

## Fordíthatóság az angol és magyar idiómák tükrében Egy angol–magyar idióma-adatbázis létrehozásáról

Simkó Katalin Iлона

SZTE BTK Nyelvtudományi Doktori Iskola  
simko@hung.u-szeged.hu

**Kivonat:** A dolgozat többszavas kifejezéseket, idiómákat vizsgál a fordíthatóságuk szempontjából. A dolgozat bemutatja egy lehetséges, fordíthatóság-alapú, idiomatikus jelentésen, lexikai elemeken és szintaktikai szerkezeten alapuló idiómacsoportosítás szempontjait. Valamint leírja a kidolgozott csoportokat és egy ezt felhasználó angol–magyar nyelvpárra létrehozott idiómaadatbázist is. A dolgozat célja annak bemutatása, hogy az idiómák fordíthatóságáról alkotott alapvető elképzelésinket a konkrét adatok sok esetben megcáfolják; az idiómák viselkedése egyes nyelvpárok között összetettebb annál, mint amit elvárunk. Az idiómák szerepe a nyelvtanításban és a fordításban is nagy, a dolgozat és az adatbázis célja segíteni ezeket a folyamatokat.

### 1 Bevezetés

Az idiómák egyik legalapvetőbb tulajdonsága a nem kompozicionális jelentésük, ami miatt egyik nyelvről a másikra fordításuk kompozicionális szövegekhez képest más megoldásokat igényel. Ennek ellenére egy adott nyelvpár esetén egyes idiómák nem is annyira nem kompozicionálisan viselkednek, mint gondolnánk, amint azt a spanyol dialektusok kapcsán Bogantes és munkatársai is megfigyelték (Bogantes et al. 2015).

Egy szöveg fordításakor különböző idiómákat különbözőképpen tudunk a célnyelvre fordítani annak ellenére, hogy minden esetben az idiomatikus jelentést próbáljuk a másik nyelven is kifejezni. Egyes idiómák esetén a fordításhoz mintha nem is kellene az idiomatikus jelentést figyelembe vennünk, és az idióma „szó szerint” lefordítható a célnyelvre, más esetekben egyáltalán nem találunk a célnyelven megfelelő jelentésű idiómát.

Dolgozatomban bemutatom az általam létrehozott, fordíthatóság szerint csoportosított angol–magyar idióma-adatbázist és a létrehozásához kidolgozott csoportosítást. Az adatbázis és a csoportosítás célja a gépi és géppel segített fordítás, valamint az idegennyelv-tanulás segítése.

A dolgozatban először bemutatom a felhasznált kiinduló adatbázist és a legyűjtés folyamatát. Ezután mutatom be az általam kidolgozott jelenlegi csoportosítást és az így létrehozott jelenlegi adatbázist. Végül a jövőbeli terveimet írom le az adatbázissal kapcsolatban, részben a témával kapcsolatos visszajelzések alapján.

## 2 Szakirodalom

Az idiómák csoportosításával sokan, számos szempontból foglalkoztak. Az idiómák és azok etimológiája kapcsán több kortárs magyar kutató is aktív kutatást végez (például Forgács 2015; Bárdosi 2017). Ezek a munkák elsősorban egynyelvű, etimológiai szempontú vizsgálatot mutatnak be. Jelen munkámban a nyelvek közötti szempontok dominálnak.

A következőkben röviden bemutatom Makkai (1972) főként angol nyelvre fókuszáló idiómaklasszifikációját, valamint Bogantes és munkatársainak spanyol dialektusokkal foglalkozó munkáját (Bogantes et al. 2015).

Makkai egy olyan strukturális keretet dolgoz ki, amelyben az angolra alapozva a lexémikus (lexemic) és szemémikus (sememic) idiómákat kezeli. Munkájában az angolra és a magyarra jellemző idiómákat is összehasonlítja.

A lexémikus idiómák esetén az idiomatikus jelentés az idiómát alkotó nyelvi elemek poliszémiájából adódik. Ezzel ellentétben a szemémikus idiómák valamilyen „tanulásra” épülnek, gyakoriságukból adódóan a hallgató erre a „tanulásra” asszociál akkor is, ha annak konkrét levezetése csak több lépésben lehetséges az eredeti idiómából kiindulva.

