebontjuk a nyelvészetben szemantikai atomoknak/ szemémáknak nevezett alkotóikra, majd a folyamatot megfordítjuk (Wolf-Terroine, 1976)

Ungváry szerint (1981) átkapcsoló nyelveken inkább technikai eszközöket értünk, segítségükkel továbbíthatóak az egyik információs rendszerből a másikba a kérdések. Függetlenül attól, hogy természetes nyelven, vagy a továbbító rendszer által használt információkereső nyelven vannak-e ezek a kérdések kifejezve. Az átkapcsoló nyelvek a közvetítő nyelv eljárásának ellentét valósítják meg, vagyis a tag rendezés során, olyan

fogalmakat határoznak meg, amelyek nagyobb kategóriákba helyezhetők, így kívülről közelítik meg a forrásnyelv fogalmait.

A közvetítő nyelv a közvetítő rendszeren belül kap helyet, ennek a rendszernek a meghatározását eredetileg a fordítás transzformációs elméletében a generatív nyelvelméleti modellben rögzítette *Noam Chomsky* (1985). A modell azon a feltételezésre alapul, hogy a különböző nyelvek az eltérő felszíni kifejezési sajátosságok mögött, az egyedi sajátosságaiktól sokkal függetlenebb közös minden nyelvre jellemző mélystruktúrát tartalmaznak. Így bármelyik nyelvpár összevetésekor meghatározhatóak olyan tartalmi kategóriák, amelybe ennek a két nyelvnek minden valóságos eleme belefoglalható. Gyakorlatban a tárgy rendezés, vagyis az átkapcsoló nyelvek rendszere könnyebben megvalósítható, ezért nemzetközi szinten is a csúcstezauruszok kerülnek előtérbe.

A többnyelvű rendszerekben a mélystruktúrába való átlépés, maga a fordítás. A mélystruktúrák szintjén, a forrásnyelv és a cél nyelv közötti fordításkor sajátos közvetítőrendszert kell használni, ezek lexikai egységei szemantikai összetevőkből épülnek fel. *Neelameghan* (1976) a közvetítő rendszerek értelmezésében *Ranganathan* fogalmi kategóriáiból indul ki, szerinte az egyik rendszerből a másikba való áttérés egy abszolút szintaxison alapszik.

Harris (1978) szerint a tezauruszok fordításakor olyan feszültségek jelentkeznek, amelyek természetes nyelvek közötti fordításkor nem állnak fenn. A tezaurusz nyelvén megfogalmazott „szöveg” ugyanis egy formális fogalmi rendszert tükröz, ami a relációjelekkel kifejezett bináris kapcsolatokon alapul. A tezaurusz fordításakor tehát nem csak a kifejezéseket, de a kifejezések közötti összekapcsoló struktúrát – a relációkat – is fordítani kell valamilyen módon. A relációszerkezet, azonban nem lesz mindig azonos a két nyelvnél.

A valóságban a mélystruktúrák ezen közvetítő szerepe, csak rejtve, a fordítás intuitív folyamatába ágyazva jelenik meg. A közvetítő nyelv a többnyelvű tezauruszban (ha csak rejtett formában is) de mindig jelen van, azonban explicit megvalósulása az automatikus többnyelvű terminológiák készítésénél fajsúlyos és megjelenése is csak leegyszerűsített formája a közvetítő nyelveknek. A közvetítő nyelv több ilyen leegyszerűsített helyettesítési formája is létezik, az egyik a fogalomazonosító (vagy más néven fordítókód) a másik valamilyen természetes nyelv formájában is megjelenhet (*Ungváry*, 1981).

4.3.1 A fordítókód

A fordítókódoknak csak az automatizált rendszerekben funkcionálisan van létjogosultságuk. A fordítókódok pusztán közvetítő elemek. Már az egynyelvű teauruszok gépi változataiban is van szerepük, de csak a többnyelvű rendszerekben válik ez jelentősebbé. A kettőnél többnyelvű rendszerekben a fordítókódok a fogalomhoz rendelve nem közvetlenül a másnyelvű ekvivalenciákat jelölik, hanem csak a kódolt címeket, amelyek egy nyelvtől független közbelső file pozíciójára utalnak. Ez a file mindazon előfordulásokat tartalmazza, amelyek alatt az ide kapcsolódó többnyelvű megfelelők vagy szóösszetételek megtalálhatóak. Az egyes nyelveknek tehát külön teauruszaik vannak, amelyeket kapcsolófile köt össze. Ebből is adódik a teaurusz további nyelvvél való bővíthetőségének viszonylag egyszerű mivolta. A fordítókód további jelentősége megnyilvánulhat az arab és a latin ábécét használó rendszerek közötti információcserében is. A gépesített teauruszkészítésben a fordítókódok leegyszerűsített közvetítőnyelvi szerepe mellett, ki kell emelnünk a hibaellenőrző funkciójukat is (Ungváry, 1981).

A fordítókódokat bizonyos esetekben hierarchikusan szerkesztik meg, ez a gyakorlat zárt teauruszokat eredményezhet. Ebben az esetben a csúcsgalmainak kódja képviseli az alapszámot, és ennek decimális vagy alfanumerikus alosztásai lesznek az alárendelt deskriptorok. Ilyenkor akár 12 jegyű képződmények is keletkezhetnek, amelyek következtében a teauruszok csak nehezen bővíthetők. Ezek a terjedős fogalomazonosítók számtalan hiba forrásai is lehetnek, mivel automatikus ellenőrzésükre felépítésük tartalmi jellege miatt nincs mód. Egy másik megoldás az ugrószámok képzése, ebben az esetben a kódok a finom szakcsoportot jelölik és azért, hogy ezekbe új lexikai egységeket is felvehessenek, üres helyeket hagynak, hogy ezt később kitölthessék. A módszer hátránya, hogy egyes helyeken a rendszer tartalma telítődik, más helyen pedig sosem telik be. A legjobb megoldás a folyamatos számozású fordítókódok alkalmazása. Ezekben vagy az előírt oszthatóságuk alapján, vagy ellenőrzőszám alkalmazásával biztosítják az automatikus hibaellenőrzést. Így nem áll fenn a kód elírásának lehetősége. (Ungváry, 1981)

A fordítókódok jelentősége az arab és latin ábécét használó nyelvek között is megnyilvánul, mivel a kódszámok sorozata alapján az egyik rendszer kulcsszavai bármely más rendszerben használt kulcsszóra lefordíthatók. Vagyis ha a kifejezések leírásában a rendszeren belül hiba történik, adott fogalom és ekvivalensei mégis megtalálhatóak lesznek. Éppen ezért, így nem jelentenek problémát azok az arab karakterek, amelyek bár rendkívül különböznek a

latinban használt betűktől, de adott nyelven már csak a központosítás megváltoztatásával is teljesen más jelentésű kifejezést kaphatunk, illetve azok a karakterek sem, amelyek a latin ábécével ki sem fejezhetőek (4. 2. táblázat).

ز	ذ	ع	ظ	س	ف	ب
z r	d d	‘ ġ	ʔ z	š s	q f	y ʔ t b

4.2. táblázat: *Központosításból eredő nehézségek bemutatása, betűpárokba állítva (s. k.)*

4.3.2 A természetes nyelv, mint közvetítő nyelv

A tezaurusz készítésekor, még ha párhuzamosan is készítik - különösen, ha kettőnél többnyelvű tezauruszról van szó – az összetevő nyelvek egybevetése után egy jó felbontóképességű közvetítőnyelvet kell kialakítani. Vagyis munkanyelvként ki kell jelölni valamelyik nyelvet. Az *UNESCO irányelvei* (1976) megkülönböztetnek azonos státuszú és vezérnyelvet tartalmazó rendszereket. Az első esetben a munkanyelvet (vagyis az összekapcsolás céljára létrehozott nyelvet) egyszerűen csere-nyelvnek nevezi. A második esetben az összekapcsolást az egyetemes közvetítő nyelv a vezérnyelv végzi.

A két-és többnyelvű rendszerek közötti minőségi különbség hátterében a közvetítő rendszer szerepének növekedése áll. A többnyelvű rendszerekben az összetevő nyelvek párjainak egyeztetése után az egyes nyelvpárok közötti kölcsönös összhang megteremtése mindenképpen megköveteli egy közvetítő eszköz használatát. Ennek a közvetítő nyelvnek a kiválasztása nagy jelentőséggel bír. *Springl* (1980) szerint, a többnyelvű rendszer szerkezeti finomságát, ez a legkisebb nyelvi kifejezésű nyelv határozza meg.

5. TÖBBNYELVŰ TEZAUROSZOK KIALAKÍTÁSÁNAK LEHETŐSÉGEI

A többnyelvű tezauszok kialakításának kezdeti lépései az egynyelvű változatokéhoz nagyban hasonló. Az előkészítő munka során meg kell határozni a célokat és a tezausz fő témaköreit és rögzíteni kell a továbbfejleszthetőség lehetséges módjait. A következő lépés az induló szóanyag gyűjtése, amelynek többféle módszere lehetséges, ez történhet a szakszövegek vagy az adott területre vonatkozó tájékoztatói eszközök elemzésével, tapasztalati módszer segítségével illetve meglévő osztályozási rendszerek és szókészletek felhasználásával is. Majd az így létrejött szókészlet elemzésével kiválasztjuk a legfontosabb deskriptorokat és elkészítjük a betűrendes mutatót (*Barátné Hajdu és Babiczky* 1998).

Ezen előkészítő munkálatok a többnyelvű tezauszok esetében azonban kiegészülnek a célstruktúra megválasztásának feladatával, mivel ez nagyban befolyásolja majd a későbbi folyamatokat. Tehát a munkacsoport lehetőségeit figyelembe véve ki kell választaniuk, hogy szimmetrikus vagy nem szimmetrikus tezausz elkészítésébe vágunk bele.

5.1 A nemszimmetrikus és a szimmetrikus tezausz jellemzői

Az egyik alternatíva egy nem azonos, nemszimmetrikus tezausz szerkezet kialakítása, ahol a deskriptorok száma nem szükségszerűen egyezik meg minden egyes nyelven és az sem, ahogyan a deskriptorok egymáshoz kapcsolódnak a különböző nyelveken. A nemszimmetrikus tezauszban a forrásnyelvi deskriptor állhat önállóan egyenértékű célnyelvi kifejezés nélkül. Az *IFLA* irányelvében (2009) a szimmetrikus szerkezethez képest is eltérés van az előjárószavas kifejezések alkalmazásánál.

Hudon (1997) szerint helyesebbnek tűnik, ha elfogadjuk a nem szimmetrikus, nem azonos struktúrákat a többnyelvű tezauszokban. A deskriptorok számát azért kell változtathatóvá tenni az egyes nyelvi változatokban, mivel a fogalmaknak, amelyek egy adott kultúrában léteznek és egy bizonyos nyelven képviseltetik magukat, nem biztos, hogy létezik ekvivalensük egy másik kultúrában és az sem, hogy a megfelelő verbális alak áll majd rendelkezésünkre. Illetve a paradigmatiszka kapcsolatokat az alá-fölé rendeltségi viszonynál, például nem feltétlenül ismerik el érvényesnek minden természetes nyelven. Egy többnyelvű tezauszban, amikor két csúcs deskriptor vagy fölérendelt kifejezés pontatlan vagy részleges nyelvi megfelelést mutat, némileg eltérő jelentésük lehet és ennek következtében eltérő lesz az alárendelt kifejezés is. Azokban a többnyelvű tezauszokban, amelyekben a struktúra nem egyforma, nagyobb valószínűség szerint hitelesen tükrözik a világegyetem fogalmainak és kifejezéseinek képviseletét minden kultúrában és nyelven.

