

A magyar frazeológiai adatbázis létrehozása és az ebből generált szinonim frazémaszótár munkálatai

Bárdosi Vilmos¹, Kiss Gábor²

¹ ELTE BTK, Francia Tanszék, 1068 Budapest, Múzeum krt. 4-6.
bardosi.vilmos@btk.elte.hu

² TINTA Könyvkiadó, 1116 Budapest, Kondorosi út 17.
kissgabo@tintakiado.hu

Kivonat: A munka során létrehozott magyar frazeológiai adatbázis lineáris elrendezésű frazémagyűjtemény, a definícióikkal együtt tartalmazza a magyar nyelv leggyakoribb kollokációit, szólásait, közmondásait és helyzetmondait. E gyűjteményből a definíciók elemzésével, kulcsszavak felhasználásával gépi transzformációval állítottuk elő a frazémák szinonimikus elrendezését. A transzformáció során első lépésként a frazémadefiníciókat tanulmányoztuk, és definiáltunk 850 kulcsfogalomsort, melyek mindegyike 2-8 szinonim szó alkotott. Az egyes kulcsfogalomsorok alá a frazémák besorolását úgy végeztük, hogy a kulcsfogalomsor szinonim szavait illesztettük az egyes frazémák definíciójára. Az automatikus elrendezést manuálisan javítottuk és finomítottuk (pl. az aspektusbeli jellemzők figyelembe vételével).

1 Bevezetés

Korábbi cikkünkben bemutattuk a magyar szavak jelentéshasonlóságának automatikus feltérképezési lehetőségeit [3]. Jelen cikkünkben frazémák hasonló osztályozására tett kísérletünket vázoljuk fel. Az automatikus szövegfeldolgozás nem nélkülözheti a szavaknál nagyobb lexikai egységek, a frazémák helyes kezelését. Frazéma gyűjtőfogalom alatt értünk minden olyan többbelemű szósort, melynek szavai együttesen mást jelentenek, mint várnánk a szavak pusztá kompozíciójából. A régebbi magyar szakirodalom terminológiája nem volt egységes, számos elnevezés volt használatos: *közmondás, példabeszéd, közbeszéd, közmondat, közszólás, közhasznlat, közpéldabeszéd, közpéldaszó, példaszó, példázat, aggszó, rólabeszéd, szólásmód, szójárás, szólásmondás, szokásmondás, mondás*, ahogy erre Tolnai Vilmos felhívja figyelmünket [10]. Voigt Vilmos A magyar folklór című könyvben a Kisepikai műfajok között tárgyalja a szavaknál nagyobb nyelvi egységeket, és a néprajzi szakirodalomban bevett proverbium terminust használja [11].

Az ezredforduló után szerencsés módon szinte egy időben több magyar frazémaszótár is napvilágot látott [2, 4, 5, 9, 8]. A pusztá leltározást többirányú feldolgozásnak kell követnie. Az egyik lehetséges feldolgozás a magyar frazémakincs szinonimikus elrendezése. Erre a feladatra Tolnai Vilmos már 1935-ben felhívta a figyelmet ezekkel a szavakkal: „A szóláskutatás feladata ... a hasonlóértelmű szólá-

sok csoportosítása, a szólás-szinonimika, melyben nem az alak, hanem az alkalmazás jelentése az osztályozás alapja, pl.: *halálos beteg: nem sokáig viszi, hálni jár belé a lélek, nem lesz többé ember belőle, ütött az utolsó órája.*” [10]. A szerző még a frazémák fogalomköri elrendezését is sürgeti ezekkel a szavakkal: „Feladatunk a fogalmi körök szerint való csoportosítás. Pl.: szólások háziállatainkról; az év napjaihoz fűződő mondások; az időjárás jóslatai; a művelődéstörténet tárgyi csoportjai.” [01]. Az ilyen jellegű hiányt az elméleti háttér feltárása után [1] pótolták a Bárdosi Vilmos szótáraiban található fogalomköri mutatók [2, 4].

2 A frazéma-adatbázis létrehozása és jellemzője

A magyar frazeológiai adatbázis több mint 25.000 frazémát tartalmaz, két nagy gyűjtemény [2, 4, 7] frazémáinak összefésülésével hoztuk létre. E lineáris elrendezésű gyűjtemény a definícióikkal együtt tartalmazza a magyar nyelv leggyakoribb kollokációit, szólásait, közmondásait és helyzetmondait.

