

Doktoranduszok tanulmányai  
az alkalmazott nyelvészet köréből



2017

MTA Nyelvtudományi Intézet,  
Budapest

ALKALMAZOTT NYELVÉSZETI DOKTORANDUSZOK

# **Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2017**

## **XI. Alkalmazott Nyelvészeti Doktoranduszkonferencia**

Budapest, 2017. 02. 03.

Sorozatszerkesztő: Váradi Tamás

Szerkesztő: Ludányi Zsófia

Lektorálta:

Balaskó Mária

Fenyvesi Anna

Fóris Ágota

Furkó B. Péter

Gósy Mária

Gréczi-Zsoldos Enikő

Heltai János Imre

Hunyadi László

Károly Krisztina

Klaudy Kinga

Markó Alexandra

Miháltz Márton

Navracsics Judit

Prószéky Gábor

Sass Bálint

Sato Noriko

Schirm Anita

Váradi Tamás

A szerkesztő munkatársa: Tóth Bianka

ISBN 978-963-9074-70-5

Kiadja: MTA Nyelvtudományi Intézet

Budapest, 2017.

## Tartalomjegyzék

|  |     |
|--|-----|
| <i>Berente Anikó</i>   |     |
| A szegeedi boszorkánypercek négy írónak nyelvadatai .....  | 4   |
| <i>Dodé Réka</i>   |     |
| Doménspecifikus korpusz építése és validálása .....  | 17  |
| <i>Dömötör Andrea</i>  |     |
| Hány VAN nincs? A létige – zéró váltakozás korpuszvezérelt vizsgálata .....  | 28  |
| <i>Drávucz Fanni, Szabó Martina Katalin</i>  |     |
| A beszélői szubjektivitás vizsgálata szentiment- és emóciókorpuszokon .....  | 39  |
| <i>Jing Feng</i>   |     |
| The Generic Structure of Chinese MA Students' Thesis Acknowledgments:<br>A Comparative Study across Contexts ..... | 50  |
| <i>Gaál Péter</i>  |     |
| Online-szótár-használati szokások vizsgálata Magyarországon.....   | 65  |
| <i>Huszár Attiláné Fábriánkovits Erzsébet</i>  |     |
| Nyelvmegtartás és identitásépítés egy nemzetiségi horvát településen.....  | 76  |
| <i>Kaba Ariel</i>  |     |
| Pokemon-go ポケモン語 – avagy hogyan kommunikálnak a Pokemonok? .....   | 87  |
| <i>Kalivoda Ágnes</i>  |     |
| Az igekötők gépi annotálásának problémái.....  | 100 |
| <i>Kondacs Flóra</i>   |     |
| Az óvodások megnyilatkozásairól az <i>izé</i> kapcsán.....   | 109 |
| <i>Kovács Viktória</i>   |     |
| Koreferenciaviszonyok vizsgálata enyhe kognitív zavarban szenvedők<br>beszédátírataiban .....                      | 120 |
| <i>Kresztyankó Annamária</i>   |     |
| Cigány identitáskonstrukciók heterogenitása egyetemisták nyelvhasználatában .                                      | 131 |
| <i>Simkó Katalin Ilona</i>   |     |
| Fordíthatóság az angol és magyar idiómák tükrében Egy angol–magyar idióma-<br>adatbázis létrehozásáról .....       | 140 |
| <i>Sitkei Dóra</i>   |     |
| Magyarul tanuló lengyel hallgatók köztesnyelvről különböző beszédaktusokban<br>.....                               | 148 |

*Szabó Martina Katalin, Lázár Bernadett, Nyíri Zsófi, Morvay Gergely*

A negatív emotív elemek vizsgálata a nemek közötti nyelvhasználati különbségek szempontjából..... 161

*Szabó Martina Katalin, Nyíri Zsófi, Lázár Bernadett*

Negatív emotív szemantikai tartalmú fokozó elemek vizsgálata az angol–orosz és orosz–angol fordítás tükrében..... 177

*Vadász Noémi*

Anaforafeloldás menet közben – névmások egy pszicholingvisztikailag motivált elemzőben..... 192

