

A metaforikus nyelvhasználat korpuszalapú elemzése

Babarczy Anna¹, Bencze Ildikó¹, Fekete István¹,
Simon Eszter^{1,2}

¹ BME Kognitív Tudományi Tanszék,
1111 Budapest, Stoczek u. 2.
{babarczy, ibencze, ifekete, esimon}@cogsci.bme.hu

² MTA Nyelvtudományi Intézet,
1068 Budapest, Benczúr u. 33.

Kivonat: Jelen tanulmány a konkrét–absztrakt (vagy metaforikus) nyelvhasználat korpuszalapú elemzésének segítségével arra keresi a választ, hogy a fogalmi metafora hipotézis milyen mértékben járulhat hozzá a metaforikus jelenségek értelmezéséhez természetes nyelvi szövegekben. A kutatás céljaira egy 300 ezer szavas magyar nyelvű korpuszt hoztunk létre különböző szövegtípusokból. [3] és [2] metaforaindexéből 12 ismert fogalmi metaforát választottunk ki, melyek mindegyikéhez két kifejezéslistát állítottunk össze – az egyik a forrástartományra, a másik a céltartományra jellemző kifejezéseket foglalta magában. A metaforák automatikus azonosításához Martin módszerét [17] alkalmaztuk, vagyis olyan mondatokat kerestünk, amelyekben mindkét tartomány kifejezései szerepeltek egyazon mondaton belül. A hipotézis alapján azt feltételeztük, hogy ha egy mondat tartalmaz egy forrás- és egy céltartományi kifejezést is, akkor az metaforikus lesz. Az eredmények azt mutatják, hogy egy forrás–cél tartománypáron belül nem bármilyen asszociáció vezet metaforikus értelmezéshez, és a valóban metaforicitásra utaló relációk mibenléte leginkább az adott szöveg nyelvi tulajdonságain múlik.

1 Absztrakt tudás: a kognitív metaforaelmélet és a statisztikai megközelítés

A metaforamegértés kérdése szorosan kapcsolódik az absztrakt tudás reprezentációjának és elsajátításának problémaköréhez. Az elvont tudás reprezentációjáról két radikálisan különböző, de elméletileg összeegyeztethető megközelítés uralkodik a kognitív tudományban: (i) a testesültség (*embodiment*) hipotézis, amely szerint absztrakt tudásunk fogalmi metaforákra vezethető vissza; és (ii) a statisztikai tanulás hipotézis, amely szerint a nyelv statisztikai tulajdonságai segítségével sajátítjuk el az elvont fogalmakat.

A testesültség nézet a kognitív nyelvészet egyik alaptétele [6, 12, 13, 14, 15]. Kognitív nyelvészeti megközelítésben az absztrakt fogalmak mentális reprezentációja fizikai, elsősorban az ember által közvetlenül tapasztalható jelenségeken alapul. A nyelvi folyamatokra vonatkozóan ez azt jelenti, hogy az elvont fogalmakat (a céltartományt) kódoló nyelv produkciója és feldolgozása konkrét fogalmak (a forrástarto-

mány) metaforikus kiterjesztésén keresztül valósulhat meg, amihez ún. fogalmi metaforák, azaz konkrét—absztrakt fogalmi megfeleltetések állnak a nyelvhasználó rendelkezésére [12, 14]. Az *időt pazarol* vagy *védi az álláspontját* kifejezések háttérben például az AZ IDŐ PÉNZ, illetve az A VITA CSATA fogalmi metaforák állnak: a forrástartományokat a pénzről és a csatáról alkotott konkrét tapasztalatunk képezi, amelyeket aztán az 'idő' és a 'vita' elvont fogalmak strukturálására használunk fel. A sokak által vitatott [pl. 5, 18] kognitív nyelvészeti álláspont erős verziója szerint az ember képtelen absztrakt fogalmakról konkrét terminusok nélkül gondolkodni.

Az absztrakt tudás mibenlétének másik magyarázata, a statisztikai tanulás elmélete szerint a nyelv statisztikai tulajdonságai segítségével sajátítjuk el és strukturáljuk absztrakt fogalmainkat [2, 16]. A nézet azt hirdeti, hogy a nyelvhasználó a számára ismert nyelvi szimbólumok disztribúciós tulajdonságaiból kiindulva értelmez és használ újabb nyelvi szimbólumokat, beleértve az elvont fogalmakat kódoló nyelvet is.

A két megközelítés elméletileg megfér egymás mellett, hiszen elképzelhető, hogy absztrakt tudásunk mindkét forrást felhasználva alakul ki. A testesültség és a statisztikai elmélet közti különbséget tulajdonképpen arra a kérdésre vezethetjük vissza, hogy lehetséges-e szimbólumlehorgonyzás (jelentés kialakulása) kizárólag nyelvi szimbólumokra épülve. A metaforamegértésre nézve pedig az a lényegi kérdés, hogy függetlenek-e az absztrakt fogalmak a konkrét fogalmaktól a nyelvhasználat során. A két szemlélet a különféle kognitív rendszerek és modalitások szerepét és súlyát vitatja az absztrakt tudásunk reprezentációjának kérdése kapcsán.