A magyar frazéma-adatbázis szerkezete

1. a frazéma; **2.** a frazéma típusa [egyszavas ekvivalenssel rendelkező szókapcsolat | kollokáció | helyzetmondat | szólás | közmondás]; **3.** a stílusminősítés [bizalmas | durva | gúnyos | hivatalos | népi | régi | ritka | rosszalló | szépítő | szleng | tájnyelvi | tréfás | választékos]; **4.** a definíció

A frazémák típusának jelzése

A frazéma-adatbázisban megjelöltük a frazéma típusát. Az alábbi kategóriákat határoztuk meg, és megadjuk, hogy az adott kategóriából kerekítve hány darab található az adatbázisunkban.

EK – Szókapcsolat, melynek van egyetlen szóból álló ekvivalense (260 darab)

Pl.: *gazdasági növény: haszonnövény*

költöző madár: vándormadár

Krisztus földi helytartója: pápa

KO – Kollokáció (1650 darab)

Pl.: *elveszti a fogadást*

kezet fog <vkivel>

szoros eredmény

HMO – Helyzetmondat (4300 darab)

Pl.: *Eszed tokja!*

Meglesz ennek még a böjtje!

Sokra megyek velem!

HMI – Helyzetmondat alete, mely az írott nyelvhasználatra jellemző (15 darab)

Pl.: *Isten dicsőségére állította:*

Kihajolni veszélyes!

Minden külön értesítés helyett.

HMM – Helyzetmondat alete: népmesékre jellemző fordulat (10 darab)

Pl.: *Boldogan éltek, míg meg nem haltak.*

Hol volt, hol nem volt.

Szerencséd, hogy öreganyádnak szólítottál.

MSZ – Helyzetmondat alete: megszólítás, köszönés (60 darab)

Pl.: *A vérfagyasztó vizontlátásra! (biz)*

Jó szerencsét!

Kezét csókolom!

SZO – Szólás (18850 darab)

Pl.: *a szavakon lovagol*

egy híron pendül <vkivel>

szélmalomharcot folytat

SZH – Szólás alete: szóláshasonlat (750 darab)

Pl.: *Áll, mintha gyökeret vert volna a lába.*

Úgy hasonlítanak egymásra, mint két tojás.

Olyan a bőre, mint a rinocérosznak.

KM – Közmondás (950 darab)

Pl.: *A rest kétszer fárad!*

Ahol nincs, ott ne keress!

Reggeli vendég nem sokáig marad (népi)

A stilisztikai minősítések

A frazéma-adatbázisban, ahol szükséges volt, a frazémákat stilisztikai minősítéssel láttuk el. Az alábbiakban közöljük a leggyakoribb minősítéseket, példákkal, és feltüntetjük, hogy az adott stílusminősítéssel hány frazémát láttunk el.

BIZ – a bizalmas nyelvhasználatra jellemző (3050 darab)

Pl.: *Akárcsak a falnak beszélnek!*

csinálja a cirkuszt

nincs egy megveszekedett vasa sem

DUR – a durva, trágár nyelvhasználatra jellemző (350 darab)

Pl.: *Hallgat, mint szar a fűben.*

kitapossa a belét <vkinek>

vén szatyor

GÚNY – gúnyos jelentésű, jelentésárnyalatú frazéma (350 darab)

Pl.: *egy kézlegyintéssel intéz el <vmit>*

feltalálta a spanyolviaszt

Nagy az Isten állatkertje, és sok bolond lakik benne.

HIV – a hivatalos, sajtónyelvi nyelv jellegzetes frazémája (50 darab)

Pl.: *benyújtja a lemondását*

halottá nyilvánít <vkit>

késedelmet szenved

NÉP – a népi és a tájnyelv jellegzetes frazémája (1250 darab)

Pl.: *Guba gubához, suba, subához*

Hencidától Boncidáig folyt a sárga lé.

Mennél koszosabb a malac, annál jobban vakarózik.

RÉG – a régi nyelvben használatos, ma már elavult frazéma (950 darab)

Pl.: *Gyepre legény!*

tetézve adja vissza a csapott vékát

Üres gyomornak nem kell prédikáció.

RIT – ritkán használt, nem gyakori frazéma (800 darab)

Pl.: *magával hord <vmit>, mint a csiga a házát*

ráfordítja a kulcsot <vkire>

zászlós bajusz

ROSSZ – rosszálló stílusértékű frazéma (80 darab)

Pl.: *a színpalak mögött*

egy követ fúj <vkivel>

tartja a markát

SZÉP – szépítő, eufemisztikus frazéma (100 darab)

Pl.: *a szerelem papnője*

magához szólít <vkit> az Úr

testi szükséglet

SZLE – a szlengben használatos frazéma (1150 darab)

Pl.: *eldobja az agyát*

lemegy hídba

tökig nyomja a gázt

TRÉF – a tréfás stílusértékű frazéma (650 darab)

Pl.: *Ha a hegy nem megy Mohamedhez, Mohamed megy a hegyhez.*

Kismise, nagymise, utoljára semmise.