## A szegedi boszorkányperek négy írnokának nyelvi adatai

Berente Anikó

SZTE BTK Nyelvtudományi Doktori Iskola  
anikbrnt2@gmail.com

**Kivonat:** A dolgozat a történeti szociolingvisztika elméleti keretében végzett vizsgálatot mutatja be. Az elemzés korpusza a XVIII. században, Szegeden keletkezett iratokból áll, melyeket négy jegyző készített (tehát nem a teljes iratanyag vizsgálatáról van szó). A különböző nyelvjárási jelenségek segítségével azt kívánom tisztázni, hogy kimutatható-e a dokumentumok lejegyzőinek anyanyelvjárása. Számtalan kérdés, módszertani probléma merülhet fel egy ilyen jellegű elemzés elvégzésekor, így ezekre is utalni szeretnék: el lehet-e különíteni a jegyzők és a tanúk, vádlottak anyanyelvjárását; mely jegyek azok, melyek a deáki nyelvváltozat részét képezhették; hogyan magyarázható olyan jelenségek felbukkanása, melyek előzetes várakozásaink ellenére mégis felfedezhetők az iratokban?

### 1 Bevezetés

Dolgozatomban a szegedi boszorkányperek négy jegyzőjének anyanyelvjárását kívánom feltárni. A szegedi dokumentumok 1726 és 1744 között keletkeztek, a kézírások sokszínűsége és az időtávlatok miatt több jegyzőt feltételezek. Az iratokra általában jellemző, hogy az írnok kielégítően utaló nyomokat nem tartalmaznak, de szerencsére ez alól van kivétel: négy lejegyzésben 3–3 aláírás, név szerepel. Az 1734-ben keletkezett CCVII. 67-es irat végén a következő nevek és titulusok találhatók: *Andreas Siskovitty m[anu] p[ropria] Senator, Ignatius Rakity m[anu] p[ropr]ia Senator* és *Joannes Széplaky m[anu] p[ropria] p[rae]libatae ci[vi]t[is] com[m]issar[ius]*. Az 1737-ben íródott CCVII. 72-es irat végén a következők olvashatók: *Joannes Széplaky m[anu] p[ropria] Senator, Josephus Pállfy m[anu] p[ropria] Senator* és *Joannes Barcza m[anu] p[ropria] P[rae]batae Civitatis Iur[at]us et Notarius*. A szintén 1737-es évszámhoz köthető CCVII. 76-os dokumentum végén lévő nevek a következők: *Franciscus Babarczy m[anu] p[rop]ria, Josephus Pállfi Senator, Joannes Kerny m[anu] p[rop]ria*. Az 1744-ben lejegyzett CCVII. 80-as irat aláírásai: *Franc[iscus] Babarczy, Josephus Pállffy, Michi[ael] Muhoray m[anu] p[ropria]*. Az ellenjegyzők a szenátorok voltak, így valószínűleg Széplaky János, Kerny János és Muhoray Mihály lehettek az iratok lejegyzői Barcza János mellett.

Az íráskép és a nyelvi jegyek alapján a következő iratokat társítottam Barczához: CCVII. 44; CCVII. 70. és másolata; CCVII. 72; CCVII. 74; CCVII. 75; CCVII. 76. egy része; CCVII.77. első oldala, CCVII. 79; CCVII. 80. egy része; Farkas Istvánné Nedelko Margit elleni tanúvallomások. A CCVII. 76. dokumentum 6. és 8. oldalán lévő rövid részletek, valamint a 80-as mappában tévesen helyet kapó néhány sort tartalmazó papírdarab is Barczához köthető. Az ő esetében egy korábbi vizsgálatomban (Berente

2017) a nyelvi adatok segítségével kimutattam, hogy minden valószínűség szerint a nyugat-dunántúli, esetleg a közép-dunántúli–kisalföldi nyelvjárási régió nyugati részéből származhatott. Ezekre az adatokra táblázatos formában a későbbiekben még kitérek.

Széplaky Jánoshoz a CCVII. 65. és CCVII. 67-es dokumentumokat, Kerny Jánoshoz a CCVII. 76-os irat nem Barczához köthető részeit és a CCVII. 77. második és harmadik oldalát, Muhoray Mihályhoz pedig a CCVII. 80-as dokumentumot társítottam. Az iratokat tehát az íráskép alapján rendeltem a nevekhez, de ez sem volt minden esetben könnyű feladat, hiszen két különböző kéz írásmódja is lehet a megtévesztésig hasonló, illetve egy kéz is írhat másképpen iratról iratra.