A leggyakoribb nézet, hogy a többnyelvű teaurusz minden egyes nyelvi változatának azonosnak kell lennie. Amennyiben adott teauruszban a deszkriptoroknak kizárólag egy egyenértékű megfelelője van minden egyes nyelven, valamint a deszkriptorok minden nyelven ugyanúgy kapcsolódnak egymáshoz, abban az esetben szimmetrikus teauruszról beszélünk. Ez a kapcsolat alapulhat teljes vagy részleges egyenértékűségeen is, illetve a nemdeszkriptorok száma eltérő lehet. A szimmetrikus teaurusz kialakítóinak célja a tökéletes egyezés létrehozása a deszkriptorok és a nemdeszkriptorok között, valamint az egyes nyelvi ekvivalensek között. Példával, ha X nyelven generikus reláció áll fenn A és B deszkriptor között, akkor generikus reláció áll fenn az egyenértékű A' és B' deszkriptorok között is, Y nyelven (IFLA, 2009).

Ennek következményeként megtörténhet és gyakran elő is fordul, hogy nyelvek közötti egyenértékűséget kényszeríteni kell, még ott is ahol, a valóságban nincsenek meg ezek az egyezések, így megkérdőjelezhető kapcsolati szerkezetek fordulhatnak elő. A szimmetrikus teauruszban a kifejezések jelentését gyakran szűkíteni, vagy tágítani kell ahhoz képest, amilyen az adott kifejezés a természetes nyelvi használatban, de még így sem lehet elkerülni a mesterséges „megalkotott” kifejezéseket, neologizmusokat. A szimmetrikus teauruszban elfogadhatatlan az a helyzet, amikor nincs ekvivalens, amely azonos jelentéssel bír a célnyelven. Ebben az esetben teaurusz alkotói általában eltávolítják a rendszerből ezt a „gazdátlan” forrásnyelvi kifejezést, különösen akkor, ha az nagyon speciális (Hudon, 1997).

5.2 A többnyelvű teaurusz kialakítása alulról-felfelé építkezve

Az IFLA irányelvében (2009) a többnyelvű teauruszok három kialakítási módszerét határozzák meg, az első lehetőség, hogy alulról-felfelé építkezve alakítsunk ki egy teauruszt. Ezen belül két variációt adnak meg: az elsőben, a munkát egy nyelvvel kezdjük majd hozzá adunk egy másik nyelvet vagy nyelveket. A másik opció, ha egyszerre több, mint egy nyelvvel kezdjük meg a struktúra kialakítását.

Különböző nyelvi változatok együttes fejlesztése egyenértékűséget biztosít minden nyelven. Minden nyelv felváltva válik forrásnyelvvé, majd célnyelvvé, ami gyakran mesterkelt és nem mindig egyező kifejezéseket eredményezhet. A teauruszban minden kultúra kifejezései hozzájárulnak az eszköz szerkezetéhez, az optimális összehangolás tevékenységét ebben a folyamatban „terminológia harmonizációjának” nevezik (Hudon, 1997).

A szakirodalom a 70-es évek közepétől azt hirdeti, hogy ez a párhuzamos (ab initio) teauruszkészítési módszer a legjobb. A párhuzamos készítésű teauruszok nemzetközi intézményeknél és világcégeknekél készülnek, mivel nehéz és költséges feladat hosszú időn át,

olyan szakértők egy csoportját összefogni és együtt tartani, akik értenek mind az adott szakterülethez, mind az adott (idegen és anya)nyelvhez és a dokumentáció szakterületéhez is. Azonban erre a módszerre jellemző a legtakarékosabb szógyűjtés, kevesebb terminológia keletkezik, így kevesebb homonima és pontatlan kifejezés is, ami rendkívül nagy pozitívum.

A természetes nyelvek jóval többek egyszerű szókészleteknél: hűen tükrözik az univerzum fogalmi rendszerét, amelyek eltérőek lehetnek az egyes kultúrákban vagy társadalmakban. A fogalmak különbségei gyakran megmutatkoznak a speciális entitások, az eljárások és kapcsolatok vonatkozásában. A teaurusz tervezők számára könnyebb lenne az élet, ha az összes használt kifejezés a többnyelvű hálózatban általános és egyszerű ismeretek formájában lenne feltüntetve, ez azonban nem így van. A modern teauruszokban a megalkotók azonban ennek ellenére is a nagyobb pontosságra és az összetett fogalmak képviselésére törekszenek (Hudon, 1997).

A gyakorlatban, tehát ez az eljárás is arra törekszik, hogy a szerkezet, minden nyelvi verziójában azonos legyen, azonban ez nagy valószínűség szerint azonban nem valósul meg indexelő szótárban egészében. Erre példa az (5.1 táblázat) angol „Education” kifejezés, amelynek jelentése tágabb, mint a francia „Éducation” kifejezés, vagy másképpen kifejezve az „Education” és „Teaching” kevésbé világosan különülnek el az angolban, mint az „Éducation” és az „Enseignement” francia kifejezések. Ez a tény hatással van ezekre a kifejezésekre és a deskriptorok alatt hierarchikusan alárendelt (szűkebb) kifejezések közti generikus relációra.

<i>angol (forrás nyelv)</i>	<i>francia (célnyelv)</i>
Education	Éducation Enseignement Formation
Teaching	Enseignement
Instruction	Enseignement Instruction

5. 1. táblázat: *A relációs hálózat bonyolultsága (Hudon, 2001):*

5.2.1 A tulajdonnevek megjelenítése

Dahlberg szerint (1989) minden egyes deskriptor, ami a teauruszban szerepel egy egyedi fogalmat jelenít meg. Ez lehet egyszavas kifejezés vagy több szavas összetett kifejezés. Elvileg a deskriptoroknak természetes nyelvi formájúnak kell lenniük, a teaurusz kifejezései nem lehetnek mesterséges eredetűek. Abban az esetben, ha adott deskriptor tulajdonnév

választanunk kell a következő három lehetőség közül. Vagy kiválasztjuk a tulajdonnév egyik nyelvi változatát és azt használjuk minden nyelven. Vagy egy tulajdonnevet használunk minden egyes nyelven, ahol az adott kifejezés szerepel, és az eredeti nyelvű tulajdonnevet használjuk ott, ahol a nyelvből ez a kifejezés még hiányzik. Vagy fordításokat használunk, amennyire ez ésszerű és lehetséges. Minden esetben a kiválasztott forma, megfelelő nemdeszkriptorokkal egészül ki, amihez ajánlatos az egyes nyelvek engedélyezett neveinek listáját használni. A (5. 2. táblázatban) az angol és a francia deszkriptorok az adott nyelven hivatalos formában vannak, a harmadik oszlopban a hivatalos holland deszkriptor található, ill. az angol és a francia kifejezések nemdeszkriptorként szerepelnek. Véleményem szerint többek között ez azért is lehet, mivel Hollandiában sokan használják a francia és az angol nyelvet is, utóbbi bizonyos tartományokban hivatalos nyelv.

<i>angol</i>	<i>francia</i>	<i>holland</i>
International Committee of the Blue Shield (ICBS)	Comité international du Bouclier bleu (CIBB)	Internationaal Comité van het Blauwe Schild UF International Committee of the Blue Shield (ICBS) UF Comité international du Bouclier bleu (CIBB)

5. 2. táblázat: Tulajdonnevek a többnyelvű tezauruszban (IFLA, 2009)

5.2.2 A homonimák kérdése

A homonimák nem csak egy nyelven belül fordulhatnak elő, hanem a nyelvek között is. Célszerű, ha egy vagy több homonimát helyettesítünk egy általánosan használt szinonimával, ha ez rendelkezésre áll, amelyet magyarázattal kell kiegészítenünk, ha ez szükséges. Illetve a homonimákat alkalmazása során, minden egyes jelentés értelmét tisztázni kell - a kétértelműség elkerülése végett - egy zárójeles minősítő hozzáadásával.

Példa az első esetre a „Cranes” angol kifejezés, amely a magyar nyelvben is teljesen egyező homonimapárt alkot, mivel „Cranes” (birds) vagy „Cranes” (lifting equipment) magyarul „Daruk” (madarak) és a „Daruk” (emelő berendezések) teljes egyenértékűi egymásnak. Az a tény, hogy a „Cranes” kifejezés homonima az angol nyelvben persze, nem feltétlenül jelenti azt, hogy ez más nyelveken is homonima lesz, így a holland nyelvben sem az. A nyelvek közötti homonimák rendszerint nem vezetnek félreértésekhez a többnyelvű tezauruszban, de kivételek is léteznek, éppen ezért a megjegyzések célszerűek vagy szükségesek lehetnek, mint ahogy ez a „Hijskranen” holland kifejezés alatt is látható (5. 3. táblázatban).

<i>brit angol</i>	<i>amerikai angol</i>	<i>holland</i>	<i>francia</i>	<i>magyar</i>
Cranes (birds)	Cranes (birds)	Kraanvogels	Grue (oiseau)	Daruk (madarak)
Cranes (lifting equipment)	Cranes (lifting equipment)	Hijskranen SN voor andere typen kranen, zie aldaar	Grue (appareil de levage)	Daruk (emelő berendezések)

5. 3. táblázat: *Homonimák a többnyelvű tezauruszban (IFLA, 2009)*

A tezaurusz megjelenítésének azon esetében, amikor a kifejezések egynél több nyelven vannak megadva alfabetikus sorrendben nem használunk megjegyzéseket. A gyakorlatban ennek a módszernek az alkalmazása nem javasolt, ha minden nyelv külön listán szerepel az jóval előnyösebb. Amennyiben ezt az opciót kell használni, a kifejezés mellett zárójelben elhelyezett nyelvmegjelölés az ajánlott A (5.4. táblázat) felső részben a nyelvek közötti ekvivalencia esetét és párhuzamait, míg az alsó cellában a kifejezések ezen típusú megjelenítését láthatjuk (IFLA, 2009).

Chance (angol) = Hasard (francia) / jelentésük: esély
Fortune (angol) = Chance (francia) / jelentésük: véletlen
Alfabetikus sorrend:
Chance (eng)
Chance (fre)
Fortune (eng)
Hasard (fre)

5.4. táblázat: *Homonimák az alfabetikus többnyelvű megjelenítésben (s.k.)*

5.2.3 Az összetett kifejezések kérdése

Összetett kifejezés, olyan kifejezés, amely több, mint egy szóból vagy összetett szóból áll. Az összetett szavak olyan szavak, amelyek több, mint egy szóból (gyakran főnévből) állnak és egy vagy több betűből álló kötőszó kapcsolhatja őket össze. Az összetett kifejezések általában prekoordináltak, vagyis előzetesen egyesítenek két vagy több egyszerű fogalmat egy összetett fogalommá. Ezen kifejezések használata a tezauruszban és a tárgyszólistákban egyre növeli a specifikusság szintjét. A specifikussággal párhuzamban nő a pontosság is, ami az adott fogalom precíz képviseletét eredményezi. Ezzel egyetemben pedig nő az eszköz hatékonysága is, mivel a rendszer a nem kívánatos, tárgykörbe nem tartozó dokumentumokat ki tudja zárni. *Taube Uniterm*-jének jellemzésénél is kiemelttem mekkora problémát jelenthet, ha egy rendszer nem képes összetett kifejezésekkel jellemezni az adott tartalmat, mivel így bizonyos speciális jelentések kiszorulhatnak, vagy torzulhatnak a struktúrán belül.