Semmi baja, csak a torka véres.

VÁL – a választékos, az emelkedettebb nyelvhasználatban, az irodalom nyelvében előforduló frazéma (1500 darab)

Pl.: *bűnnek ereszti a fejét*
kiissza a méregpoharat
pusztába kiáltott szó

A többjelentésű frazémák feldolgozása

A nyelvben nem csak a szavak, hanem a frazémák is sok esetben többjelentésűek. Éppen ezért minden frazéma mellett pontosan feltüntetjük minden jelentését. Így vált lehetségessé, hogy a frazémát az adott jelentésének megfelelő szinonimikus csoportba soroljuk be. Mintaképpen bemutatunk néhány többjelentésű frazémát és közreadjuk az egyes jelentésekhez tartozó definíciókat.

1. *benne van a buliban* = előnyös munka, vállalkozás részese
2. *benne van a buliban* = együtt szórakozik a társasággal

1. *előkészíti a talajt* = a szükséges mezőgazdasági munkálatokat a talajon előre elvégzi

2. *előkészíti a talajt* = (átvitt értelemben) tevékenységével v. hatásával vkit, vmit fogékonyra, alkalmassá tesz vmire

1. *felszaggatja a sebet* = a rajta levő kötést és vart letépi

2. *felszaggatja a sebet* = (átvitt értelemben, választékos) vmely már-már elfelejtett lelki fájdalmat felújít, felszínre hoz

1. *Isten malmai lassan örölnek.* = {ha későn is, de utoléri a gonosz embert a méltó büntetés}

2. *Isten malmai lassan örölnek.* = (tréf) {vmely hivatalban lassan intézik az ügyeket}

1. *járatja a pofáját* = dicsekszik

2. *járatja a pofáját* = (gyakran badarságokat fecsegve) feleslegesen és sokat beszél

1. *jó füle van* = élesen hall; kifogástalan zenei hallása van

2. *jó füle van* = (átvitt értelemben) azt is meghallja, amit nem neki szántak

1. *kitér a hitéből* = régi vallását elhagyja és esetleg más felekezet tagja lesz

2. *kitér a hitéből* = (tréf) (vminek hallatán) nagyon bosszús, dühös lesz v. nagyon megbotránkozik

1. *kitöri a nyakát* = baleset (különösen lezuhanás) következtében szörnyethal

2. *kitöri a nyakát* = (nagyra törő v. hatalmon levő személy) vállalkozására csúnyán ráfizet

1. *komoly szándékai vannak* = komolyan gondol a házasságkötésre
2. *komoly szándékai vannak* = nagyon szeretne vmít, erősen törekszik vmire

1. *kosarat ad* = <nő> a férfi házassági ajánlatát visszautasítja
2. *kosarat ad* = vmely ajánlatot visszautasít, elhárít magától

1. *Néha még a kapanyél is elsül.* = {veszélyre való figyelmeztetés kifejezése: a veszélytelennek látszó dolgok is veszélyessé válhatnak}

2. *Néha még a kapanyél is elsül.* = {bizonytalanság kifejezése: semmi sem lehetetlen}

1. *nem a mai világba való* <vki> v. <vmi> = élehetetlen, gyámoltalan, ügyetlen személy

2. *nem a mai világba való* <vki> v. <vmi> = túlhaladott, korszerűtlen dolog

1. *nem fér a fejébe* <vmi> = nem tud megérteni, fölfogni, megtanulni <vmi>

2. *nem fér a fejébe* <vmi> = nem tud elfogadni, elképzelni <vmi>

3 A kulcsfogalmak rendszere és a kulcsfogalmak szinonimáinak illesztése a definíciókra

A frazéma-adatbázisban található definíciók szemantikai elemzése után 850 kulcsfogalomsort határoztunk meg. A kulcsfogalomsorok szinonim szavakból épülnek föl. Példa kulcsfogalomsorokra (feltüntetve a sorszámukat is):

456: HARAGSZIK, MEGHARAGSZIK, MEGSÉRTŐDIK, NEHEZTEL, MEGNEHEZTEL, DUZZOG

518: MENTEGETŐZIK, MAGYARÁZKODIK

774: RAVASZ, AGYAFÚRT, FORTÉLYOS, CSELES, HUNCUT

812: SZÍVESEN, ÖRÖMMEL, KÉSZSÉGESEN

Minta a magyar szinonimikus frazemaszótárból

Az automatikus illesztés után manuális szerkesztés következett, melyek során az illesztés hibáit javítottuk. Ugyancsak kézzel dolgoztuk fel és bontottuk szét igei aspektus szerint a frazémákat. Egy-egy kulcsfogalomsor összetartozó, aspektusbeli aleleteit megkülönböztetjük. Ezeket az alábbi mintában csillaggal (*) jelöljük.