Mivel korábban a tallózó gyűjtési módszerrel a másik három jegyző esetén is találtam olyan nyelvjárási jegyeket, melyek nem köthetők a dél-alföldi régióhoz, így az a hipotézisem, hogy ők sem Szegeden vagy környékén születtek és nevelkedtek. Jelen dolgozatban tehát arra keresem a választ, hogy kimutatható-e a másik három írnok esetében is nyelvi adataik alapján anyanyelvjárásuk, mint Barcza esetében.

## 2 Elméleti keret

Elméleti keretem a történeti szociolingvisztika, melynek megítélése nem egységes a magyar nyelvudományban. Vannak kutatók, akik úgy vélik, a történeti jelző eleve szükségtelen, hiszen a szociolingvisztika önmagában már a diakrón szemléletet vallja, míg mások úgy látják, hogy éppen ez a jelző az, ami megfelelően utal arra a metodikai különbségre, melyre szükség van a szinkrón és diakrón szociolingvisztikai vizsgálatokban (Grécsi-Zsoldos 2013: 173). A két módszer eltérésére a Nevalainen–Raumolin-Brunberg (1996: 18) szerzőpáros világított rá, ezt Németh (2008) ültette át magyar nyelvre.

A magyar nyelvre vonatkozó elméleti kérdéseket taglaló tanulmányok száma elmarad a történeti szociolingvisztika módszerével dolgozó gyakorlati vizsgálatok számától, jelen tanulmány is az utóbbi kategóriába tartozik. A kutatás módszertana többretű, hiszen a történeti dialektológiai és szociolingvisztikai vizsgálat a klasszikus nyelvtörténeti megközelítéssel és a paleográfiaival is kiegészül.

## 3 Munkamódszer

A korábbi tallózó gyűjtés helyett arra vállalkozom, hogy a másik három lejegyző nyelvjárási adatait is teljeskörűen bemutatom úgy, ahogy Barcza János esetében is tettem. A jelenségek kiválasztásában a MDial. *Magyar nyelvjárások területi egységei* c. fejezet (Juhász 2003: 262–324) volt a segítségemre. Barczánál a következő nyelvjárási jelenségeket vizsgáltam: illeszkedéses *ö*-zés, erős *i*-zés, erős *l*-ezés (amely a nyugat-dunántúli régióban nemcsak a *ly*, hanem az *l + j* hangkapcsolat helyén is jelentkezik), *á* utáni *o*-zás, a főnévi igenév *-nyi* és *-nya* formája, explozív *gy*-zés, *ty*-zés. Ezek a jegyek együttesen a nyugat-dunántúli nyelvjárási régió jellegzetességei, de az MNyA. egyes térképlapjainak (761. *tőle*, 478. *feleségem*, 550. *mosolyog*, 937. *választás*, 915. *fürödni*, 917. *inni*, 992. *borjú*, 439. *gyapjú*) tanúbizonysága szerint a közép-dunántúli–kisalföldi régió nyugati része is érintett valamelyest a jelenségeket illetően, így ez a terület sem zárható ki a jegyző anyanyelvjárásának tisztázásakor.

Ezekhez a jegyekhez szorosan kapcsolódva kiválasztottam az  $n \sim ny$  palatalizációt,  $ly$ -ezést (mint az  $l$ -ezés ellentéte), az  $o$ -zást, az  $a$ -zást, illetve az  $é$ -zést, melyekről a későbbiekben még részletesebben szólni fogok. A jelenségek kiválasztásában a *Magyar nyelvjárások területi egységei* c. fejezet (Juhász 2003: 262–324) volt segítségemre.