Az összetett kifejezések részeit az alábbi elvek szerint lehet megkülönböztetni. A középpont, vagyis a főnév, az az összetevő, amely azonosítja a fogalmak általános osztályát, amelyre a kifejezés egésze vonatkozik. „Printed indexes” kifejezés főnévi komponense az „indexes”. A „Hospitals for children” kifejezésben a „hospitals” ugyanezt a szerepet tölti be. Illetve a „Broadband” összetett szó, főnévi része a „band”. Az összetett kifejezések másik része, a minősítő. Az az egy vagy több komponens, ami arra szolgál, hogy szűkítse a kiterjeszthetőség középpontját és így megadja az alosztályokat. Ilyen például a „printed” melléknév a „Printed indexes” kifejezésben. Az előljáró-kötőszó-főnév összetételben, vagyis a „Hospitals for children” kifejezésben a „for children”, míg a „Broadband” kifejezésben a „Broad” tölti be ezt a szerepet. Általában, ha egy összetett kifejezés két vagy több kifejezésből áll egy nyelven, valószínűleg a vele egyenértékű idegen nyelvű kifejezések is ilyen formában lesz fellelhető (IFLA, 2009).

A különböző nyelvű deskriptorok formája a gyakorlatban azonban nem mindig egyezik, mert pl. az angol „Rail safety” kifejezés holland megfelelője, az előjárószavas „Veiligheid bij spoorwegen” kifejezés. Egy nem szimmetrikus tezaurusban lehetőség van a USE indikátor, mögötte pedig az összetevő deskriptorok plusz jeles kapcsolására (5. 5. táblázat). Ennek az eshetőségnek, viszont komoly következményei vannak a tezaurus szerkezetére (IFLA, 2009).

angol	holland
Rail safety	Veiligheid bij spoorwegen USE Veiligheid + Spoorwegen

5.5. táblázat: *megoldás az összetett kifejezések alaki eltérésére a nemszimmetrikus struktúrában* (IFLA, 2009)

Az összetett kifejezések előfordulásuknak két formája van, a jelzős kifejezések és az előjárószavas kifejezések és a jelzős kifejezések. Az utóbbiak jelenlétét lehetőleg kerülni kell a tezaurusban. Éppen ezért az angol „Carbohydrate metabolism” (szénhidrát anyagcsere) kifejezést részesítjük előnyben, a megegyező jelentésű „Metabolism of carbohydrates” kifejezéssel szemben. A jelzős kifejezések felépítése az angol és a magyar nyelven azonos, vagyis a jelző a főnév előtt elkülönülten helyezkedik el. A szavak elhelyezkedése a hollandban ezzel megegyező, azonban a két szó egybeolvad. Az olasz nyelvben a jelző a főnév mögött kap helyet és egy kötőszó is beékelődik közéjük, csak ezzel a harmadik taggal együtt alakul ki a jelzős kifejezés (példa a 5.6. táblázatban).

angol	holland	olasz	magyar
Library catalogues	Bibliotheekcatalogi	Cataloghi di biblioteca	Könyvtári katalógusok

5.6. táblázat: *a jelzős kifejezések formai különbségei az egyes nyelveken (IFLA, 2009)*

Az összetett kifejezések és a prekoordináció használata előidézi, hogy így nehezebb megállapítani az egyenértékűséget. Amennyiben a tezaurusban, az előjárószavas kifejezésnek még sincs közvetlen egyenértékűje a másik nyelven, ezt eltérő indikátor használat jellemzi a szimmetrikus és a nemszimmetrikus tezaurusokban (5.7. táblázat).

Megoldás a szimmetrikus tezaurusban:	
angol:	holland:
Accessories UF Accessories after the fact	Medeplichtigen
Megoldás a nemszimmetrikus tezaurusban:	
angol:	holland:
Accessories NT Accessories after the fact	Medeplichtigen

5.7. táblázat: *pontatlan ekvivalencia a szimmetrikus és a nemszimmetrikus tezaurusban (s. k.)*

5.2.4 Az ekvivalenciák

A nyelvek közötti ekvivalenciának három megközelítése van: szemantikai, kulturális és strukturális. A szemantikai és a kulturális szempont a kifejezések jelentésére és arra vonatkozik, ahogyan ezeket egy adott nyelven vagy kultúrában használják. Ez a két megközelítés csak a deszkriptorokkal kapcsolatban releváns, a legtöbb többnyelvű tezaurusban ugyanis nem tesznek kísérletet arra, hogy egyenértékű kifejezéseket rendeljenek minden nemdeszkriptorhoz. Ez a gyakorlat a neologizmusok elkerülése végett ajánlott. A deszkriptorok közötti interlingvális ekvivalenciának öt különböző fokát lehet megkülönböztetni az alulról-felfelé kialakított tezaurusokban (IFLA, 2009).

A pontos, a pontatlan, a részleges, az egy a többhöz egyenértékűséget, illetve azt az esetet, amikor nem létezik ekvivalens a másik nyelven. A pontos egyenértékűség, más néven nyelvek közötti szinonima, abban az esetben áll fenn, amikor X és Y nyelven is a deszkriptorok szemantikai és nyelvi egyenértékűséget egyaránt mutatnak (5. 8. táblázat).

német	angol	francia	holland	magyar
Schienenennetz	Rail network	Résau ferroviaire	Spoorwegnet	vasúthálózat
UF Eisenbahnlinie Eisenbahnstrecke	UF Railway line Railway track	UF Ligne ferroviaire Voie ferrée	UF Spoorlijn Spoorweg	UF vasútszakasz vasútvonal vasútvonalszakasz

5. 8. táblázat: *Pontos egyenértékűség az EuroVoc-ban (s. k.)*

A következő típus a pontatlan, vagy közeli egyenértékűség, amely a nyelvek közötti kvázi-szinonimák, vagyis adott szakterület azonosnak tekinthető kifejezéseinek nézőpontbeli különbségeinél merülnek fel. Ebben az értelemben, a deskriptorok X és Y nyelven ugyanazt az általános fogalmat fejezik ki, de a kifejezések jelentése X és Y nyelven nem teljesen azonosak. Ezek a differenciák gyakran inkább kulturális jellegűek, mint szemantikaiak, azaz jelentésárnyalatbeli vagy felfogásbeli különbségről beszélhetünk. Így például: A „Strong-willed” és a „pig-headed” angol kifejezések jelentése egyaránt „stubborn” vagyis magyarul makacs, de míg az első kifejezés pozitív jelentéssel bír, addig a második negatívval. Az IFLA irányelvei alapján a pontatlan egyenértékűség esetében a kifejezéseket úgy kell kezelni, mintha pontos egyenértékűség állna fenn a kifejezések között (5. 9. táblázat).

angol	spanyol	francia
Historic settlements	= Asentamientos historicos	≈ Site de peuplement

5. 9. táblázat: *Pontatlan egyenértékűség (HEREIN Tezaurusz, 2003)*

Az egyenértékűség következő megkülönböztetési fokán a részleges ekvivalenciák helyezkednek el, amelyek a nyelvek közötti kvázi-szinonimáknál a specifikusságbeli differencia esetében állnak fenn. Ebben az esetben a deskriptor a nyelvek egyikén egy kissé tágabb vagy szűkebb értelmű, mint ugyanaz a deskriptor egy másik nyelven (IFLA, 2009).

A részleges ekvivalencia esetében az IFLA irányelv (2009) három megoldást kínál, az egyik a már fentebb is használt megoldás, vagyis az, hogy kezeljük úgy a kifejezéseket, mintha egymás egyenértékűi lennének. A második az, hogy minden egyes nyelvi változatot tüntessünk fel a tezauruszban és rendezzük ezeket hierarchikusan, vagyis a tágabb értelmű kifejezést fölérendeltként, míg a szűkebb értelmű kifejezést alárendeltként adjuk meg (5. 10. táblázat). Illetve kezelhetjük úgy az esetet, mintha egy a többhöz egyenértékűség lenne.

német	angol
Wissenschaft	Wissenschaft SN Loan term adopted from German NT Science
Science D Lehnwort aus dem Englischen OB Wissenschaft	Science BT Wissenschaft

5. 10. táblázat: *Részleges ekvivalencia (ISO 5964,1985)*

Hudon szerint (1997) az egy a többhöz ekvivalencia, röviden abban az esetben áll fenn, ha túl sok vagy túl kevés egyenértékű kifejezés áll a rendelkezésünkre. Ennek az ekvivalenciának tehát két altípusa van, az egyik, amikor forrásnyelvi kifejezésnek több, mint egy egyenértékű kifejezése van a célnyelven. A másik pedig ennek az ellentéte, vagyis ha nincs elég célnyelvi kifejezés. Túl sok célnyelvi kifejezés esetén három lehetséges megoldás létezik. Ha létre hozunk egy előre összevont célnyelvi deszkriptort (5. 11. táblázat). Vagy, ha módosíthatjuk, vagy specifikálhatjuk a forrásnyelvi kifejezést például egy minősítő hozzáadásával (5. 12. Táblázat). Illetve, ha létrehozhatunk egy vagy több célnyelvi nemdeszkriptort a deszkriptóhoz kapcsolódóan (5. 13. táblázat). A harmadik megoldás a leginkább elfogadható, a nyelvi ekvivalencia szempontjából, mivel így nem keletkeznek mesterségesen létrehozott kifejezések.

angol	francia
Listed building	Édifice inscrit + Édifice classé EP Édifice inscrit EP Édifice classé

5.11. táblázat: *Egy a többhöz ekvivalencia, összevont célnyelvi deszkriptor létrehozása. (HEREIN tezaurusz)*

angol	német
Festival (feast) RT Holiday	Fest RT Feiertag
Festival (series of performances) RT Event	Festival RT Event

5.12. táblázat: *Egy a többhöz ekvivalencia, specifikáció minősítővel (Schott, 2006)*

angol	német
Festival	Festival BF Fest VB Holiday
	Fest BS Festival

5.13. táblázat: Egy a többhöz ekvivalencia, célnyelvi nemdeszkriptorral (Schott, 2006)

Az egy a többhöz ekvivalencia másik altípusa, abban az esetben áll fenn, ha a célnyelvi kifejezés csak forrás kifejezések kombinációja révén képviselheti a fogalmat, vagyis nincs megfelelő célnyelvi kifejezés. Erre a problémára négy megoldás létezik. Az első, ha létrehozunk egy összevont célnyelvi kifejezést (5. 14. táblázat). Vagy létrehozhatunk egy általunk megalkotott kifejezést ún. neologizmust (5. 15. táblázat). Használhatunk fölérendelt célnyelvi kifejezést minősítővel együtt (5. 16. táblázat). Vagy létrehozhatunk egy vagy több forrásnyelvi nemdeszkriptort, amely(ek) kapcsolatban állnak a deszkriptorral (5. 17. táblázat).

német	angol
Berufspraktikum	Vocational education + Internship

5.14. táblázat: Egy a többhöz ekvivalencia, összevont célnyelvi kifejezésekkel (Schott, 2006)

német	angol
Berufspraktikum	Vocational training SN Use this descriptor for a combination of vocational education and internship

5.15. táblázat: Egy a többhöz ekvivalencia, neologizmus (Schott, 2006)

német	angol
Gymnasium	Secondary school (gymnasium)

5.16 táblázat: Egy a többhöz ekvivalencia, fölérendelt kifejezés minősítővel. (Schott, 2006)

angol	német
Town redevelopment	BS Stadt und Sanierung
Town	Stadt
Redevelopment	Sanierung BF Stadtsanierung