Az angol és magyar összehasonlításban Makkai megállapítja, hogy a lexémikus és szemémikus idiómák és azok alosztályai mindkét nyelvben jellemzően előfordulnak az idiómák kifejezésére. Megjegyzi továbbá, hogy egyes konkrét idiomatikus jelentést kifejező idiómák között is találunk párhuzamot a két nyelvben: a *don't count the chickens before they hatch* és *ne igyál előre a medve bőrére* szemémikus idiómák mindkét nyelvben a [TILTÁS] + [ÜNNEPLÉS] + [TÚL HAMAR] + [EREDMÉNY] elemeket tartalmaznak ugyanannak a „tanulásnak” a megfogalmazása céljából (Makkai 1972).

Bogantes és munkatársai a Makkai-munkában is felvetett kérdéssel foglalkoznak: az egyes többszavas kifejezések és azon belül is az idiómák milyen variációt mutatnak egy nyelv dialektusain keresztül? A munka egy spanyol nyelvű idióma-adatbázist, annak felépítését és egy korpuszvizsgálatot mutat be négy latin amerikai dialektusban: a kolumbiai, a Costa Rica-i, a mexikói és a perui spanyolban.

Az adatbázis 250 idiómát tartalmaz, és jelöli, hogy az adott kifejezéshez melyik nyelvjárásban járul idiomatikus jelentés és mi az. A kutatás során három kategóriát különböztetnek meg: egyes esetekben a különböző nyelvjárásokban különböző idióma fejezi ki ugyanazt az idiomatikus jelentést; a második esetben az idióma több nyelvjárásban is ugyanazt az idiomatikus jelentést fejezi ki; végül az az eset is megtalálható, amikor ugyanaz az idiomatikus kifejezés az egyes nyelvjárásokban más-más idiomatikus jelentést hordoz.

Ezután egy webkorpuszban annotálták a kifejezések háromféle előfordulását: idiomatikus jelentésében áll, kompozicionális/szó szerinti jelentésében áll, vagy az idióma összes összetevője szerepel a mondatban, de szintaktikai struktúrájuk nem ugyanaz, mint ami a többszavas kifejezés idiomatikus jelentéséhez szükséges (Bogantes et al. 2015).

### 3 Felhasznált adatbázis

Munkám kiindulópontjaként egy angolul tanuló magyaroknak szánt „idiómalistát” használtam fel.<sup>1</sup> A lista 2463 kifejezést tartalmaz. Minden kifejezés egy magyar fordítással, az angol definíciójával és egy angol nyelvű példamondattal áll. Választásom azért erre a listára esett, mert ez digitalizált formában, ingyenesen elérhető.

Az „idiómalista” elnevezés ellenére a lista nem csak idiómákat tartalmaz: a listán megtalálhatóak vonzatos igék és egyéb kollokációk is, amint az az eredeti lista alapján készített 1. táblázatban is látszik.

Kifejezés	Magyar fordítás	Angol definíció	Angol példamondat
bring up	felhoz témát	introduce a subject into a discussion	They brought up the subject at the meeting but nobody wanted to talk about it.
bring up the rear	sereghajtó	be at the end of the line or in the last position	The runner from the other school was bringing up the rear in the school relay race.
broke	le van égve (anyagilag)	have no money	I spent all of my money on my holiday and I am now broke.
brush up on something	felfrissít (tudást)	review something one has already learned	I'm going to brush up on my English before my trip to New York.
bury/hide one's head in the sand	homokba dugja a fejét; nem vesz tudomást a problémákról	refuse to see or face something, keep from seeing or knowing something unpleasant	He always buries his head in the sand and never wants to deal with his family problems at all.
bury the hatchet	elássa a csatabárdot, kibékül	stop quarreling and become friendly with someone	He decided to bury the hatchet with his brother and they are now on friendly terms again.
by and large	többé-kevésbé, nagyjából	on the whole, considering everything	By and large we had a good meeting even though it was a little short.
by and by	most már nemsokára, lassacskán	before long	By and by they will come and we can go out for dinner.

1. táblázat. Részlet a Sulinet adatbázisból

<sup>1</sup> Elérhető: <http://www.sulinet.hu/nyelvek/?p=content&id=274>.  
Letöltve: 2017. 07. 17.

Munkám során a Sulinet adatbázisból először kiszűrtem a vonzatos igéket (például *bring up*) és kollokációkat (például *by and large* vagy *by and by*). Idiómának a Makkai által szemémikusnak nevezett kategóriát tekintettem, ahol az idiomatikus jelentés valamiféle „tanulság” útján vonható le a szó szerinti jelentésből.

## 4 A csoportosítás

A fordíthatósági csoportosítás kidolgozásánál három szempontot vettem figyelembe: az idiomatikus jelentést, a szintaktikai szerkezetet és a lexikai tartalmat. Idiomatikus jelentés alatt a kifejezés nem szó szerinti jelentését értem, a szintaktikai szerkezet és lexikai tartalom egyezését pedig természetesen ezek „lefordítása” után értelmezem.