Jelen tanulmány a konkrét—absztrakt (vagy metaforikus) nyelvhasználat korpuszalapú elemzésének segítségével arra keresi a választ, hogy a fogalmi metafora hipotézis milyen mértékben járulhat hozzá a nem szó szerinti (vagy metaforikus) jelentések értelmezéséhez természetes nyelvi szövegekben. [12] és [14] metaforaindexéből 12, a nemzetközi szakirodalomban ismert fogalmi metaforát választottunk ki, amelyeket különböző szövegtípusokon teszteltünk. A hipotézis szerint egyrészt azt feltételeztük, hogy ha egy adott mondat tartalmaz egy elvont céltartományhoz tartozó fogalmat kódoló kifejezést, akkor a megfelelő forrástartományhoz tartozó kifejezést is tartalmaznia kell [17]. Másrészt, ha egy mondat tartalmaz egy forrás- és egy céltartományi kifejezést is, akkor a mondat nagy valószínűséggel metaforikus lesz.

2 Korábbi eredmények a testesültség hipotézis mellett és ellen

2.1 Pszicholingvisztikai kísérletek

Gibbs és Matlock [8] a testesültség hipotézist, pontosabban a metaforikus szimuláció (*metaphorical simulation*) jelenségét vizsgálja. Eszerint a metaforikus kifejezések értelmezése során egyfajta fizikai mozgásszimulációt végzünk, vagyis elképzeljük a metaforikusan használt szó által leírt konkrét cselekvést vagy eseményt. Kísérleti bizonyítékaink vannak arra, hogy például a szenzomotoros tapasztalat nagyban befolyásolja az időről szóló metaforikus kifejezések értelmezését [1], akárcsak a párkapcsolatokat utazásként leíró szövegek megértését [8]. A szerzők értelmezése szerint a testi szimuláció, azaz a szó szerinti jelentés mentális aktiválódása segíti a figuratív nyelv feldolgozását.

Ezek az eredmények azt a feltételezést támasztják alá, miszerint a metaforák megértésekor valóban az absztrakt céltartománynak a konkrét forrástartományra való leképezése történik meg. Ugyancsak ezt a feltételezést erősítik meg egyes lexikális döntési feladatok, melyek során azt figyelték meg, hogy az A DÜH EGY TARTÁLYBAN LÉVŐ FELHEVÍTETT FOLYADÉK (ANGER IS A HEATED FLUID IN A CONTAINER) vagy az AZ OPTIMIZMUS FÉNY (OPTIMISM IS LIGHT) konceptuális metaforákat tartalmazó kifejezések után a kísérleti alanyok gyorsabban döntöttek arról, hogy a *hőség*, illetve a *fény* lexémák valódi szavak-e, mint az ilyen metaforákat nem tartalmazó kifejezések után [7].

Akadnak azonban olyan vizsgálatok is, amelyek az eddig elmondottakkal ellentétes következtetésre jutottak. Keysar és munkatársai [10] olyan szövegek megértésének feldolgozási idejét mérték, amelyek az A SZERELEM EGY BETEG PÁCIENS (LOVE IS A PATIENT), az A VITA UTAZÁS (ARGUMENT IS JOURNEY) és az A GONDOLATOK EMBEREK (IDEAS ARE PEOPLE) fogalmi metaforákat tartalmazták. A kísérlet olyan újszerű metaforikus kifejezések megértését tesztelte, amelyeket vagy a már említett metaforatípus konvencionális példái előztek meg, vagy pedig nem metaforikus mondatok. Az eredmények szerint az első esetben a megértés nem volt gyorsabb, mint a másodikban, ami arra utal, hogy a konvencionális kifejezések feldolgozása során nem aktiválódtak a megfelelő fogalmi leképezések. Az AZ IDŐ TÉR (TIME IS SPACE) fogalmi metafora vizsgálatát célzó kísérletek eredményei szintén nem adnak egyértelmű választ arra a kérdésre, hogy a téri sémák feltétlenül szükségesek-e az időről való gondolkodáshoz [21].

2.2 Korpuszelemzések eredményei

A korpusnyelvészeti módszereket segítségül hívó kutatók az elméleti megközelítések sokféleségének és a pszicholingvisztikai kísérletek ellentmondásos eredményeinek problémáját általában abban látják, hogy azok egyrészt túlságosan a metaforák fogalmi természetével vannak elfoglalva, és így figyelmen kívül hagyják a nyelvi tényezőket, másrészt nem természetes nyelvi adatokat használnak a kísérletek lebonyolításához, hanem nyelvi intuíciókon alapuló kitalált példákat, amelyek félrevezetőek lehetnek.