5.17. táblázat: Egy a többhöz ekvivalencia, forrásnyelvi nemdeszkriptorokkal (Shcott, 2006)

A nyelvek közötti ekvivalencia utolsó típusa, abban a helyzetben áll fenn, ha nem létezik a forrásnyelvével egyenértékű kifejezés, amely azonos jelentéssel bírna a célnyelven. Ez a helyzet elfogadhatatlan a szimmetrikus tezauszban. *Hudon* szerint (1997) ebben az esetben a szimmetrikus struktúra érdekében, a legegyszerűbb megoldás az, ha eltávolítjuk a „gazdátlan” forrásnyelvi kifejezést, különösen akkor, ha ez egy nagyon speciális kifejezés. Azonban, ha az eltávolítás nem célszerű, további három lehetőség áll rendelkezésünkre. Átalakíthatjuk a „gazdátlan” kifejezést nemdeszkriptorrá és ezt összekapcsolhatjuk egy olyan deszkriptorral amellyel számos tulajdonságban osztozik (5.18. táblázat). Ez a megoldás nincs különösebb hatással a szerkezetre, így ez lehet az ideális választás, feltéve, ha a „gazdátlan” kifejezés valójában közeli rokona az adott deszkriptornak a többnyelvű tezauszban. Vagy importálhatjuk a forrás kifejezést a célnyelvre (5.19. táblázat). Azonban a felhasználók egy adott kultúrából talán így olyan információhoz juthatnak a szervezetekről, a folyamatokról és a kapcsolatokról, amelyek csak egy másik társadalomban léteznek. Vagy létrehozhatunk egy egyenértékű új szót vagy kifejezést (5.20. táblázat). A neologizmusok létrehozása azonban soha nem a legjobb megoldás. A tezausz nem egy terminológiai kifejezésbank, a tezausz önmagában nem idéz elő lényeges változásokat a nyelvben, hanem inkább tükröznie kell a nyelvek specializált használatát a társadalom bizonyos szegmenseiben.

angol	francia
Classroom environment USE Learning environment	
Learning environment UF Classroom environment UF Educational environment	Milieu éducatif EP Conditions d'apprentissage EP Milieu pédagogique EP Milieu d'apprentissage

5.18. táblázat: *ha nincs ekvivalencia, átalakítjuk a kifejezést nemdeszkriptorrá (Canadian literacy thesaurus, 1996)*

német	angol
Diakonisches Werk	Diakonisches Werk SN Form of social work done by protestant churches in Germany

5.19. táblázat: *ha nincs ekvivalencia, importálhatjuk a forrás kifejezést a célnyelvre. (Schott, 2006)*

német	angol
Hochrechnung	Projection (statistical)

5.20. táblázat: *ha nincs ekvivalencia, neologizmust alkalmazhatunk (Schott, 2006)*

5.3 Már kialakított teauruszok összeolvasztása, összekapcsolása

Arra is lehetőség van, hogy ötvözzünk két vagy több már kialakított teauruszt egyetlen új többnyelvű változatba. Illetve már kialakított teauruszok és tárgyszójegyzékek összekapcsolásával is létrehozhatunk többnyelvű struktúrát. A cél az, hogy felhasználjuk a meglévő szótárak szókincsét amennyire csak lehetséges. A gyakorlatban ez a módszer egy már létező szótárakon alapuló új teauruszt eredményez. Az így kapott teauruszt a felhasználók alkalmazhatják mind az indexelésnél és a keresésnél. Az ötlet mögött az az elképzelés áll, hogy a használók továbbra is azzal az ismert szókinccsel dolgozhatnak, amivel már korábban is találkoztak, így egy ismerős szókészlethez egy új funkciót rendeltünk (IFLA, 2009).

Az összekapcsolás több különböző formában lehetséges. Az első esetben két vagy több megegyező egynyelvű szótárt kapcsolunk össze, erre példát találunk az orosz nyelvterületen. Vagy különböző nyelvű, egynyelvű szótárak összekapcsolását végezzük el, például német és francia szótárak összekapcsolásának esetében. Illetve többnyelvű szótárak összevonásával és a fent említett lehetőségek kombinációjával is megvalósítható egyetlen új többnyelvű szótár létrehozása. Az összekapcsolás általában a teauruszok kifejezéseinek és a tárgyszójegyzékek szavainak feltérképezésével kezdődik, innen a 'mapping' kifejezés a szabványban. A találatokat sokféleképpen jeleníthetik meg, például egy táblázatban, amely annyi oszlopot tartalmaz, mint a felhasznált szótárak száma. A táblázat minden egyes sorában pedig a különböző szótárak adott egyenértékű kifejezései helyezkednek el. A táblázat egy cellája nulla, vagy annál több kifejezést is tartalmazhat (IFLA, 2009).

Hudon (1997) az alábbiakat írja az összekapcsolásról. Bár már jóval elfogadhatóbb ez a módszer, mint a fordítás, de nagyon komoly gyakorlati problémákkal jár együtt. Ezt az eljárást a legnehezebb intellektuálisan kezelni, mivel minden egyes teaurusz eltér egymástól egy bizonyos mértékben, lefedettség mélységben, a prekordináció fokában, a specifikusság szintjén stb. Emellett valószínű, hogy a legnagyobb és legfejlettebb teaurusz nyelve és struktúrája lesz a domináns és az is, hogy a többi teaurusz struktúráját fogják igazítani/egyeztetni ehhez az egyetlen dominánshoz.

5.3.1 Az ekvivalenciák:

Az összekapcsolással kialakított többnyelvű teauruszokban három ekvivalencia típust különböztetünk meg egymástól. A teljes és a hiányos ekvivalenciát, valamint azt a már említett esetet, amikor nincs ekvivalense a keresett kifejezésnek. Ebben az esetben tájékoztatni kell róla

a felhasználót, hogy az általa keresett kifejezésnek nincs ekvivalense a másik nyelv(ek)en. A teljes ekvivalenciánál, minden szókészletben adott kifejezésre az ekvivalensek keresése során egy és csak egy deskriptor létezik (5.21. táblázat). Ebben az esetben az összekapcsolás lehet hozzátétőleges. Az egyenértékűség nem azt jelenti, hogy itt a kifejezések azonos nyelvi jelentésűek, hanem azt, hogy várhatóan az eltérő szókészletekből származó kifejezések megegyező témájú forrásokhoz vezetnek.

angol	francia	német
Decathlete	Décathloniens	Zehnkämpfer
Decathlon	Décathlon	Zehnkampf
Discus throwing	Lancer du disque	Diskuswurf
Divers	Plongeurs	Kunstspringer
Diving	Plongeon	Wasserspringen
Hammer throwing	Lancer du marteau	Hammerwurf
Hurdle racing	Course de haies	Hürdenlauf

5.21. táblázat: teljes ekvivalencia, az összekapcsolás eljárásánál (MACS, 2004)

Hiányos ekvivalencia esetében, legalább az egyik szókészletben két vagy több kifejezést kell kombinálnunk, amelyek egymás egyenértékű kifejezéseiként használhatóak. Ez a kombináció alapulhat a Boole algebra „ÉS” illetve „VAGY” operátorain. A „VAGY” összekapcsolás akkor szükséges, ha a deskriptor az adott szótárban két kifejezést jelöl, az első kifejezés együttes kiterjesztését használni a másik szókészletben (5. 22. táblázat). Az ÉS operátorra akkor van szükség, amikor az egyik szókészlet kifejezése hierarchikusan alacsonyabb szinten létezik, mint a releváns kifejezés legalább egy a készítés során felhasznált másik szókészletben (5. 23. táblázat).

angol	német	francia
Jumping	Hochsprung OR Sprung	Saut en hauteur OR Sauts (athlétisme)
Long jump	Weitsprung	Saut en longueur

5.22. táblázat: hiányos ekvivalencia, VAGY operátorral (MACS,2004)

angol	német	francia
Cycling	Radsport	Cyclisme
Cycling–Training	Radsport AND Training	Cyclisme AND Entraînement

5.23. táblázat: hiányos ekvivalencia, ÉS operátorral (MACS, 2004)

5.4 A tezaurusok fordítása

A fordítás eljárása igen vitatott, éppen ezért ezt a megközelítést az *IFLA* irányelve (2009) már nem is tartalmazza, noha a meglévő tezaurusz egy vagy több új nyelvre fordítása igen népszerű volt a múltban, főleg gazdasági okokból. Ez a módszer nem teszi lehetővé az összes érintett nyelv közötti egyenlő kezelés lehetőségét, mivel itt a forrás nyelv lesz a domináns nyelv és az így előállított termék nem tükrözi megfelelően a célkultúrákat, vagyis az egynyelvű tezaurusz kulturálisan mindig elfogult. Az egyenes fordítás egyfajta kulturális imperializmushoz vezethet (*Hudon, 1997*).

A szakirodalomban vannak, akik azt állítják, hogy lehetetlen fordítás útján tezauruszt készíteni, azonban a gyakorlatban a két- és többnyelvű tezauruszok többnyire úgy keletkeznek, hogy egy forrástezauruszt a célnyelvre fordítanak. Fordítás esetében is ugyanolyan a szakembergárdára van szükség, mint a párhuzamos eljárásnál, legfeljebb térben és időben vannak egymástól elválasztva a specialisták. Másrészt a fordítási eljárás számos nehézség forrása lehet, mivel alapvetően sokkal nagyobb kifejezés halmaz keletkezik, illetve ezzel együtt rendkívül megnő a homonimák és a pontatlan szakkifejezések száma. Ebben az esetben sokkal több egyeztetésre van szükség az egyes munkafolyamatok között, megnő az ellenőrzés szerepe, ami még több munkát generál. Az eljárás másik problematikus területe lehet, hogy kimarad sok, csak a célnyelven létező kifejezés, viszont rendkívül sok összetett szó keletkezhet a forrásnyelven (*Ungváry, 1981*).

Hudon (1997) a következő három problematikus elemre hívja fel a figyelmünket, amelyek hagyományosan kapcsolódnak a többnyelvű tezauruszok fordításához. Az első problémát korábban már említettem, ami a következő, a kifejezések szó szerinti fordítása a forrás nyelvről semmitmondó kifejezéseket eredményezhet a célnyelven. A második, egy adott nyelv tágítása, annak érdekében, hogy illeszkedjen egy idegen fogalmi struktúra azonos pontjaihoz, azt eredményezheti, hogy a nyelv alig felismerhetővé válik a beszélői számára. Az utolsó intés a fordítással kapcsolatban az, hogy mielőtt áthelyeznénk egy teljes fogalmi struktúrát egyik kultúrából a másikba jól meg kell gondolnunk ez vajon egyáltalán célszerű lesz-e.

Doerr (2011) szerint, még ha a terminológiai források szemantikai (jelentéstani) különbözőségére gyakran is hivatkozunk úgy, mint egy problémára, ennek a szellemi alapnak és a szerkezetnek a szisztematikus/rendszeres elemzése ezt nem képes kiküszöbölni. Véleménye szerint a fordított tezaurusz is tezaurusz, ahol minden fogalom optimálisan

értelmezett több nyelven, ami lehetővé teszi a nyelvet beszélőknek, hogy a tezaurust jobban megérthessék és eredményesebben használhassák. Azt is hangsúlyozza, hogy az ilyen fordítások általában nem megalapozottak a célnyelv indexelési kifejezéseinél.