Ezek alapján öt kategóriát állapítottam meg:

1. csoport: Azonos idiomatikus jelentés mindkét nyelvben azonos szintaktikai szerkezet és lexikai tartalom mellett
2. csoport: Azonos idiomatikus jelentés mindkét nyelvben, eltérő szintaktikai szerkezet VAGY lexikai tartalom
3. csoport: Azonos idiomatikus jelentés mindkét nyelvben eltérő szintaktikai szerkezet és lexikai tartalom mellett
4. csoport: Azonos szintaktikai szerkezet és lexikai tartalom mindkét nyelvben idiomatikus, de eltérő idiomatikus jelentéssel
5. csoport: Az idiomatikus jelentés csak az egyik nyelvben fejezhető ki idiomatikusán

Az idiómák ilyen típusú csoportosítását hasznosnak vélem a nyelvtanulásban, nyelvtanításban, ahol ezen kategóriák megismerése elősegítheti helyes használatukat a tanult nyelven, valamint géppel segített fordításban, ahol különböző kontextusokban egy-egy idióma különböző típusú fordítása megfelelőbb lehet.

### 4.1 Az 1. csoportba tartozó idiómák

Az első csoportba tartozó idiómák lexikai és szintaktikai felépítése megegyezik a két nyelvben, valamint a két kifejezés ugyanazt az idiomatikus jelentést is hordozza. Ezen idiómák esetén a fordítás megegyezik a nem idiomatikus kifejezések fordításával: a két kifejezést akár egy egyszerű gépi fordítóprogram is megfelelteti egymással, mivel azok szó szerinti és idiomatikus jelentése párhuzamos a két nyelvben.

Ebbe a csoportba tartoznak például angol–magyar nyelvpár esetén az [1], [2] és [3] példákban szereplő idiómák.

- [1] in a nutshell – dióhéjban
- [2] to break the ice – megtöri a jeget
- [3] to throw the first stone – az első követ veti rá

### 4.2 A 2. csoportba tartozó idiómák

A 2. csoport két alcsoportra bontható: az idiomatikus jelentés mindkét esetben megegyezik a két nyelvben, viszont vagy a lexikai, vagy a szintaktikai felépítés megváltozik. Az idiómák között még mindig fellelhető valamilyen kapcsolat, így a

csoportot nyelvtanulás szempontjából fontosnak tartom. Emellett a géppel segített fordításban is hasznos lehet: ha több lehetséges fordítása is van egy adott idiómának, elképzelhetőnek tartom, hogy egyes esetekben a szintaktikai szerkezet, máskor a lexikai elemek megtartása a fontosabb az adott kontextusban.

A 2. csoport első alcsoportjába tartozó idiómapárok például angol–magyar nyelvpárra a [4] és [5] példák. A másik alcsoport ennek az ellentettje: például a [6], [7] és [8] pár.

- [4] one's bark is worse than one's bite – amelyik kutya ugat, az nem harap
- [5] tail between one's legs – behúzza fülét-farkát
- [6] pot calling the kettle black – bagoly mondja verébnek, hogy nagyfejű
- [7] make a mountain of a molehill – bolhából elefántot csinál
- [8] six of one and half-a-dozen of the other – egyik kutya, másik eb

### 4.3 A 3. csoportba tartozó idiómák

A 3. csoportba tartozó idiómák felelnek meg leginkább az idiómákkal kapcsolatos előzetes elvárásainknak: ugyanazt az idiomatikus jelentést teljesen más szerkezet fejezi ki az egyik nyelvben, mint a másikban. A két nyelvben az adott idiomatikus jelentéshez tartozó idiómák lexikai és szintaktikai felépítése is eltér, viszont ezt a jelentést mindkét nyelvben kifejezhetjük idiómával.

Ide tartoznak angol–magyar összehasonlításban a [9], [10] és [11] példák.