Stefanowitsch [20] az érzelmekkel kapcsolatos metaforák korpusalapú elemzése során azt találta, hogy az ún. metaforikus sablon módszer (*metaphorical pattern method*), amely a metaforák céltartományára jellemző szavak korpusokban való vizsgálatát jelenti, jóval hasznosabb lehet az elméleti kutatók által használt introspekciónál – két okból is: az egyik, hogy ezzel a módszerrel olyan metaforatípusokat is fel lehet lelteni, amelyekről eddig nem esett szó a szakirodalomban, a másik pedig, hogy a gyakorisági mutatókat figyelembe véve meg lehet határozni, hogy az egyes céltartományokat mely leképezések jellemzik leginkább. A fogalmi metafora hipotézis szerint például a BOLDOGSÁG céltartományt a következő forrástartományok strukturálják: FENT, FÉNY, MELEGSÉG, TERMÉSZETI ERŐ stb. Stefanowitsch az általa használt módszerrel további forrástartomány-típusokat határozott meg, amelyek szintén a BOLDOGSÁG absztrakt kategóriát írják le: FOLYADÉK, ÖSSZETÖRHETŐ TÁRGY, BETEGSÉG, AGRRESSZÍV ÁLLATI VISELKEDÉS, ORGANIZMUS stb.

A nyelvi metaforák grammatikai viselkedésének vizsgálata is olyan fontos részletekre világít rá, amelyeket a fogalmi metafora hipotézisben figyelmen kívül hagynak. Deignan [4] elemzéseiből kiderül, hogy a különböző szavak, kifejezések többnyire más-más grammatikai jellemzőkkel, illetve logikai relációkkal rendelkeznek a szó szerinti és a metaforikus használatban. Az AZ EMBERI VISELKEDÉS ÁLLATI VISELKEDÉS fogalmi metafora esetén például azok a szavak, amelyek a forrástartományban szerepelnek, és entitásokat jelölnek, metaforikus használatukban többnyire igeiként vagy melléknévként fordulnak elő. A szerző egyéb metaforatípusok vizsgálata alapján számos példával mutatja meg, hogy metaforikus használatban a szavak jóval kevesebb grammatikai szabadsággal rendelkeznek, mint amikor szó szerinti jelentésükben jelennek meg. Ez azt jelenti, hogy a forrástartományban lévő entítások közti logikai reláció nem egyszerűen megismétlődik a céltartományban, ahogyan azt a kognitív metaforaelmélet jósolná, hanem át is alakul: a szavak metaforikus jelentésükben önálló életet kezdenek élni.

Természetesen olyan elemzések is léteznek, amelyek alátámasztják a fogalmi metaforákon alapuló megközelítést. Martin [17] a metaforákat megelőző kontextusokat vizsgálva azt találta, hogy azok a kontextusok jósolják be legmegbízhatóbban a célmetaforát, amelyek ugyanolyan típusú metaforikus kifejezéseket tartalmaznak, a legkevésbé pedig azok, amelyekben a forrástartomány szavai szó szerinti jelentésükben fordulnak elő. A szerző szerint ez az eredmény azt a korábbi kísérletet erősíti meg, amelyben azt találták, hogy a metaforikus kontextus felgyorsítja a célmetafora megértését, a forrástartomány szavainak szó szerinti jelentésben való használata pedig gátolja azt.

A fentebb bemutatott korpusznyelvészeti elemzések alapján a fogalmi metafora hipotézist és a pszicholingvisztikai kísérleteket célzó egyik legfontosabb kritika abban áll, hogy nem fektetnek elég hangsúlyt a metaforikus nyelvhasználat nyelvi jellemzőire. Ezek az adatok azonban, mint kiderült, igen fontosak, hiszen rámutatnak, hogy olyan egyéb tényezők is szerepet játszanak a figuratív nyelvhasználatban, mint a gyakoriság, a kollokáció, a nyelvi sablonok, a grammatikai formák, továbbá a nyelvi és szövegtípusbeli változatosságok.

3 A korpuszpépítés

3.1 A korpuszelemzés módszertani kérdései

A metaforák korpuszalapú vizsgálata során több nehézséggel állunk szemben. Egyrészt a korpusz kiválasztása önmagában is meghatározó jelentőségű lehet, másrészt pedig a metaforikus kifejezések szövegekben való azonosítása sem problémamentes. Ez utóbbi azért okoz nehézséget, mert a kognitív szakirodalomban tárgyalt fogalmi leképezések általában nincsenek sajátos nyelvi formákhoz kötve, és így nem könnyű meghatározni azokat a nyelvi jegyeket, amelyek leginkább jellemezhetik az egyes tartományokat. Az egyik lehetséges módszer így a kézi keresés, amely során a kutatók saját nyelvi intuícióikra támaszkodva próbálják összegyűjteni egy adott korpuszból a szerintük metaforikusnak ítélt kifejezéseket. Mivel ez az eljárás meglehetősen idő- és munkagényes, legalább részben automatizált módszerekkel is érdemes pró-