Bassnet (1995) arra, hívja fel a tezausz kialakításának ezen formáját kivitelezők figyelmét, hogy a fordítás művészete már *Horatiusnál* és *Cicerónál* is a forrásszöveg gondos értelmezését és a szöveg nem szó szerinti átvitelét jelentette. Vagyis szerinte, a fordító még mindig felelős a célközönségnek.

Bár most úgy tűnhet, hogy általános megegyezés van abban, hogy nincs értelme a tükörfordításnak, de a fordítást nem tekinthetjük egy egyszerű aktusnak. Az elméletek a fordítót gyakran a saját kultúrája „fogjának” tartják (*Nida*, 1981). A fordító általában a saját nyelvi-kulturális kontextusában érti meg az üzenetet, majd megpróbálja megoldani a problémát idegen eredetű kifejezések használatával. Függetlenül attól, hogy egy forrásnyelvi fogalom szemantikailag nem mindig egyenértékű a célnyelv megegyező fogalmával. Általában egy másik kultúrában a magyarázó kifejezések ezért sokkal jobbak, egyenértékűbbek, mint a lefordított idegen nyelvű kifejezések. Ezt az elvet tükrözi a tezausz készítése szabványokban és irányelvekben is, mivel az ebből a hibából keletkező jövevényszavak, kölcsönzött kifejezések, neologizmusok használata ezekben sem ajánlott.

A korai elvárások alapján készülő fordítások még nem feleltek meg a nyelv természetének. Mára a fő trend megfigyelhető a modern fordítási elméletekben, a terminológiai útmutatókban, valamint a tezausz szabványokban is. Az egyenértékűség típusainak felosztása a szabványokban, útmutatókban és a fordítási elméletekben ebből a felismerésből fakad. A funkcionális, kommunikatív fordítási elméletek közé tartozik többek között a dinamikus ekvivalencia elmélete, a funkcionális ekvivalencia elmélete és a SKOPOS elmélet (*Keränen*, 2005).

A dinamikus ekvivalencia elméletét *Nida* fejlesztette ki a 1969-ben, ez volt az első kommunikatív fordítási elmélet. A dinamikus szó arra az elképzelésre utal, hogy a fordításnak nem kell ugyanazt a hatást okoznia a célközönségnél, mint a forrásszövegnek az eredeti kontextusban. *Nida* fordítási meghatározása szerint a fordítás újra alkotja a célnyelvben a forrásnyelv legközelebbi természetes ekvivalensét, először a kifejezések értelmében, másodszor a kifejezések stílusában. A rendeltetési vagy funkcionális ekvivalencia teóriájának az az elképzelés az alapja, hogy a forrásszöveg rendeltetése megegyezik, vagy hasonló a célszöveg rendeltetésével. Más szavakkal a fordítás rendeltetése hozzá igazodik a forrásszöveg rendeltetéséhez (*Vehmas-Lehto* 1999).

A SKOPOS elmélet a fordítási műveletet hangsúlyozza, tehát a fordítás cselekvési folyamatát. Vermeer (1989) szerint a fordítási művelet, olyan, mint egy bizonyos fajta viselkedés, ezt a viselkedést a cselekmény szempontjából cselekedetnek nevezhetjük, mivel az azt végző személynek kell (potenciálisan) képesnek lennie arra, hogy megmagyarázza miért járt el úgy ahogy, bár más módon is eljárhatott volna. Továbbá a cselekedet valódi okai mindig kifejezhetőek a célok szándékainak vagy állításainak segítségével a kifejezésekben, mint egy jó okkal elkövetett cselekvés (Reiss, 1986).

Vehmas-Lehto (1999, 12. o) gondos meghatározást adott a fordítás meghatározására: *fordításkor valaki kifejez valamit a célnyelv eszközeivel, amit korábban már kifejeztek a forrásnyelv eszközeivel.* A meghatározás jelzi, hogy a leginkább a fordítás tartalma a fontos, a jelentés és nem maga a nyelv.

5.4.1 A fordítást kiegészítő statisztikai eljárások

Keränen (2005) a fordítási eljárást statisztikai módszerekkel is támogatná, azért, hogy a megfelelő szókészlet feltérképezésén túl, az adott nyelv kifejezései közötti kapcsolatok is átláthatóbbá váljanak a fordító számára. Az általa javasolt eljárások a társ-szó analízis, a szó társítási módszer és a szemantikus alkotórész analízis. A társ-szó elemzés nagyon hasonlít az együtt idézés elemzésére, a kifejezések együttes előfordulásával foglalkozik a dokumentumokban. A társ-szó analízist úgy lehet jellemezni, mint egy kétdimenziós térképet. A térképen a relatív kör mérete összefügg az előfordulások számával, ez alatt azt érthetjük, hogy minél nagyobb a körön belüli deskriptorok száma, értelem szerűen annál nagyobb lesz a körünk, viszont ha kisebb a kör, akkor több a kapcsolat lesz az egymással relációban álló deskriptorok között. A társ-szó analízis eredménye két dimenziós térképen is bemutatható.

A szó társítási módszert már korábban is alkalmazták a tezausz készítés keretein belül. Nielsen (2002) szerint a szó társítási módszer egy használható és működőképes tezauszt eredményezhet, amely jobban teljesít, mint a hagyományos tezausz készítési módszerekkel kialakított társai. A szó társítási módszer előnye az, hogy gazdagíthatja a tezausz építést az aktuális, releváns és terület-specifikus információk biztosításával.

A szemantikai komponens analízist gyakran említik jelentés analízisként is. Az eljárás a szó jelentésváltozatait osztja kisebb elemekre, szemantikai jellemzőkre. Az összetevők elemzése különösen hasznos a poliszémiák esetében (Keränen, 2005).

6. A TÖBBNYELVŰ TEZAUROSZOK MAI MEGJELENÉSI FORMÁI ÉS A JÖVŐBELI ELKÉPZELÉSEK

6.1 A többnyelvű tezauruszok formai elrendezése

A többnyelvű tezauruszok szerkezete a korábban részletezett kialakítási eljárások bármelyikének alkalmazása után a következő formákban jelenhet meg. Akárcsak az egynyelvű változatok, két nagy részegységből a bevezető és szótári részből állnak. A bevezető rész tartalmazza a címlapot és a bevezetést, utóbbiban a tezaurusz céljáról, szerkezeti felépítéséről, készítésének módjáról, mennyiségi jellemzőiről, a benne használt rövidítések (jelölők) szerepéről és jelentéséről stb. találunk információt, ezt a tezaurusz minden nyelvén elérhetővé kell tenni. A szótári rész a lexikai egységek különböző ismérvek szerint rendezett jegyzékeinek együttese, ennek a fő résznek két alapvető formája jellemző.

A megjelenítés egyik verziója a többhasábos, ahol hasábonként más-más nyelven szerepelnek a teljes tezauruszcikkek egymás mellett. Vagyis a nyelvi verziók párhuzamos bemutatása egy oldalon (6.1. táblázat). Ez csak akkor fogadható el, ha ez a tezaurusz minden egyes nyelvén megjelenik, miközben a bal oszlop az utasításokat tartalmazza. Illetve ehhez kapcsolódóan létezik a többhasábos hierarchikus megjelenítés is, amikor két párhuzamos hasábbal a csúcshierarchia teljes hierarchikus lebontása látható két nyelven (Springl, 1980).

FAMILLE EQ Families TS Famille a faible revenu Famille d'accueil Famille monoparentale TA Droit de la famille Finances familiales Violence familiale	FAMILIES EQ Famille NT Low income famille Foster families Single parent families RT Family law Family finances Domestic violence
FAMILLE A FAIBLE REVENU EQ Famille a faible revenu	LOW INCOME FAMILIES EQ Low income families
FAMILLE D'ACCUEIL EQ Foster families TA Adoption	FOSTER FAMILIES EQ Famille d'accueil RT Adoption

6.1. táblázat: A tezaurusz adatainak megjelenítése – többhasábos formában (Hudon, 1997)

Az egyhasábos megjelenítés egyik formája az, amikor az összekapcsolódó különböző nyelvi kifejezéseket egy hasámban tüntetik fel, ez jellemző az *UNESCO* teauruszára és az *APT*-re is (6. 2. táblázat). Az egyhasábos megjelenítés másik változata az, amikor valamennyi nyelvi változatot fizikailag szétválasztanak a különálló szekciók minőségében (6.3. táblázat), ebben az esetben minden nyelvi változatnak van saját bevezetése, indexe stb. A megjelenítés ezen típusát részesítetik előnyben Kanadában, valamint ez az alapmodell jelenik meg a leggyakrabban elektronikus változatokban is (Hudon, 1997).

arab kifejezés (arab nyelven)
osztály szintje és kódja (angol)
angol ekvivalens (angol)
(USE) (UF) utalók (arab)
(SN) megjegyzés (arab)
(BT) tágabb kifejezés (arab)
(NT) alárendelt kifejezés (arab)
(RT) kapcsolódó kifejezés (arab)

6.2. táblázat: *Az APT kétnyelvű terminológiájának egyhasábos megjelenítése (s.k.)*

FINANCEMENT	
EQ	Funding
EP	Autofinancement
	Programme de financement
TA	Aide financiere
	Levee de fonds
FINANCEMENT DE PROGRAMME	
EQ	Program funding
FINANCES	
EQ	Finances
TA	Budget
	Comptabilite
FINANCES	
EQ	Finances
RT	Accounting
	Budget
FINANCIAL NEEDS	
EQ	Besoins financiers
BT	Needs

6.3. táblázat: *A teaurusz adatainak megjelenítése – a verziók szétválasztása (Hudon, 1997)*

6.1.1 Az UNESCO többnyelvű tezaurusának megjelenítési lehetőségei

Az UNESCO tezaurusz, az *Egyesült Nemzetek Oktatási, Tudományos és Kulturális Szervezete* által fejlesztett ellenőrzött szótár, amely magába foglalja az alábbi területek kifejezésit: oktatás, tudomány, kultúra, természettudományok, társadalom- és bölcsészettudományok, tájékoztatás és kommunikáció, politika, jog és közgazdaságtudomány, továbbá az országok neveit és az egyes politikai, gazdasági, földrajzi, etnikai, vallási és nyelvi csoportosulásokat. A fent említett területek 7 nagy témakörbe vannak csoportosítva, számozásuk 1 és 7 közötti egész szám. Ezeken belül a speciálisabb témakörök mikrotezauruszokban szerepelnek (számozásuk a följük tartozó témakör egész számának és a mikrotezauruszok közötti megkülönböztetést szolgáló további két számjegyből áll). A tezauruszt az UNESCO azzal a céllal fejlesztette ki, hogy az UNESCO *Intergált Dokumentációs Hálózatában* indexelésre és az információ visszakeresésére használhassák. Mára az UNESCO tezaurusz megkönnyíti mind a könyvtárak, a levéltárak és más hasonló intézmények indexelését. Jelenlegi feladata a szervezet céljainak megfelelően feltárni a fejlődést szolgáló dokumentumokat és a mintaalkalmazásokat (Severino, 2007).

A terminológiai szótárát 1977-ben publikálták először, a második kiadás 1995-ben került közzétételre. Az UNESCO tezaurusz 7000 angol és orosz, illetve 8000 francia és spanyol kifejezést tartalmaz, ezen kívül pedig a magas hierarchia szintek arab és kínai kifejezései is szerepelnek benne. Ezek az adatok arra engednek következtetni, hogy a tezaurusz kidolgozását, a francia és a spanyol rész megalkotásával kezdték és nyelvpárokban dolgoztak/dolgoznak a kialakítók. A tezaurusz főrészből, permutált betűrendes részből és hierarchikus részből áll. Mint más többnyelvű tezaurusznál itt is a lexikai egységeket három relációtípus kapcsolja össze, a hierarchikus, az asszociatív és az ekvivalencia reláció, ezen kívül pedig a megjegyzések (SN) segítik a kifejezések jelentésének megértését.