- [9] let the grass grow under one's feet - megalszik a tej a szájában
- [10] count one's chickens before they hatch - előre iszik a medve bőrére
- [11] wasn't born yesterday – nem most jött le a falvédőről

### 4.4 A 4. csoportba tartozó idiómák

A 4. csoport a legritkább típus az idiómák fordíthatóságának szempontjából. A csoportba olyan idiómapárok tartoznak, amelyek megegyező lexikai és szintaktikai szerkezettel mindkét nyelvben idiomatikus jelentéssel bírnak, ám ez az idiomatikus jelentés eltér a két nyelvben. A csoport nagy szerepet kaphat a nyelvtanításban. Emellett a gépi fordításban is fontos tiltólistaként: ezeket az idiómákat még a többen is fontosabb lehet felismernünk: mert míg más esetekben egy idióma szó szerinti lefordítása értelmetlen kimenethez vezet, ami nagy eséllyel felhívja a felhasználó figyelmét legalább a probléma jelenlétére, addig az ezen csoportba tartozó idiómák szó szerinti lefordítása egy teljesen más olvasathoz vezethet, mint ami az eredeti szövegben szerepelt.

A csoportba tartoznak például angol–magyar esetén a [12], [13] és [14] idiómapárok.

- [12] one's heart is in the right place („jóindulatú”, „együttérző”) – helyén van a szíve („bátor”)
- [13] hold one's breath („vár valamire”) – eláll a lélegzete („nagyon meglepődik”)
- [14] jump out of one's skin („nagyon megijed”) – majd kiugrik a bőrből („nagyon izgatott”)

#### 4.5 Az 5. csoportba tartozó idiómák

Az 5. csoport a 3.-hoz hasonló, a két nyelvben itt sem találunk lexikai vagy szintaktikai kapcsolatot az idiomatikus jelentés kifejezései között. Ebben a csoportban viszont nem is találjuk a kiinduló nyelv idiómájának idiomatikus megfelelőjét a célnyelvben. A kiinduló nyelvi idióma idiomatikus jelentését a célnyelv csak körülírással vagy szó szerinti kifejezéssel tudja leírni.

Erre jó példák azok az angol idiómák, amelyeknek nem találtam magyar idiomatikus megfelelőjét: a [15], [16] és [17] példák. Fordítottan pedig a [18] példa, ahol a magyar idiómának nem találunk angol nyelvű idiomatikus megfelelőjét.

[15] penny-wise, but pound foolish („kis dolgokban spórol, de nagy dolgokban költekezik”)

[16] little pitchers have big ears („a kisgyerekek mindent meghallanak”)

[17] cross that bridge when we come to it („ezt majd akkor eldöntjük, ha ehhez a problémához értünk”)

[18] nem erőszak a disznótor („nem kell elfogadnod, amit kínálok, bár jót akarok neked”)

A 2. táblázat a csoportok tulajdonságait szemlélteti. A táblázatban X jelöli, ha az idiómapárok az adott idiómacsoport esetén a sorban található tulajdonság szempontjából megegyeznek a két vizsgált nyelvben.

	1. csoport	2. csoport		3. csoport	4. csoport	5. csoport
		A	B			
<b>Idiomatikus jelentés</b>	X	X	X	X		
<b>Lexikai elemek</b>	X	X			X	
<b>Szintaktikai szerkezet</b>	X		X		X	

2. táblázat. A csoportok tulajdonságai

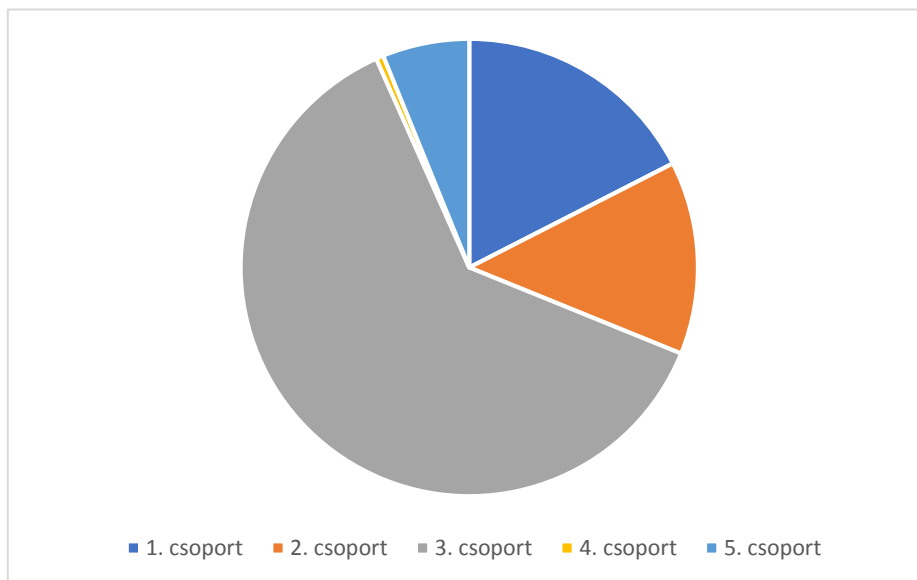
## 5 Az adatbázis

Saját adatbázisom a Sulineten található kifejezések listájából indul ki. A listában található 2463 kifejezést kézzel annotáltam: a kifejezések többségét kizártam, mivel azok vonzatos igék vagy kollokációk voltak, amelyeket nem tartottam relevánsnak a vizsgálat szempontjából.