bálkozni. Ilyen módszer a forrástartomány szókincsére való rákeresés (pl. Deignan elemzése). Ebben az esetben összegyűjtjük egy adott metaforatípus forrástartományára potenciálisan jellemző szavakat, majd megnézzük, hogy milyen arányban fordulnak elő ezek metaforikus értelemben. Egy harmadik módszer a céltartomány szókincsére való rákeresés (pl. Stefanowitsch elemzése), amely talán azért lehet sikerebb, mint az előző kettő, mert azokban a metaforikus mondatokban, amelyek tartalmaznak egy céltartományi kifejezést, általában egy forrástartományi kifejezés is megjelenik, s így nagyobb az esély az ún. metaforikus sablonok fellelésére. Végül negyedik módszerként olyan mondatokra is rákereshetünk, amelyek egy adott metaforának mind a forrás-, mind pedig a céltartományára jellemző szavakat is tartalmazzák (pl. Martin módszere). Ennek az eljárásnak az a hátránya, hogy így csak előre meghatározott metaforikus leképezéseket tudunk tesztelni, és a Stefanowitsch-féle módszerrel szemben az új metaforatípusok fellelése eleve kizárt. Ezzel szemben nagy előnye, hogy gyorsabban megy az annotálás, így nagyobb szövegeken is alkalmazható.

Természetesen mindegyik esetben szükség van egyrészt megfelelő szolisták összeállítására, másrészt pedig annak explicit meghatározására, hogy mi számít metaforikus kifejezésnek, és mi nem.

Az eddigi korpusznyelvészeti kutatások nagyrészt a metaforikus kifejezések nyelvi jellemzőire voltak kíváncsiak, ezért általában az első három elemzési módszer valamelyikét alkalmazták. Ezzel szemben jelen tanulmány elsősorban arra keresi a választ, hogy a fogalmi metaforáknak szövegekben való automatikus megtalálása mennyire sikeres a testesültség hipotézisét alapul véve. A megfelelő korpusz és elemzési módszer kiválasztására nézve ez a következőket jelentette:

- többféle fogalmi metaforát tesztelni;
- olyan korpuszt vizsgálni, amely többféle szövegtípusból áll;
- olyan példákat találni, amelyek mind a forrás-, mind pedig a céltartomány jellemző szavait tartalmazzák;
- kimerítő listát összeállítani mindkét tartományra vonatkozóan.

Ennek megfelelően [12] és [14] metaforaindexéből 12 széles körben ismert fogalmi metaforát választottunk ki, melyek közül az egyiknek mindkét irányú megvalósulását külön vizsgáltuk (a több fent van/a kevesebb lent van), így tulajdonképpen 13-féle annotáció lehetséges (a példák az általunk annotált szövegekből származnak):

1. A VÁLTOZÁS MOZGÁS (CHANGE IS MOTION): *jön a hideg; rohamléptekkel közeledik a szünidő; mélységes szomorúság járta át a lelkem*
2. AZ IRÁNYÍTÁS FENT VAN (CONTROL IS UP): *magas rangú katonatiszt; az amerikai parti őrség és a haditengerészet járőrei felügyelik a houstoni csatornát*
3. A TÖBB FENT VAN (MORE IS UP): *magasabbra kúszik az átlaghőmérséklet; mértéken felül bosszantott az ismeretlen ideetlen tréfája*
4. A KEVESEBB LENT VAN (LESS IS DOWN): *mély hangját lehalkítva folytatta; le szállították igényüket kétszáz rúpiáról egy fém pumpára; lelohad a szerelem*
5. A HALADÁS ELŐRE MOZGÁS (PROGRESS IS MOTION FORWARD): *a műszaki haladás [...] előrevihet bennünket ezen az úton; rendbe jönnek a dolgok*

6. AZ ERŐFORRÁSOK ÉTELEK (RESOURCES ARE FOOD): *rengeteg áramot fogyaszt; finom artériahálózat táplálja vérrel a sebészek beható vizsgálata alatt álló régiót*
7. AZ ELME GÉPEZET (THE MIND IS A MACHINE): *fokozni akarják szellemi kapacitásukat; kattant valami Mihail Alekszandrovics fejében*
8. AZ IDŐ PÉNZ (TIME IS MONEY): *nem pazarolom az időmet; mennyi időbe kerül a kivitelezés*
9. A DÜH HŐSÉG (ANGER IS HEAT): *a vita hevében elfelejtettem bemutatkozni; a lobbanékony helytartó milyen különös formában torolja meg*
10. A KONFLIKTUS TŰZ (CONFLICT IS FIRE): *le akartad rombolni a templomot, s erre tüzelted a népet; kitört a háború*
11. AZ ELMÉLETEK ÉPÜLETEK (THEORIES ARE BUILDINGS): *a genetika alapjai; az öreg előbb megdöntötte mind az öt bizonyítékot, és aztán (...) ő maga felállított egy hatodikot*
12. AZ ALKOTÁS ÉPÍTÉS (CREATION IS BUILDING): *alapjaiban kell átalakítanunk az életünket; így formálhattak jogot a meghódított területekre*
13. A POLITIKA HÁBORÚ (POLITICS IS WAR): *a velejéig korrupt kormány és rendőrség kirabolja a népet; csakis az ellenséges propaganda állíthatja*