A különböző nyelvjárási jegyek összegyűjtése nem jelent különösebb problémát – hacsak a grafémák esetleges olvasati nehézségeit nem vesszük számításba –, lényegesebb kérdés azonban, hogy a talált adatainkat mihez képest értékeljük. Egyrészt adódik a lehetőség, hogy a mai sztenderd nyelvi formákkal vessük össze őket, másrészt az összehasonlítást elvégezhetjük egy másik nyelvjárási régió jellemző alakjaival. Mivel a kutatásom kezdetén az volt az egyik hipotézisem, hogy a szegedi keletkezésű iratokban nagy számban fognak előfordulni a dél-alföldi régióra, szűkebben a Szeged környéki nyelvjárásra jellemző elemek, így talán nem meglepő, ha a nyugat-dunántúli/közép-dunántúli–kisalföldi jegyeket a Szeged környéki jelenségekkel vetem össze. Az egyéb kiválasztott jelenségek, melyeket bevontam a vizsgálatba ( $n \sim ny$  hangkapcsolat,  $a$ -zás stb.), külön táblázatban közlöm, de itt már nem az egyes dunántúli jegyekből indulok ki, hanem magukból a nyelvjárási formákból, és a mai sztenderdnek megfelelő számadatokkal hasonlítom össze őket. Erre azért volt szükség, mert pl. az  $a$ -zás nem az eddig említett területeken jellemző, vagy az  $n \sim ny$  kapcsolata paradigmatiszós kiegészítődést mutat a nyelvterület több részén is. Ezek a jegyek tehát általánosságban véve az írrok nyelvjárásiasságát mutathatják meg.

### 3.1 Néhány kutatási probléma

A négy írrok közül csak Barcza neve kapcsán említi bizonyosnak a jegyzőséget a Szeged történetével foglalkozó irodalom (vö. Vass 1985), és láthattuk, hogy egy helyütt ő maga írja alá jegyzőként a dokumentumot. Széplaky, Kerny és Muhoray esetében azonban nem találunk olyan utalást, miszerint írkokok lettek volna a szegedi tanácsban, egyéb titulusok kapcsán viszont felmerül a nevük, így tudható, hogy nem töltöttek be magas szenátori tisztséget. Más lényeges független változó értéke, mint pl. a koruk, iskolázottságuk szintje, társadalmi rangjuk nem derül ki – annyi bizonyos, hogy magasabb műveltségi rétegbe tartozó, írástudó férfiak voltak.

A periratok gazdag nyelvi forrásnak tekinthetők, hiszen a műfaj nagyon közel áll az élőbeszédhez. Mivel sok tanút szólaltatnak meg, a nyelvi anyag is igen változatos lehet. A tanúvallomásokban feltűntetik a tanúk legfontosabb adatait, nevüket, korukat, olykor születési helyüket, jelenlegi, esetleg korábbi lakhelyüket. Bizonyos értelemben tehát sok lényeges információt megtudunk róluk, gyakran azonban csak egy nagyon rövid szövegrészlet kapcsolódik egy-egy személyhez, így még nehezebb helyzetbe kerülünk, ha el szeretnénk különíteni a tanúk és a lejegyzők nyelvjárását. Szerencsésebbek vagyunk azokban esetekben, amikor pl. a vádlott önvallomásai szerepelnek, így nagyobb eséllyel kerülhetett a lejegyzésbe a vádlott nyelvhasználatából is. Ugyanakkor kiforrott módszer hiányában szinte lehetetlen a tanúk, vádlottak és jegyzők nyelvhasználatát egymástól elkülöníteni.

Mindig figyelembe kell vennünk azt is, hogy a jegyző egy hivatalos pozíciót betöltő személy, akinek feladata, hogy a lehető leghívebben visszaadja az elhangzottakat, ugyanakkor stilizálja is azokat. Utóbbira utal pl. az *s. v. [salva venia]* jelölés, amely trágár szavak, megnyilatkozások előtt szerepel, illetve azok az esetek is, amikor a lejegyző E/3-ban adja vissza a tanú vallomását. Van azonban egy további tényező, amely befolyásolhatta a XVIII. századi lejegyzéseket: a sztenderdizálódási törekvések,

folyamatok. Véltetően feltételezhetünk egy deáki, írnoeki nyelvvaltozatot, melybe különböző elemek ajánlott lejegyzési sémája is beletartozhatott – ez pedig már a sztenderdizálódás folyamata felé mutat (ld. Németh 2004).

Láthatjuk tehát, hogy a boszorkánypercek történeti nyelvészeti vizsgálata milyen sokrétű problémakörrel rendelkezik. Kutatásom elején járva alapvetően még abból indulok ki, hogy a lejegyzésre leginkább az írnoek nyelvhasználata lehetett hatással, így arra egyelőre nem vállalkozom, hogy megpróbáljam egyértelműen elkülöníteni a tanúk, vádlottak és a lejegyzők anyanyelvjárását.