Az UNESCO tezauruszában az alfabetikus és a hierararchikus megjelenítés között választhatunk. Mindkét esetben egyhasábos megjelenítést kapunk, a különbség a kettő között abban áll, hogy a hierarchikus megjelenítés során először az adott témakör kifejezéseinek fölérendelt kifejezéseit látjuk betűrendben és csak egy lépéssel később jutunk el arra a szintre, amit a másik lehetőség közvetlenül felkínál nekünk. Itt a keresett kifejezést az összes nyelvi egyenértékűvel együtt szerepel, illetve a keresés nyelvének megfelelően a szűkebb és tágabb kifejezések is egyaránt szerepelnek.

6.1.2 Az Eurovoc megjelenítési lehetőségei

Az *EuroVocot* eredetileg kifejezetten az *EU* intézményeinek dokumentációs iratkezelésére tervezték. 1982-ben az intézményeknél használt dokumentációs nyelvek komparatív vizsgálata után az Európai Parlament és az Európai Bizottság egy olyan többnyelvű tezaurusz létrehozása mellett döntött, amely felöleli az európai intézményeket érdeklő területeket, és amely megfelel a vonatkozó nemzetközi szabványoknak. A munkát az *OECD* makrotezauruszának felhasználásával elkészített egynyelvű tezaurusz kialakítása alapozta meg, továbbá figyelembe vették a Bizottság Központi Könyvtárában használt, 1978-as *ECLAS* tezaurusz legelső változatát és a Központi Automatizált Dokumentációs Szolgálat (SCAD) által javasolt deszkriptorjelölteket is (Koreny, 2005).

Az *EuroVoc* tezaurusz első kiadása 1984-ben került a nyilvánosság elé hét nyelven (angol, dán, francia, görög, holland, német és olasz), két kötetben: egy betűrendes és egy tematikus tezaurusz formájában. A nyelvi és terminológiai bővítések után a legújabb változat az *EuroVoc* 4.3. kiadása 2008-ban, a 72. témakör (Földrajz) és a 76. témakör (Nemzetközi szervezetek) alapos felülvizsgálatát követően jelent meg az interneten. Az új bulgáriai és romániai régiókkal együtt 6797 fogalmat tartalmaz. Az *EuroVoc* mára az Európai Unió 23 hivatalos nyelvén, valamint szerbül is elérhető. A tezaurusz 21 témakört fed le, amely kétszintű hierarchia szerint tagozódik: az első szinten a témakörök kapnak helyet, melyeket egy kétjegyű szám és egy cím kísér pl. 10 EURÓPAI UNIÓ. A második szint a mikrotezaurusz szintje, amelyet egy négyjegyű szám és egy cím kísér pl. 1011 az Európai Unió joga. A témakörök és a mikrotezauruszok számozása valamennyi nyelvi változatban egységes (Severino, 2007).

Hátrányai közé tartozik, hogy mivel a tezauruszt az EU tevékenységi területeihez igazították, a speciális dokumentumok indexelésére és keresésére nem alkalmas. Illetve az egyes nyelvek, mint pl. az angol, francia, német sokkal kimunkáltabb, mint pl. a bolgár vagy a román. Mindez abból fakad, hogy a tezaurusz kialakítását annak idején a „Hatok” számára kezdték el parallel formában kialakítani, majd ehhez a struktúrához kapcsolták a később tagországgá vált államok hivatalos nyelveit.

Az *EuroVoc* elektronikus változatának főoldalán a megjelenítés több lehetősége közül választhatunk. A böngészés kiválasztása után az egyes témaköröket kapjuk meg, amelyek a jobb oldalon beállított nyelv függvényében a kívánt változatokban jelennek meg. A 21 témakörön belül, több speciális területet különböztetnek meg, azonban, ha ezek közül kiválasztunk egy alkategóriát, az adott területhez kapcsolódó összes fogalmat megtekinthetjük,

de csak egy bizonyos nyelven. Tehát az EuroVocban az egyhasábos nyelvenkénti megjelenítés dominál. Ha kiválasztunk egy konkrét kifejezést, akkor a kéthasábos megjelenítéssel találkozhatunk, de nem a klasszikus formában, hanem a bal oldalon kap helyet az általunk kiválasztott kifejezés és a hozzá kapcsolódó fölérendelt kifejezés. A jobb oldalon pedig adott kifejezés összes nyelvi egyenértékűjét találjuk. Ha a keresés opcióját választjuk, két hasábos formában jelennek meg kifejezések, az egyik oldalon a keresési találatok, míg a másik sorban a hozzájuk kapcsolódó fölérendelt fogalmak.

6.2 A többnyelvű teauruszok problematikus területei

A 21. században a többnyelvű eszközök egyre nagyobb teret nyernek és egyre több és színesebb felhasználói csoporthoz jutnak el, akik eltérő kulturális és nyelvi háttérrel rendelkeznek és hasonlóan változatos információ típusokhoz szeretnének hozzájutni. *Jorna* (2001, 284. o) úgy véli, hogy a többnyelvű teauruszok leggyakoribb formái jelenleg nem megfelelőek erre a szerepre. Véleménye szerint a többnyelvű eszközök ezen típusának új formájára van szükség. Ennek érdekében foglalkozik az *InfoDEFT* nevű kísérleti teaurusz kialakításával, amit az online teauruszok lehetséges új modelljének tart. Az *InfoDEFT* egy szemantikailag strukturált, enciklopédikus jellegű és többnyelvű teaurusz, amely felhasználható a mesterséges tanulást segítő programok esetében is, ezzel csökkentve az eszköz manuális előállításának költséges mivoltát.

A többnyelvű információkeresés biztosítására nincs egyszerű és kézenfekvő megoldás. A többnyelvű teauruszok összetett kialakítási problémáira *Hahn* és *Angelova* szerint (1996) ideális opció lehet, egy független nyelvi tudásbázis bevezetése, amely közös fogalmi struktúrát alkotna, minden egyes nyelven. *Jorna* szerint (2001) egy ilyen tudásbázis kiépítése azonban rendkívül költséges lenne, és nagy erejű nemzetközi összefogást kívánná meg. Éppen ezért a problémára inkább, a három nyelvű kísérleti teauruszában az *InfoDEFT*-ben látja a választ, mely rendkívül hasznosan alkalmazható az információ visszakeresésénél és a felhasználók számára is a tények, kifejezések megértéséhez. Erre a modern értelmezésre *Schmitz-Esser EXPO 2000* nevű teaurusza is rávilágított már, de a teauruszban sajnos nem hívják fel a használó figyelmét az egyes nyelvek közti kulturális különbségekre. Az egyes kifejezések szoros ellenőrzésére, különösen az osztályozáshoz tartozó fogalmaké, hívta fel *Jorna* figyelmét arra, milyen fontosak ezek a differenciák. Ezért az *InfoDEFT*-et kialakító csoport facetták alkalmazását találta a legideálisabbnak, vagyis egy olyan szerkezet létrehozását, amely eltér az

egyes nyelvektől, ezzel is felhívva a figyelmet a különbségekre, nem csupán a szókészletben, hanem a fogalmi kontextusban is.

Jorna (2001) három fő sarokpontot határoz meg, amelyek egy jó tezaurusz alapkövei kell, hogy legyenek. Először is a tezaurusznak többnyelvűnek kell lennie, hogy támogassa az információhoz való hozzáférést az egyre inkább globálissá váló információs társadalomban. Másodszor ezeknek a rendszereknek az alfabetikus forma helyett, szemantikailag strukturáltaknak kell lenniük, ez a szempont részben lehetővé teszi a felhasznált nyelvek egyenlő képviselését. Harmadszor, segíteniük kell a felhasználók számára a megértést, a felidézést, és a precíz információ visszakeresést, a definíciók és a megjegyzések használatával.

Míg a modern információs és kommunikációs technológia megkönnyíti a globális adatok transzferét, ugyanez nem könnyíti meg a globális kommunikációt. Jelenleg mind az internet és a nemzetközi vállalkozások nagyrészt az angol nyelvet használják a nemzetközi kommunikációban, amely problémákat vet fel a kultúrák közötti kommunikációban, az angol anyanyelvűek és a más ajkúak körében. A legújabb kutatások szerint, az angol nyelv hamarosan megszűnik uralni a globális kommunikációt, helyét Ázsia és Latin-Amerika egyes nyelvcsoportjai fogják átvenni (*Graddol*, 1999).

A fent említett előrejelzések tükrében, új többnyelvű tezauruszra lesz szükség, hogy támogassa a későbbi uralkodói státuszt nyerő nyelvcsoportok közötti kommunikációt, illetve a nyugat-európai nyelvekhez kapcsolódó diskurzust. Manapság sok úgynevezett többnyelvű tezauruszhoz férhetünk hozzá az interneten, vagy a könyvtárak OPAC-jában, valójában ezek azonban egynyelvű tezauruszok, amelyek két- vagy többnyelvű szótárakkal együtt működnek. A többnyelvű információ visszakereső rendszerek ezen formái gyakran csődöt mondanak, amit igazolni lehet a *LISA* (*Library and Information Science Abstracts*) részletes vizsgálatával. A *LISA* tezaurusza a címkéző kifejezések legnagyobb csoportját angol nyelven tartalmazza, annak ellenére, hogy több, mint 60 ország szakterméséből tartalmaz absztraktokat és teljes szövegű hozzáférést. A *LISA*-nál minden absztraktot angol nyelvre fordítanak, ami súlyosan ellentmondásos keresési eredményekhez vezet, abban az esetben, ha nem adottak a pontos ekvivalensek a nyelvek között. A francia „documentalistes” a német „Dokumentar” és az angol „documentalists” egymás egyenértékű kifejezéseinek vannak feltüntetve a tezauruszban, de különböző keresési eredményekhez vezetnek, a keresések esetében 55, 4, és 110 rekordot kapunk. Ha az „information professionals” kifejezéssel keresünk, akkor 3372 rekordot kapunk, ha a kifejezés nemdeszkriptorával „documentalists” fogalommal ugyanezt végrehajtjuk, csak 590 rekordot kapunk. Az ilyen következetlenségek, legyőzhetőek abban az esetben, ha egyenlő

hozzáférést nyújtunk minden nyelven, ahelyett, hogy az angol nyelvet részesítenénk előnyben. Ez nem csak a kultúrák és így az emberek közötti egyenlő bánásmódot jelenti, hanem lehetővé teszi, a tezaurusz tárgyainak elfogulatlan kezelését is (Hudon, 1997. 84. o.).

Sajnos minden nyelv egyenlő kezelése a rendszeren belül nem könnyű feladat. Valaha még voltak fizikai korlátai a többnyelvű megjelenítésnek, de mára ezeket le lehet küzdeni az elektronika segítségével. A technológia azonban nem tud győzedelmeskedni a szemantikai egyenértékűség intellektuális problémái fölött. A természetes nyelvek komplexitása, amely minden terminológia részei, ritkán engedik pontos ekvivalenciák kialakítását a nyelvek között, amelynek kérdése igen sok gondot okoz a szemantikailag strukturált tezauruszokban. Ez a probléma kevésbé lényeges az alfabetikus tezauruszok esetében, de egy jó tezaurusz nem is alapulhat ilyen elrendezésen (Cabré, 1996).