Mivel széles körben használt fogalmi metaforákat választottunk a testesültség hipotézisének korpuszalapú vizsgálatára, minél reprezentatívabb korpuszt kellett építenünk, amiben mindegyik választott metaforatípus elég sokszor fordul elő. A projekt eredeti célkitűzései között szerepelt, hogy a metaforákat többnyelvű párhuzamos korpuszon vizsgáljuk, ezért olyan szövegeket kellett szereznünk, melyek mind a 4 előirányzott nyelven (magyar, angol, spanyol, olasz) elérhetőek és szabadon felhasználhatók kutatási célokra. Jogtiszta szövegeket gyűjteni mind a 4 nyelven meglehetősen nehéz, sokszor kivitelezhetetlen feladatnak bizonyult, így végül magyar nyelvű korpuszunk csak 3 szövegtípust tartalmaz: regények, *National Geographic* cikkek és filmfeliratok az alábbi arányban:

1. táblázat: A korpusz összetétele.

Szövegtípus	Szövegszavak száma
National Geographic cikkek	68 997
Filmfeliratok	32 148
Regények	208 384
Összesen	309 529

Mivel a szövegek különböző formátumokban kerültek a birtokunkba, először is egységesíteni kellett őket: minden dokumentumot UTF-8 karakterkódolású sima szöveggé alakítottunk. A korpuszt szövegszavakra és mondatokra bontottuk, majd minden szövegszóhoz egyértelmű morfológiai elemzést rendeltünk a HunPos morfológiai egyértelműsítő [9] segítségével.

3.2 A gold standard korpusz

A cél tehát olyan korpusz létrehozása volt, amelyben a metaforikus mondatok meg vannak jelölve azzal a címkével, amelyik fogalmi metaforához tartoznak a kiválasztott 13 közül. Feltételezésünk szerint az a mondat metaforikus, amelyben egyaránt megtalálható egy fogalmi metafora forrás- és céltartományához tartozó szó. Módsze-reink teszteléséhez szükségünk volt egy *gold standard* korpuszra, melyben a metafo-rikus mondatok kézzel be vannak jelölve.

Erre a célra építettünk egy minikorpuszt a teljes korpusz 10 százalékából. A minikorpusz tehát kb. 30000 szövegszóból áll, és a teljes korpuszt arányosan repre-zentálja. A minikorpuszt 3 részre osztottuk; mindegyik részben 2 annotátor kézzel bejelölte az általa metaforikusnak ítélt mondatokat. A metaforikusság meghatározásá-hoz [19] kritériumait vettük alapul, néhány ezek közül: az állandósult kifejezéseket, „halott metaforákat” vagy azokat, amelyek csak etimológiai szempontból számítanak metaforáknak, nem vettük figyelembe (pl. a *depresszió* nem számít metaforikusnak); az igekötők számítanak (a *le* vagy *fel* mint FENT, illetve LENT forrástartományok); az allegóriák nem; ha egy metaforának az ellentettjét találtuk, akkor azt nem vettük bele az adott metaforatípusba. Ezenkívül az összes vizsgált típusnál külön-külön röviden össze is foglaltuk a fontosabb útmutatásokat. Például az A TÖBB FENT VAN fogalmi metaforánál a következő útmutatót használtuk: „Minden olyan mennyiséget jelentő kifejezést annotálunk, amit vertikális skálán képzelünk el, pl. *ár, bér, hőmérséklet*. Minden olyat annotálunk, amiben szerepel a *csúcs* szó: *csúcstermelés, csúcstechnoló-gia* stb. Az olyan kifejezések, amelyek arról szólnak, hogy valamiből sok van, és az egy nagy kupacot alkot – pl. *hegyekben áll, tornyosul* –, nem metaforák, nem annotál-juk.”

Az annotátorok közötti egyetértés megállapításához a legegyszerűbb egyetértési mértéket használtuk, ami azt számolja, hogy az esetek hány százalékában ítélték azo-nosan az annotátorok. Mivel rendkívül alacsony értékeket kaptunk – ami arra utal, hogy a metaforikusság definíciója eleve kérdéses, nehezen meghatározható –, úgy döntöttünk, hogy az annotátorok által metaforikusnak ítélt mondatok unióját vesszük. Így 155 metaforikusnak annotált mondat szerepel a *gold standard* minikorpuszunkban.