## 4 Eredmények

### 4.1 Barcza János

A Barcza János esetében gyűjtött adatok táblázatos összefoglalása az alábbiakban látható, a pontos nyelvi adatok közlésétől a nagyszámú előfordulás miatt e jegyző esetében viszont eltekintek. (A nyelvjárási adatok részletes bemutatása egy korábbi esettanulmányban kapott helyet, l. Berente 2017).

A nyugat-dunántúli, illetve a közép-dunántúli–kisalföldi régió egy részében is jellemző jelenségek Szeged környékén nem fordulnak elő (eltekintve néhány *i*-ző szótól), lényegében tehát a Szeged környéki területen jellemző adatok a mai sztenderdnek is megfelelnek. Megjegyzendő továbbá, hogy vannak olyan jelenségek, melyek más nyelvjárási területeken is felbukkannak, mint pl. a főnévi igenév *-nya* formája a palócoknál vagy az *i*-zés a Tiszántúlon, de az alább feltüntetett jelenségek együttesen a nyugat-dunántúli, esetleg a közép-dunántúli–kisalföldi régió nyugati felére mutatnak.

| Jelenség  | Nyugat-dunántúli/közép-dunántúli–kisalföldi régióra jellemző adatok száma | Szegedi környéki régióra jellemző adatok száma |
|---|---|--|
| Erős <i>i</i> -zés  | 98  | 37   |
| Erős <i>l</i> -ezés   | 67  | 49   |
| Illeszkedéses <i>ö</i> -zés                                 | 53  | 2  |
| Főnévi igenév képzőjének <i>-nyi</i> és <i>-nya</i> formája |   |  |
| <i>Á</i> utáni <i>o</i> -zás                                | 13  | 10   |
| Explozív <i>gy</i> -zés                                     | 3   | 3  |
| Explozív <i>ty</i> -zés                                     | 0   | 2  |

1. táblázat. A nyugat-/közép-dunántúli–kisalföldi és Szeged környéki nyelvjárásra jellemző adatok számszerű megoszlása Barczánál

Ahogy az 1. táblázat mutatja, Barczánál túlnyomó többségben jelenik meg az erős *i*-zés, *l*-ezés és illeszkedéses *ö*-zés, aminek alapján azt mondhatjuk, hogy a jegyző minden valószínűség szerint a nyugat-dunántúli, esetleg a közép-dunántúli–kisalföldi régió nyugati részéből érkezhetett Szegedre. Véleményem szerint a többi jelenség (a



főnévi igenév *-nyi* és *-nya* formája, *á* utáni *o*-zás stb.) e megállapítást erősíti még abban az esetben is, ha jóval kisebb számban fordulnak elő, mint pl. az *ö*-zés.

Az 1. táblázatban a nyugat-/közép-dunántúli–kisalföldi és a szegedi régióban jellemző adatokat hasonlítottam össze, ugyanakkor vannak olyan jelenségek, amelyek több régióban is előfordulhatnak. A köznyelvi *n* és nyelvjárási *ny* megfelelés az egész nyelvterületen jelentkezhet, de legerősebbnek a Dunántúl nyugati és középső részein tekinthető, míg a köznyelvi *ny* és nyelvjárási *n* megfelelés inkább a nyugati megyékben és Erdély keleti-északkeleti pontjain jelenik meg (vö. Imre 1971: 250–3). Kálmán (1966: 46) viszont kiemeli, hogy egyes nyelvjárásokban az *n*-es és *ny*-es paradigmák kiegyenlítődnek, így bizonyos hangtani helyzetben palatalizáció, ill. depalatalizáció valósul meg a köznyelvi formához képest. Végh (1971: 822) Tápé nyelvét vizsgálva arra a megállapításra jut, hogy az *n* ~ *ny* kiegyenlítődés a Szeged melletti település nyelvjárásának részét képezi, tehát hasonló jelenséget feltételezhetünk a szegedi nyelvjárásban is. A *ly*-ezés a palóc típusú nyelvjárásokban fordul elő leginkább, a *ly* fonéma helyén a Dunántúl nagy részén inkább az *l*-es megoldás a tipikus (Imre 1971: 243). Kálmán (1966: 42) még azt is megemlíti, hogy néhány keleti nyelvszigeten is jellemző a *ly* fonéma. Az *o* ellenében jelentkező *a*-zás Erdély középső területén tekinthető kiterjedtnek (Imre 1971: 228), tehát feltételezhetjük azt, hogy a jelenség nem fog a szegedi iratokban nagy számban előfordulni. Az *o*-zás viszonylag erős hatású a Dunántúl nyugati és Erdély keleti részén, független helyzetben azonban néhány morféma érintett a Duna-Tisza közén, Szeged környékén és a Tiszántúl déli részén is (Imre 1971: 193). Az *é*-zés a Dunántúl déli részén és Erdély területén jellemző, de kisebb mértékben a Duna-Tisza közén, Szeged környékén, sőt a középpalócság egyes részein is megjelenhet egy-egy morfémaiban (Imre 1971: 146). Ugyanakkor több kutató is úgy véli – pl. Papp 1961; Gréczi-Zsoldos 2007 –, hogy az *é*-zés az *í*-zést ellensúlyozhatta, így az írnokok nyelvhasználatában megtalálható *é*-zés a deáki nyelvváltozat részét képezte. Annak megállapítása, hogy az *é*-ző adatok közül melyek vezethetők vissza a lejegyző/tanú anyanyelvjárására, s melyek a deáki írásgyakorlatra, természetesen további vizsgálatot igényel egy nagyobb korpusz elemzésével.