6.2.1 Az EXPO 2000 és az InfoDEFT újszerű megoldásai

A facetták vagy más szemantikai struktúrák segítik a felhasználót a kifejezések megértésében, és annak az átlátásában, hogyan viszonyulnak egymáshoz a mögöttes fogalmi kapcsolatok. A Schmitz-Esser és csapata által Hannoverben kialakított *EXPO 2000* többnyelvű tezaurusz, általános ismeretkörrel ölel fel, öt nyelven. A *Jorna* nevéhez fűződő *InfoDEFT* pedig egy három nyelvű, csak az információ visszakeresés kifejezéseit feltáró tezaurusz. Mindkettő elektronikus formájú. A két terminológiai szótár azért érdekes számunkra, mert a saját területükön legyőzik a fogalomtárak leggyakoribb nehézségeit.

A legnagyobb különbség ezen példányok és a korábbi tezauruszok között az, hogy túlmutatnak az alapvető szemantikai viszonyokon a tágabb, a szűkebb és a kapcsolódó kifejezések szempontjából. Schmitz-Esser-féle tezaurusz a hierarchikus kapcsolatokban 13 alosztást különböztet meg, míg az *InfoDEFT* facettás alapú. Schmitz-Esser bevezetett egy alapvető szemantikai referencia szerkezetet a bizonytalan asszociatív kapcsolatok helyett, amely hat fő kategóriát tartalmaz, a fogalom, a név, az esemény, a hely, a szempont, és az idő megkülönböztetéseit. Ez a rendszerezés részben megegyezik az *InfoDEFT* öt fő kategóriájával, mely a következő kifejezések alapján válogat, dolgok, emberek, tevékenységek, helyek és környezet. Mindkét tezaurusz tehát azt mutatja be, milyen értelemben kapcsolódnak egymáshoz a kifejezések (Jorna, 2001).

Jorna ugyan egyetért Schmitz-Esser besorolási rendszerével az általános tárgykörű tezauruszban, azzal azonban nem, hogy az *EXPO 2000*-ben azt az elvet követik, hogy a lexikális hézagokat parafrázisokkal kell kitölteni ahol ez szükséges. Mivel szerinte lexikai

hiányosságokra inkább fel kell hívni a figyelmet, mint azokat elrejtteni. A felhasználói tudatosság fokozásával ugyanis, nő a fogalmi megértés és javul a kommunikáció. Az *InfoDEFT* azonban, egy második és lényegesebb tekintetben is különbözik *EXPO 2000*-től. Utóbbiban ugyanis minden deskriptor egy kétdimenziós rácsszerkezetben kerül elhelyezésre, melynek egyik dimenziójában a tizenhárom hierarchia típus, a másikban pedig a hat alapvető szemantikai referencia található. *InfoDEFT*-ben facetták szerepelnek abban az értelemben, hogy a gráfok felépítését követik, vagyis számos horizontális kapcsolatot tartalmaznak, melyek a fák ágainak felelnek meg. A *Schmitz-Esser*-féle táblázatos megjelenítésnek az az előnye, hogy átláthatóbb és következetesebb a felhasználók számára, vagyis inkább enciklopédikus áttekintést nyújt, amely gyakran megy a nyelvi pontosság rovására. Az *InfoDEFT*-re ezzel ellentétben az intuitív megközelítés a jellemző, a cél a kifejezések jelentésének jobb megértése a különböző nyelveken, amely az esetleges enciklopédikus áttekintés rovására megy. Végeredményben megállapítható tehát, hogy a tezauruszok különböző funkciói indokolják ezeket az eltérő modern módszereket (Jorna, 2001, 287. o).

6.2.2 Az interoperabilitás, a jövő tezauruszaiban

Az interoperabilitás a kulcsa az egyidejű keresés és visszakeresés képességének két vagy több adatbázisban egy vagy több nyelven. Az interoperabilitás a különböző rendszerek és szervezetek képessége az együttműködésre. A kifejezést általában technikai rendszerekben mérnöki értelemben használják, vagy alternatív módon tágabb értelemben véve, a társadalmi, politikai és szervezeti tényezők figyelembevételével. Kezdetben az interoperabilitást informatikai rendszerek definiálták, a rendszerek közti korlátozott információcserére. A szemantikus interoperabilitás a rendszer azon képessége, hogy értelmesen és pontosan feldolgozza a kicserélt információt, a rendszer végfelhasználóinak érdekében. Alapfeltétele, hogy mind az átadó és a befogadó oldalon is alkalmaznia kell a referencia modellt, vagyis azt, hogy, amit elküldünk egyezzen meg azzal, amit megértünk belőle (Neelameghan és Lalitha, 2013).

A rendszerek interoperabilitása, egy másik modern megközelítése annak, hogy az eszköz használata egyszerűbbé váljon a felhasználó számára. Régen, miután megnéztünk egy nyomtatott tezauruszt, tovább haladtunk a kiválasztott kifejezések nyomán az indexelő rendszerben, vagy a kereső felületeken. Manapság a kifejezéseket csak át kell másolni az elektronikus tezauruszba, vagy gyakrabban csak rájuk kattintani. A kereső rendszer csak akkor lesz képes kapcsolatba lépni a tezaurusz adatbázisával, ha az interoperabilitás fennáll, Vagyis a

felhasználók böngészhetnek, miközben keresnek, kiválaszthatják a tágabb kifejezéseket, egy általánosabb keresésre és így tovább. Ezek az átvitelek nem bonyolultak, feltéve, ha alkalmazzuk a szabványos adatcsere formátumokat és protokollokat.

A tezauruszok vagy más információkereső rendszerek interoperabilitása, több létfontosságú folyamatot tesz lehetővé. Az egyik az, hogy biztosítja az adatok szabványos formában való behozatalát és használatát más rendszerekben. A másik az, hogy lehetővé teszi a rendszerek között a kifejezések leképezését. Valamint biztosít bármilyen más típusú információcserét a rendszerek között (Aitchison és Clarke, 2004).

Az interoperabilitást elérni, azonban nem olyan könnyű. A szemantikus weben az ontológiák esetében szintén ilyenfajta együttműködésre van szükség. Hamarosan arra is igény lesz, hogy a tezauruszokat összekapcsolják az ontológiákkal és/vagy a taxonómiákkal. Jelenleg ennek megoldásával foglalkozik az ISO 25964 szabvány második „*Interoperabilitás más szótárakkal*” részén dolgozó munkacsoportja is. Így ez lehet a többnyelvű tezauruszok jövője.

7. ÖSSZEGZÉS

Egy hatékony és felhasználóbarát többnyelvű tezaurusban az alábbi szempontokat kell érvényre juttatni. A többnyelvű tezaurus tehát egy-egy szakterület több, mint egy természetes nyelven kifejezett, teljes fogalmi gyűjteménye, ahol a fogalmak közötti kapcsolatokat is feltüntetik. Fontos, hogy a tezaurusban a relációk és a fogalmak azonos értékűek legyenek. A többnyelvűség maga biztosítja az információhoz való hozzáférés széles skáláját, mivel teljes fogalmi és terminológiai készletet kínál, minden képviselt nyelven. Emellett fontos, hogy a tezaurus fogalmi értelmezésének középpontjában a válogatás helyett a terminológiai ellenőrzés álljon. A tezaurusnak az alfabetikus szerkezet helyett alapvetően szemantikailag strukturálnak kell lennie, persze, magában foglalhat egy betűrendes részt is, de fontos, hogy ne ez legyen benne a fő rendezési elv.

A deskriptorok optimális számát véleményem szerint nem lehet meghatározni, mivel ez függ a rendszer által felölelt témakörök nagyságától és jellegétől. A nemdeskriptorok nélküli tezaurus, azonban semmilyen körülmények között nem ajánlott, mivel ez alapvetően megnehezíti a feldolgozó és/vagy a kereső dolgát, hiszen tevékenységük végzése közben egy az egyben tisztában kell lenniük a rendszer által használt kifejezésekkel, mely lényegesen nagyobb szakterületi tájékozottságot igényel. A tezaurusok mélystruktúráiban mindenképpen használni kell egy közvetítő rendszert, amelynek véleményem szerint természetes nyelvnek kell lennie, mivel úgy gondolom a tezaurusok kialakítása egy olyan összetett, intuitív folyamat, amely csak emberi közreműködés során válhat minőségivé. A gépi fordítás sokszor erőltetett lehet, azonban az eltérő írásrendszerű nyelvek használata során, gépi segítséget is alkalmaznunk kell, fordítókódok formájában, melyek az ellenőrzés szempontjából válnak fontossá. Ebben az esetben nem a hierarchikusan képzett, hanem az ugrószámos változatokat kell preferálnunk.

A többnyelvű tezaurusok kialakítási eljárásai közül az *ab initio* nyelvpárokkal dolgozó eljárását preferálják a szakirodalomban és a szabványok is alapvetően ehhez a folyamathoz nyújtanak iránymutatást. Ez az eljárás a szimmetrikus tezaurus struktúrára törekszik és takarékos szógyűjtés jellemzi. Az alulról-felfelé építkezésnél sorolják az ekvivalenciákat a legtöbb kategóriába, vagyis a legjobban kidolgozott az egyenértékűségek megkülönböztetése is. Éppen azért egyértelmű számomra, hogy az így kapott tezaurusok lesznek a legideálisabbak minden szempontból. Az eljárás legnagyobb kritikája az, hogy komoly szakembergárdát és költségbeli ráfordításokat kíván.

A már kialakított teauruszok és/vagy tárgyszólisták összekapcsolásának előnye az, hogy a felhasználók továbbra is azzal az ismert szókinccsel dolgozhatnak, amivel már korábban is találkoztak, vagyis egy ismerős szókészlethez egy új funkciót rendelnek hozzá. A kialakítás ezen típusa kevesebb ekvivalencia típust használ, mivel a szókészletek „összefésülése” során alapvetően már struktúrákhoz kell alkalmazkodni, így nincs meg az a szabadság, mint az alulról-felfelé történő építkezés során. Hátránya, hogy sokkal több kifejezés gyűlik össze és aszimmetrikus szerkezet eredményez. Az eljárás ugyanúgy, mint az első esetben megfelelő szakembergárdát és költségbeli ráfordításokat követel, azonban nem eredményez akkora pontosságot.

A fordítás mind közül a legvitatottabb eljárás, a szakirodalomban egyes vélemények szerint nem is lehet ilyen formában teauruszt kialakítani. Ezt tükrözik a szabványok is, mivel a két korábban már említett változattal ellentétben, erről nincs szó bennük. A gyakorlat ennek ellent mond, mivel számos esetben így pl. az *EuroVocnál* is kezdetben ugyan követték az alulról-felfelé építkezést, később azonban az egyes tagországok csatlakozásával már fordítást használtak. Úgy gondolom ezt az eljárást is lehet jól kivitelezni, amennyiben emberi erőforrás végzi a munka nagy részét és ha párosítjuk a Keränen által is javasolt módszerekkel (a társ-szó analízissel, a szó-társítási módszerrel, vagy a szemantikus alkotórész analízissel). Ezekre azért van szükség, hogy a megfelelő szókészlet feltérképezésén túl, az adott nyelv kifejezései közötti kapcsolatok is átláthatóbbá váljanak a fordító számára.