3.3 Az automatikus azonosításhoz használt szólisták összeállítása

A metaforák automatikus azonosításához Martin módszerét [17] alkalmaztuk, vagyis olyan mondatokat kerestünk, amelyekben mindkét tartomány kifejezései szerepeltek egyazon mondaton belül. A hipotézis alapján azt feltételeztük, hogy ha egy mondat tartalmaz egy forrás- és egy céltartományi kifejezést is, akkor az jó eséllyel metafo-rikus lesz. Ehhez szükségünk volt forrás- és céltartományi szavakat tartalmazó szólis-tákra. Illusztrációként álljon itt egy minta az AZ ELME GÉPEZET fogalmi metaforához tartozó listákból:

2. táblázat: Az AZ ELME GÉPEZET fogalmi metaforához tartozó forrás- és céltartományi szólisták részlete.

the_mind_is_a_machine_source	the_mind_is_a_machine_target
erő	képzelet
kapacitás	szellemi
élesít	memória
működik	agykéreg
feldolgoz	információ
aktiválódik	agyterület
végrehajt	homloklebeny
létrehoz	fej

A forrás- és a céltartomány szólistáinak összeállítását három különböző módszerrel végeztük: a) asszociációs kísérlet alapján, b) szinonimaszótár alapján és c) referenciakorpusz alapján.

Az első módszer esetében a pszicholingvisztikai vizsgálatok körében bevett asszociációs kísérletet választottuk. 138 egyetemi diák végezte el a kísérletet, ami a következőképpen zajlott: a kiválasztott fogalmi metaforák hívószavai megjelentek a képernyőn, majd a kísérleti személynek egy perc állt a rendelkezésére, hogy olyan szavakat írjon, amelyek a tesztszóról eszébe jutnak. Például az A VÁLTOZÁS MOZGÁS metafora esetében a *változás* szó jelent meg a képernyőn mint forrástartományi, és a *mozgás* szó mint céltartományi tesztszó. Az így kapott listákat normalizáltuk: kiszűrtük a többszavas kifejezéseket, a tulajdonneveket és az ellentéteket, feloldottuk a rövidítéseket, majd töveltek a szavakat a Hunmorph morfológiai elemző [22] segítségével.

A második módszer során az asszociációs kísérletből nyert szólistákat kibővítettük a szavak szinonimáival a *Magyar szókincstár* [11] alapján. Ennek hatására a listák mérete természetesen sokszorosára nőtt, annak ellenére, hogy a szinonimák közül a népnyelvi, szleng és ritkán használt szavakat kihagytuk.

A harmadik módszer során [17] alapján tudatosan válogattunk össze szavakat mindegyik forrás- és céltartományhoz az előzőleg kézzel annotált *gold standard* minikorpuszból. Ebből következőleg ezt a módszert a későbbiekben a korpusznak egy másik 10 százalékán teszteltük.

Mindhárom szólista esetében a következő lépésben az eredeti és a morfológiailag egyértelműsített szövegekből, valamint a szólistákból XML-fájlokat állítottunk elő, amelyekben az eredeti szövegek az egyes szólistáknak megfelelően annotációkkal vannak ellátva, azaz a szólisták alapján feltételezett fogalmi metaforák jelölve vannak. Az XML-fájlok minden további korpusznyelvészeti feldolgozó lépését a GATE alkalmazás segítségével végeztük, amely egy könnyen kezelhető grafikus felülettel ellátott szövegfeldolgozó szoftvercsomag [3]. Az automatikusan annotált szöveget kézzel ellenőriztük, és korrigáltuk a szavak többértelműségéből adódó hibákat, azaz kitöröltük azokat a címkéket, amelyek a szólistán szereplő szóval megegyező alakú, de más jelentésű és/vagy szófajú szót jelöltek, pl. az A DÜH HŐSÉG fogalmi metafora esetében az *ég* és *nap* szavaknak az 'égbolt' és '24 óra' jelentésű előfordulásait.

3.4 Eredmények

A három módszer eredményeit az általánosan alkalmazott pontosság és fedés alapján értékeltük ki. A pontosság ebben az esetben azt mutatja meg, hogy az automatikus felismerő rendszer által metaforikusnak ítélt mondatoknak mekkora hányada ténylegesen metaforikus. A fedés értékéből pedig azt tudhatjuk meg, hogy az emberi elemzők által metaforikusnak ítélt mondatok közül hányat talált meg a rendszer. Az F-mérték pedig ezek súlyozott harmonikus közepe, vagyis a hatékonyság végső mérőszáma.

3. táblázat: A három módszer eredményei.