Láthatjuk tehát, hogy vannak olyan jelenségek, amelyeket több nyelvjárási régióhoz is tudunk kötni, s több Szeged területén is megjelenik, így nem tartottam célszerűnek az előzőekben látott módon elvégezni az összehasonlítást (nyugat-dunántúli/közép-dunántúli–kisalföldi vs. dél-alföldi), hanem inkább a mai sztenderdnek megfelelő adatokat hasonlítottam össze általában véve a nyelvjárási formákkal, természetesen a pótkompetencia fogalmát szem előtt tartva (Forgács 1993–1994).

| Jelenség                                | Nyelvjárási adatok száma | Mai sztenderdnek megfelelő adatok száma |
|---|--------------------------|---|
| Depalatalizáció: $ny > n$               | 49                       | 57                                      |
| Palatalizáció: $n > ny$                 | 16                       | 33                                      |
| Palatalizáció: $l > ly$                 | 6                        | 9                                       |
| Magánhangzó nyíltabbá válása: $o > a$   | 42                       | 135                                     |
| Magánhangzó zártabbá válása: $a > o$    | 10                       | 2                                       |
| Magánhangzó nyíltabbá válása: $i/i > é$ | 30                       | 54                                      |

**2. táblázat.** A több nyelvjárási régióhoz is kapcsolható és a mai sztenderdnek megfelelő adatok számszerű megoszlása Barczánál

Ahogy a 2. táblázatból kiderül, a sztenderdnek megfelelő adatok jóval többször mutathatók ki írnökunknál, mint a nyelvjárási formák, ennek ellenére beszédesnek tartom utóbbiak számát. Egyrészt azt mutatja, hogy noha ekkoriban már megindulhattak bizonyos sztenderdizációs törekvések, melyeknek az a céljuk, hogy minél inkább a kialakuló norma felé közelítsenek a lejegyzések, mégis képesek beszüremkedni ezek a formák is a hivatalos regiszter irataiba. Másrészt ehhez szorosan kapcsolódva közvetve következtetések vonhatók le az írnök tudatosságáról is, hiszen a jegyző a vélhető normahatás ellenére mégsem tartotta ezeket üldözendőnek. Az *a*-zás kapcsán megjegyezném, hogy sem a nyugat-/közép-dunántúli–kisalföldi, sem a szegedi régióban nem jellemző, mégis különös, hogy ekkora számban előfordul ez a jegy is. Erre többféle magyarázatot is adhatunk, akár kombinálva is ezeket. Egyrészt az *a*-zás része lehetett a tanúk nyelvhasználatának is – akik pl. más településekről érkezettek Szegedre –, így meg is őrződhetett a tanúvallomások lejegyzésekor. Másrészt a XVIII. században a jelenség még terjedőben lehetett, így noha később nem honosodott meg a vizsgált nyelvjárási területeken, a korábbi időszakban élhetett az oralitásban ezekben a régiókban is – gondoljunk csak a török utáni betelepítésekre és vándorlásokra! Harmadrészt a jegyző anyanyelvjárás-keveredése is magyarázhatja a jelenséget, hiszen elképzelhető, hogy olyan területen élt, ahol az *a*-zás jellemző volt, így az ő nyelvhasználatának is részét képezhette. Ugyanakkor meg kell említenem még egy fontos mozzanatot: Barczánál sok esetben nem lehetett egyértelműen eldönteni, hogy az adott grafémát *a*-ként vagy *o*-ként kell-e olvasni, így érdemes az *a*-zó és *o*-zó számadatokat némi fenntartással kezelnünk.