480 IHLET*a múzsa csókja**isten sugallat*

* IHLETET KAP

*homlokon csókol a múzsa <vkit> (tréf)**megszáll a szentlélek <vkit> (tréf)***593** MUNKANÉLKÜLI, ÁLLÁSTALAN*állás nélküli**maga ura**senkinek sem parancsol**munka nélküli**pénztáros az ingyen uszodában (szleng)*

* MUNKANÉLKÜLIVÉ VÁLIK, ELBOCSÁTJÁK, KIRÚGJÁK, FELMONDANAK <VKINEK>

*kirepül az állásából (biz)**lapátra kerül (biz)**repül az állásából (biz)***801** SÁPADT, FALFEHÉR*annyi vére sincs, amennyivel egy szúnyog jóllakhatna**fehér, mint a meszelt fal**fehér, mint a kísértet (nép)**fehér, mint a lárva (nép)**minden csepp vér kiszaladt az arcából**nincs rajta emberi szín**olyan a képe, mintha megette volna a szappant**olyan a színe, mint a fagyos ing (nép)**olyan fehér, mint akit a meszesgödörből húztak ki**olyan, mint a napon sült málé (rég)**pápista a színe**sápadt, mint a holdvilág**sápadt, mint a hulla**szép, mint a tejbetők*

* ELSÁPAD, ELFEHÉREDIK

*kifut az arcából a vér**kiszalad az arcából a vér***821** SZABAD, ÖNÁLLÓ, FÜGGETLEN*a maga embere**a maga ura**kirepült már a fészekből (biz)**maga gazdája**maga kenyerén van (nép)**megáll a maga lábán**nem szorul senkire**senki nem parancsol neki*

szabad, mint a madár

* ÖNÁLLÓSODIK

szárnyra kel (vál)

* FELSZABADUL, MEGSZABADUL

széttöri az ígát

letépi a láncait

széztúzza a láncokat

Meggyőződésünk, hogy a munka során létrehozott frazéma-adatbázis és a belőle generált frazéma-szinonimaszótár a magyar frazeológiai kincs mélyebb összefüggéseinek feltárása mellett magyar nyelvű szövegek automatikus feldolgozásához, szemantikai elemzéséhez is hatékonyan használható fel.

Bibliográfia

1. Bárdosi, V.: Problèmes posés par le traitement lexicographique des figés dans les dictionnaires français. In: Fremdsprachen Lehren und Lernen, Jg. 21 (1992) 104–116
2. Bárdosi V. (főszerk.): Magyar szólástár. Szólások, helyzetmondatok, közmondások értelmező és fogalomköri szótára. Tinta Könyvkiadó, Budapest (2003)
3. Bárdosi V., Kiss G., Kiss M., Rapcsák T.: Kísérlet magyar szavak jelentéshasonlóságának meghatározására a Magyar szókincstár segítségével. In: II. Magyar Számítógépes Nyelvészeti Konferencia. Szegedi Tudományegyetem, Informatikai Tanszékcsoport, Juhász Nyomda (2004 27–37)
4. Bárdosi V.: Magyar szólások, közmondások értelmező és fogalomköri szótára. TINTA Könyvkiadó, Budapest (2009)
5. Forgács T.: Magyar szólások és közmondások szótára. Mai nyelvünk állandósult szókapcsolatai példákkal szemléltetve. TINTA Könyvkiadó, Budapest (2003)
6. Kiss G.: Magyar szókincstár. Rokon értelmű szavak, szólások és ellentétek szótára. Tinta Könyvkiadó, Budapest (1998)
7. Országh L. (főszerk.): A magyar nyelv értelmező szótára (I–VII.). Akadémiai Kiadó, Budapest (1959–1961)
8. Szemerényi Á.: Szólások és közmondások. Osiris Kiadó, Budapest (2009)
9. T. Litovkina A.: Magyar közmondástár. Közmondások értelmező szótára példákkal szemléltetve. TINTA Könyvkiadó, Budapest (2005)
10. Tolnai L.: A magyarság szellemi néprajza. Harmadik kötet. Királyi Magyar Nyomda, Budapest (1933–1938) 397–433
11. Voigt V. (szerk.): A magyar folklór. Osiris Kiadó, Budapest (1998)