A megmaradt 944 angol idiomatikus kifejezést magyar fordításuk alapján a 4. fejezetben bemutatott csoportokba osztályoztam. A 3. táblázat az egyes csoportok számosságát és az adatbázis egészéhez viszonyított százalékos arányát tartalmazza, az 1. ábra pedig kördiagramon szemlélteti a megoszlást.

1. csoport	2. csoport	3. csoport	4. csoport	5. csoport
165	129	587	5	58
17,5%	13,5%	62%	0,5%	6%

3. táblázat. A csoportok megoszlása az adatbázisban



1. ábra. Az egyes csoportok megoszlása az angol–magyar adatbázisban

A legnépesebb egyértelműen a 3. csoport, az idiómák 62%-a ebbe a legintuitívabban idiómákra jellemző kategóriába tartozik. Elgondolkodtató azonban, hogy az idiómák majdnem fele valamelyik másik kategóriába tartozik, így más stratégia alkalmazására lehet szükség a fordításban, illetve tudatos elkülönítésük segítheti a nyelvtanulást is.

## 6 További lehetséges kutatási irányok

A jelenlegi adatbázist korpuszvizsgálattal tervezem kiegészíteni: az idiómákat és azok tényleges fordítását egy párhuzamosított, filmfeliratokat tartalmazó korpuszban vizsgálnám (Lison–Tiedemann 2016).

A korpuszelemzéssel célokom annak vizsgálata, hogyan jelennek meg az idiómák a két nyelv tényleges nyelvhasználatában és fordításában.

Emellett az előadás visszajelzései alapján tervezem a csoportosítás kisebb átszervezését. A 4. csoportot külön célszerűbb lenne a többitől elkülönítve kezelni, mivel itt nem az adott angol idióma magyar megfelelőiről van szó; a csoport tényleges fordításait pedig besorolnám a megfelelő másik csoportba.

Az adatbázist tervezem tovább-bővíteni: az egyes idiómák különböző fordítási lehetőségeit is fel szeretném venni a megfelelő csoportokba, és egy azonosítóval összekapcsolnám az egy idiómához tartozó különböző fordításokat. Az adatbázis

idiomaszótárakból és egyéb forrásokból is bővíthető. Ezen felül magyar nyelvű idiómákból kiindulva is kiegészítem az adatbázist.

## 7 Összegzés

Dolgozatomban idiómákat vizsgáltam fordíthatóságuk szempontjából, főként angol–magyar nyelvpáron keresztül.

Bemutattam az általam létrehozott idiómacsoportosítást, amely a két nyelv idiómáinak lexikai felépítésére, szintaktikai szerkezetére és idiomatikus jelentésére épül. Emellett az általam épített angol–magyar idiómákból álló adatbázis leírását adtam.

A témában tervezem az adatbázis kibővítését és kisebb átalakítását az előadás visszajelzései alapján, valamint a kutatás korpuszvizsgálattal való bővítését.

## Irodalom

- Bárdosi, V. 2017. The New Etimological Dictionary of Hungarian Idioms and Proverbs. *Open Linguistics*, 3: 86–99.
- Bogantes, D., Rodríguez E., Arauco A., Rodríguez A., Savary A. 2015. *Towards Lexical Encoding of Multiword Expressions in Spanish Dialects*. Elhangzott: *PARSEME 5th General Meeting*, Iași, 2015. szeptember 23–24.
- Forgács, T. 2015. „Közmondás nem hazug szólás”: A frazeológiai egységek osztályozásának nehézségei. In: Szeverényi, S., Szécsényi, T. (szerk.) *Érdekes nyelvészet*. Szeged: JATEPress Kiadó. 23–26.
- Lison, P., Tiedemann, J. 2016. OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In: In: Calzolari, N., Choukri K., Declerck T., Goggi S., Grobelnik M., Maegaard B., Mariani J., Mazo H., Moreno A., Odiik J., Piperidis S. (szerk.) *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Párizs: European Language Resources Association (ELRA). 923–929.
- Makkai, Á. 1972. *Idiom Structure in English*. Hága–Párizs: Mouton.