A teauruszokat bírálók kritikája általában az, hogy az előállítás költséges és nehéz a bővítés. Azonban egy jól kialakított teauruszban már a kidolgozás előtt biztosítanak arra módszereket, hogy a frissítés folyamatos lehessen. Tény az is, hogy a teaurusz alapú rendszerek rendelkeznek a legjobb pontossági aránnyal, így hiba volna nem használni őket. A véleményem az, hogy a kialakítási eljárást olyan alternatívákkal kell kiegészíteni, amelyek áthidalják a jelenlegi kulturális és nyelvi különbségekből adódó problémákat. Ezek lehetnek statisztikai módszerek, egy egyedi facettás struktúra alkalmazása, illetve más az elektronikus rendszerek átjárhatóságát biztosító lehetőségek. A lényeg az, hogy ne csak a klasszikus eljárások módszereit alkalmazzuk, hanem kombináljuk a lehetőségeinket!

Köszönetnyilvánítás

Ezúton szeretném megköszönni témavezetőmnek Barátné habil. dr. Hajdu Ágnesnek a dolgozat írása során nyújtott hasznos tanácsait és türelmét!

8. BIBLIOGRÁFIA

Aitchison, (1977): *UNESCO Thesaurus: A Structured List of Descriptors for Indexing and Retrieving Literature in the Fields of Education, Science, Social Science, Culture, and Communication*. UNESCO, Paris.

Jean Aitchison, (1986): *A Classification as a Source for a Thesaurus: The Bibliographic Classification of H. E. Bliss as a Source of Thesaurus Terms and Structure*. Journal of Documentation, 42. 11. sz. 160-181.

Aitchison, (1996): *International Thesaurus of Refugee Terminology*. United Nations High Commissioner for Refugees, New York and Geneva.

Aitchison, Gilchrist, Bawden, (2000): *Thesaurus Construction and Use: A Practical Manual*. Aslib, London.

Aitchison és Clarke (2004): *The Thesaurus: A Historical Viewpoint, with a Look to the Future*. Hawoth Press. 5-21.

ASTIA (1953): *Documentation Incorporated, Installation Manual for the Uniterm System of Coordinate Indexing*. ASTIA Document Service Center, Dayton.

Atherton, (1965): *FID/CR Committee on Classification Research: Elsinore. 24–28. Sept. 1964*. FID, Copenhagen, Munksgaard.

Austin (1976) *Guidelines for establishment and development of multilingual thesauri*. Unesco, Paris.

Austin és Dale (1981): *Guidelines for the Establishment and Development of Monolingual Thesauri*. UNESCO, Paris.

Austin, (1984): *PRECIS: A Manual of Concept Analysis and Subject Indexing*. British Library Bibliographic Services Division, London.

Barátné Hajdu és Babiczky, (1998): *Bevezetés az információkereső nyelvek elméletébe és gyakorlatába*. Universitas, Budapest.

Bassnet, (1995): *Teoksesta toiseen. Johdatus kirjallisuuden kääntämiseen*. Jvaskylä, Vastapaino Routledge.

Batten, (1947): *A Punched-Card System of Indexing to Meet Special Requirements*. Proceedings of the 22nd Aslib Conference, Aslib, London. 37-39.

BS 5723 (1987): *British Standard Guide to Establishment and Development of Monolingual Thesauri*. British Standards Institution, London.

BS 6723 (1985): *British Standard Guide to Establishment and Development of Multilingual Thesauri*. British Standards Institution, London.

Cabré, (1996) *A hypermedia project on terminology*. TKE'96, Terminology and Knowledge Engineering. INDEKS-Verlag, 424–430.

Canadian literacy thesaurus (1996): Canadian Literacy Thesaurus Coalition, Toronto.

Chemical Engineering Thesaurus (1961): *A Wordbook for Use with the Concept Coordination System of Information Storage and Retrieval*. American Institute of Chemical Engineers, New York.

Chomsky (1985): *The logical structure of linguistic theory* Chicago. Univ. of Chicago Press, Chicago.

Clarke (2001): "Thesaural Relationships," in *Relationships in the Organization of Knowledge*. Kluwer, Dodrecht 37-52.

Dahberg, (1975) *Concerning thesaurus terminology*. *International Classification*, 2. 2. sz. 111-112.

Diekema, (2012): *Multilinguality in the digital library. A review*. *The Electronic Library*, 30. 2. sz. 165–181.

Doerr, (2001): Semantic Problems of Thesaurus Mapping. *Journal of Digital information*, 1, 8. sz.

Eurovoc thesaurus. A weblap elérhetősége: <http://eurovoc.europa.eu/>
A letöltés ideje: 2015. 01. 06.

Gilchrist (1971) *The Thesaurus in Retrieval*. Aslib, London.

Graddol (1998): *Will English be enough?* Nuffield Foundation, London.

Guidelines for Multilingual Thesauri, (2009). Working Group on Guidelines for Multilingual Thesauri IFLA Classification and Indexing Section. Hagues.

Guidelines for subject authority and reference entries, (1992). IFLA, Frankfurt.

Harris, (1978): *On the translation of classification instruments*. *T. A. Informations*, 57-67.

Hudon (2001): *Relationships in multilingual thesauri*. *Relationships in the organization of knowledge*, Kluwer Academic Publishers, 67-80.

Hudon (1997): *Multilingual thesaurus construction : integrating the views of different cultures in one gateway to knowledge and concepts*. *Information Services & Use*, 17. 2/3. sz. 111-123.

ISO 5954 (1985), *Documentation-Guidelines for the Establishment and Development of Multilingual Thesauri*. International Organization for Standardization.

ISO 2788 (1986), *Documentation-Guidelines for the Establishment and Development of Monolingual Thesauri*. International Organization for Standardization.

ISO 25964 (2011), *The international standard for thesauri and interoperability with other vocabularies*. International Organization for Standardization.

Jorna és Davies, (2001): *Multilingual thesauri for the modern world: No ideal solution?* Journal of Documentation, 57. 2 sz. 284–295

Keranen, (2005): *Multilingual thesaurus construction: Homemakers in Finnish and British discourses*, ISKO, Capítulo Español, Congreso Barcelona 357-379

Keynes, (1985): *BSI ROOT Thesaurus*. British Standards Institution, London.

Klücklich, (1974) *O postroenii tezaurusa dlâ*. Problemy razvitiâ. 1. 73-75.

Kovácsné Koreny, (2005): *Európai uniós információforrások, visszakereshetővé tételük, és a katalógizálás forrásai* :Az Eurovoc -TMT, 52. 9. sz. 43- 47.

Krooks és Lancaster, (1993) *The Evolution of Guidelines for Thesaurus Construction*. Libri 43, 4. sz. 326-342.

Kunz, (2002): *Sachliche Suche in verteilten Ressourcen : ein kurzer Überblick über neuere Entwicklungen*. IFLA. Glasgow.

Landry, (2004): *Multilingual subject access : the linking approach of MACS*. Cataloging & classification quarterly 37. 4. sz. 177-191

Lancaster, (1986): *Vocabulary Control for Information Retrieval*. Information Resources Press.

Luhn, (1957): *A Statistical Approach to Mechanized Encoding and Searching of Literary Information*. IBM Journal of Research and Development 14, 2. sz. 309-317.

Mooers,(1963): *The Indexing Language of an Information Retrieval System*. Information Retrieval Today, 21-36.

Macrothesaurus for Information Processing in the Field of Economic and Social Development (1998) OECD Development Centre, Paris.

Medical Subject Headings (MeSH). National Library of Medicine. Bethesda.

MSZ 3418 (1987): *Magyar nyelvű információkereső teauruszok szerkezete, részei és formája*. Magyar Szabványügyi Hivatal, Budapest.

Neelameghan, (1976): *Absolute syntax and structure of an indexing and switching language*. Ordering system, 165-176.

Neelameghan és Lalitha (2013): *Multilingual Thesaurus and Interoperability*. Journal of Library & Information Technology, 33. 4. sz. 289-294

- Nida és Taber, (1969): *The theory and practice of translation*. E.J. Brill, Leiden.
- Nida és Reyburn, (1981) *Meaning Across Cultures*. Orbis Books, New York.
- NTP MCNTI 3–75, (1976): *Többnyelvű információkereső teaurusz: Tartalom, szerkezet, megjelenési forma*. OMKDK, Budapest.
- NTP MCNTI 15–76. (1978) *A többnyelvű információkereső teauruszok kidolgozásának szabályai: Az NTMK normatív-műszaki előírása*. Budapest: OMKDK, Budapest.
- Nielsen és Lykke (2002): *The word association method: a gateway to work-task based retrieval*. Lbo Akademi University Press, Lbo.
- Reiss és Vermeer (1986): *Mitä kääntäminen on*. Gaudeamus, Helsinki.
- Roberts (1984): *The Pre-history of the Information Retrieval Thesaurus*. Journal of Documentation 40. 4. sz. 271-285.
- Roget, (1852): *Thesaurus of English Words and Phrases*
- Rolling (1976): *Concerning thesaurus terminology. International Classification*, 3. 1.sz 37-39.
- Rosenbaum, (1974): *Methodische Grundsätze für das Erarbeiten mehrsprachiger Thesauren der Mitgliedländer des RWG*. Dokumentation, 24.
- Severino, (2007): *What thesaurus to define EU/ACP relations*. European journal of developmental research, 19. 2 sz. 327-251.
- Shawky, (1991): *Computerized Bilingual Thesauri: Problems of Thesauri Construction and Development in Arabic Languages*. Microcomputers for Information Management 8. 1. 27-44.
- Schmitz-Esser, (1999): *Thesaurus and beyond: an advanced formula for linguistic engineering and information retrieval*. Knowledge Organization, 26. 1. sz. 10–22.
- Soergel, (1974): *Indexing languages and thesauri: Construction and maintenance*. Melville, 1974. p. 293-297.
- Sonnevend Péter, (2013): *Hatvanéves a VINITI. Könyvtári Figyelő* 2013. 59. 2. sz. 315-321.
- Springl, (1980): *Die Erarbeitung eines mehrsprachigen Informationsrecherche-Thesaurus*. Dokumentation/Information, 49. sz. 26-31.
- Arlington, (1960): *Thesaurus of ASTIA Descriptors*. Armed Services Technical Information Agency.
- Thesaurus of Engineering Terms*, (1964) Engineers Joint Council, New York.
- Thesaurus of Engineering and Scientific Terms* (1976) Engineers Joint Council and US Department of Defense, New York.

Schott, (2006): *Thesaurus Sozialwissenschaften = Thesaurus for the social sciences*. Informationszentrum Sozialwissenschaften, Bonn.

Unesco Thesaurus: A weblap elérhetősége: <http://databases.unesco.org/thesaurus/>
A letöltés ideje: 2015. 01. 06.

Ungváry, (1981): *A két- és többnyelvű tezauruszok irodalma*. OSZK KMK. Budapest.

Ungváry, (2001): *Osztályozás és információkeresés: az osztályozás és elmélete*. OSZK, Budapest.

Z39.19 (1993) National Information Standards Organization, Guidelines for the Construction, Format, and Management of Monolingual Thesauri, NISO Press, Bethesda

Vickery, (1960): *Thesaurus: A New Word in Documentation*. Journal of Documentation 16, 4. sz. 181-189.

Vehmas-Lehto, (1999): *Kopiointia vai kommunikointia: Johdatus käännteoriaan*. Oy Finn Lectura Ab, Helsinki.

Vermeer, (1989): *Skopos and commission in translational action*. Readings in Translation Theory, Oy Finn Lectura, Loimaa 173- 187.

Von Hahn és Angelova, (1996): *Combining terminology, lexical semantics and knowledge representation in Machine Aided Translation*. Terminology and Knowledge Engineering. INDEKS-Verlag, Frankfurt. 304–331.

Wolff-Terroine, (1976): *Multilingual systems*. Second European Congress of the European Community, Verlag Dokumentation, 149 – 169.