Módszer	Fedés	Pontosság	F-mérték
Asszociáció	6/155 (3,8%)	6/80 (7,5%)	5,65%
Szótár	28/155 (18,06%)	28/617 (4,5%)	11,28%
Korpusz	41/131 (31,29%)	41/74 (55,4%)	43,34%

Az eredményekből azt láthatjuk, hogy az asszociációs módszerrel lényegesen kevesebb olyan mondatot találtunk, amely forrás- és céltartományi kifejezést is tartalmaz, mint a másik két módszerrel. Pontosság tekintetében az asszociációs kísérlet valamivel jobb eredményre vezetett, mint a szótáralapú módszer, de mindkettő messze elmarad a korpuszalapú annotáció eredményétől. Ez utóbbi módszer bizonyult a legeredményesebbnek a fedés tekintetében is, vagyis a metaforikusság gépi azonosításában nem célravezető az asszociációs módszeren alapuló pszicholingvisztikai megközelítés, ellentétben a célzott kézi válogatással.

3.5 Problémás esetek

Az eddigiekből is tisztán látszik, hogy nem könnyű feladat egy mondatról eldönteni, hogy metaforikus-e vagy sem. Általános tapasztalat, hogy ha emberi erővel nehéz megtalálni egy szövegben bizonyos elemeket, akkor azok automatikus azonosítása sem fog jó eredményt hozni. Összességében ki kell mondanunk, hogy a testesültség elméletén alapuló feltételezésünk, miszerint egy metaforikus mondatban meg kell jelennie mindkét tartományi kifejezésnek, nem helytálló. Ezt erősítik a korpusz kézi annotálása során gyűjtött példák is, melyek metaforikusak ugyan, de nem szerepel bennük mindkét tartományhoz tartozó kifejezés, vagyis a fedés szempontjából problémásak.

Bizonyos mondatokban csak forrástartományi szót találunk: *Aztán egy nap lelépett* (A VÁLTOZÁS MOZGÁS). Itt csak egy mozgást kifejező szó van a mondatban, míg a változásra expliciten nem utal semmi, mégis pontosan tudjuk, hogy nem a járdáról való lelépésről van szó, hanem az esemény szereplőinek életében bekövetkezett változásról. Más esetekben a céltartományi kifejezés szerepel ugyan a szövegben, de nem a célmondatban, hanem az azt megelőző szöveggörnyezetben: *Van Toch kapitány majd megfullad a felháborodástól [...] arca bíborszínt öltött [...] feje elkékült* (A DÜH HŐSÉG).

Olyan esetek is léteznek, amikor legkevésbé sem a mondaton múlik annak metaforikussága, hanem csupán egy szón, amely magába foglalja a forrás- és céltartományi jelentést is: *előléptetés* (A HALADÁS ELŐRE MOZGÁS).

Továbbá olyan mondatból is sok van, amelyben szerepel ugyan mindkét tartományi kifejezés, mégsem metaforikus: *Mérnökök és vezetők tanakodnak kisebb csoportokban a 23 emelet magas fűrótorony tövében* (AZ IRÁNYÍTÁS FENT VAN). Ez utóbbi mondatok felelősek az alacsony pontossági értékekért.

4 Tanulságok, összegzés

Mivel kutatásunk célja elsősorban a fogalmi metaforák szövegekben való automatikus azonosítására irányult, nem tértünk ki a talált példák grammatikai elemzésére és a szövegek típusainak a különböző metaforákkal való összefüggéseire sem. Első ránézésre azonban úgy tűnik, vizsgálatunk eredményei megerősítik az előzőekben bemutatott korpusznyelvészeti elemzések eredményeit, főként a kollokációkat és a metaforikus kifejezések nyelvi formáját illetően. Erre utal az is, hogy míg az asszociációs kísérlet segítségével összeállított listák bejósoló ereje nagyon gyenge volt, addig a tartományokra jellemző szavak szövegből való célzott összeválogatása hozta a legjobb eredményt, ami tehát azt jelenti, hogy nem bármilyen asszociáció vezet metaforához, hanem csak bizonyos szavak, kifejezések együttes előfordulása. Például a *piazarol* és *idő* vagy a *gerjeszt* és *harag* szavak egy mondaton belül szinte mindig metaforát eredményeznek. Ezenkívül a grammatikai forma fontosságát érintő adatokat is említhetünk: az AZ ERŐFORRÁSOK ÉTELEK fogalmi metaforánál a referenciakorpusz alapján a forrástartományt leginkább igék jellemzik (pl. *fogyaszt*, *felfal*, *táplál*); ezzel szemben az asszociációs módszerrel összegyűjtött szavak többsége főnév (pl. *edény*, *fagylalt*, *reggeli*). Ez megint csak azt a Deignan [4] által kapott eredményt támasztja alá, miszerint a metaforikus kifejezések jelentős hányadában a forrástartományt képviselő szavak többnyire igeeként vagy melléknévként jelennek meg, aminek az lehet a magyarázata, hogy a metaforikus beszédben általában az absztrakt entitásokat próbáljuk leírni, és így a konkrét forrástartományból leginkább viselkedést, tulajdonságokat, cselekvést leíró szavakat veszünk át. Az asszociációs kísérlet egyik gyenge pontja tehát az lehetett, hogy bármilyen szót figyelembe vettünk, függetlenül annak szófajától.

Természetesen ezeknek a feltevéseknek az alátámasztásához az alátámasztásához a talált metaforák teljesebb elemzésére van szükség. Ugyanakkor az eredeti célkitűzést követve ugyanazeknek a szövegeknek az angol, spanyol és olasz nyelvű változatát is érdemes lesz a jövőben megvizsgálni, és az eredményeket összevetni a magyar adatokkal, hiszen abból újabb következtetéseket vonhatunk le a nyelvi tényezőkre és a fogalmi metaforák természetére vonatkozóan.

Bibliográfia

1. Boroditsky, L., Ramscar, M.: The roles of body and mind in abstract thought. In: *Psychological Science* Vol. 13 (2002) 185–188
2. Burgess, C., Lund, K.: Representing abstract words and emotional connotation in high-dimensional memory space. In: *Proceedings of the Cognitive Science Society*, Lawrence Erlbaum Associates, Hillsdale, NJ (1997) 61–66
3. Cunningham, H., Maynard, D., Bontcheva, K., Tablan V.: GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In: *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*. Philadelphia (2002)
4. Deignan, A.: *Metaphor and corpus linguistics*. John Benjamins, Amsterdam/Philadelphia (2005)
5. Gentner, D., Bowdle, B., Wolff, Ph., Boronat, C.: Metaphor is like analogy. In: Gentner, D., Holyoak, K. J., Kokinov, B. N. (szerk.): *The analogical mind: Perspectives from cognitive science*. MIT Press, Cambridge MA (2001) 199–253
6. Gibbs, R. W.: *Embodiment and cognitive science*. Cambridge University Press, New York (2006)
7. Gibbs, R. W.: Experimental tests of figurative meaning construction. In: Radden, G., Köpcke, K-M., Berg, Th., Siemund, P. (szerk.): *Aspects of meaning construction*. John Benjamins, Amsterdam/Philadelphia (2007) 19–33
8. Gibbs, R. W., Matlock, T.: Metaphor, imagination and simulation. Psycholinguistic evidence. In: Gibbs, R. W. (szerk.): *The Cambridge Handbook of Metaphor and Thought*. Cambridge University Press, Cambridge (2008) 161–176
9. Halácsy, P., Kornai, A., Oravecz, Cs.: HunPos - an open source trigram tagger. In: *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics. Companion Volume Proceedings of the Demo and Poster Sessions*. Association for Computational Linguistics, Prague, Czech Republic (2007) 209–212
10. Keysar, B., Shen, Y., Glucksberg, S., Horton, W. S.: Conventional language: How metaphorical is it? *Journal of Memory and Language* Vol. 43 (2000) 576–593
11. Kiss G.: *Magyar szókincstár*. Tinta, Budapest (2007)
12. Kövecses, Z.: *Metaphor. A Practical Introduction*. University Press, Oxford (2002)
13. Kövecses, Z.: *A metafora. Gyakorlati bevezetés a kognitív metaforaelméletbe*. Typotex, Budapest (2005)
14. Lakoff, G., Johnson, M.: *Metaphors we live by*. University of Chicago Press, Chicago (1980)
15. Lakoff, G., Johnson, M.: *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books, New York, NY (1999)
16. Landauer, T.K., Dumais, S.T.: A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. In: *Psychological Review* Vol. 104 No. 2 (1997) 211–240
17. Martin, J.H.: A corpus-based analysis of context effects on metaphor comprehension. In: Stefanowitsch, A., Gries, S.Th. (szerk.): *Corpus-based approaches to metaphor and metonymy*, Mouton de Gruyter, Berlin/New York (2006) 214–236
18. Murphy, G. L.: On metaphoric representation. *Cognition* Vol. 60 (1996) 173–204
19. Pragglejaz Group: MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol* Vol. 22 No. 1 (2007) 1–39
20. Stefanowitsch, A.: Words and their metaphors: a corpus-based approach. In: Stefanowitsch, A., Gries, S. Th. (szerk.): *Corpus-based approaches to metaphor and metonymy*. Mouton de Gruyter, Berlin/New York (2006) 63–105

21. Szamarasz, V. Z.: Az idő téri metaforái: a metaforák szerepe a feldolgozásban. *Világosság* Vol. 47 (2006) 99–109
22. Trón, V., Németh, L., Halácsy, P., Kornai, A., Gyepesi, Gy., Varga, D.: Hunmorph: open source word analysis. In: *Proceedings of the ACL 2005 Workshop on Software* (2005) 77–85