#### 4.2 Széplaky János

Széplaky lejegyzéseiben a nyugat-dunántúli, illetve a közép-dunántúli–kisalföldi nyelvjárási régió nyugati felében jellemző jegyek száma Barczához képest kisszámúnak mondható, hiszen Széplakynál egyik jelenség sem tekinthető markánsnak.

Az *i*-zés a következőkben mutatható ki: *beszid*, *beszilllette*, *beszillettenek*, *égirtek*, *egisségében*, *egisségiben*, *egissigében*, *felesigié*, *felesigihez*, *fenyegéttel*, *Károlyi Pila*, *Károlyi Pilához*, *kereskedisben*, *kérisére*, *megh beszilllette*, *megh terisében*, *megh terisében*, *megh térisiben*, *szintin*, *tiged*, *veszekedisét*. Megjegyzendő, hogy az *égirtek* adat esetében felmerül, hogy betűcseréről van szó, valamint a *Károlyi Pila* név kapcsán

az elírás lehetősége, hiszen két *i*-ző forma ellenében 28 *é*-ző adat szerepel. Továbbá egy kérdéses *i*-ző adatom is van: *ennjhanj*. Mivel ez az írnok a *j*-t az *i* és az *y* jelölésére is használja, így az adat olvasata problematikus lehet – itt vélhetően mégis az *i*-zés egyik adatával állunk szemben.

Az *l*-ezés szintén kevés adatban található meg ennél az írnoknál: *mihelest* (négyszer), *mullék*, *nyavalájában*, *paszullall*, *paszult*, *valamellike*. Az illeszkedései *ö*-zés 15 helyen fordulhatott volna elő, ez azonban mégsem realizálódott a lejegyzésekben, de a főnévi igenév is csak egyszer szerepelt *-nya* formában: *megh innya*. Az *á* utáni *o*-zás hét esetben mutatható ki: *bocsánott* (kétszer), *házos társul*, *megh tanítanások*, *ne bántona* 'ne bántaná', *társosságban* (kétszer).

| Jelenség  | Nyugat-dunántúli/közép-dunántúli–kisalföldi régióra jellemző adatok száma | Szeged környéki régióra jellemző adatok száma |
|---|---|---|
| Erős <i>i</i> -zés  | 22  | 42  |
| Erős <i>l</i> -ezés   | 9   | 28  |
| Illeszkedései <i>ö</i> -zés                                 | 0   | 15  |
| Főnévi igenév képzőjének <i>-nyi</i> és <i>-nya</i> formája | 1   | 24  |
| <i>Á</i> utáni <i>o</i> -zás                                | 7   | 10  |
| Explozív <i>gy</i> -zés                                     | 0   | 1   |
| Explozív <i>ty</i> -zés                                     | 0   | 1   |

**3. táblázat.** A nyugat-/közép-dunántúli–kisalföldi és Szeged környéki nyelvjárásra jellemző adatok számszerű megoszlása Széplakynál

A nyugat-dunántúli, illetve a közép-dunántúli–kisalföldi régió nyugati részén jellemző nyelvjárási jegyek alapján nem lehet megalapozottan kijelenteni, hogy Széplaky is e két terület valamelyikéről származhatott, bár a lehetősége kétségteljesen felmerül.

Lássuk most a további jelenségeket is! Úgy tűnik, az *ny* depalatalizációja viszonylag nagy számban fordul elő, de megjegyzendő, hogy 15 adat csak az *asszony* szóhoz kapcsolódik: *Aszonak*, *aszonnak*, *aszontul*, *boszorkánságot*, *ennjhanj*, *gazda aszont*, *Halasy Aszont*, *hánta* (négyszer), *kapitánné*, *karácsonj diót*, *koldus aszonnak*, *koldus aszont*, *leánkáknak*, *legénnek*, *Ober kapitánn[e]*, *Ober kapitánné*, *ven aszon*, *vén aszon*, *vén aszonnak*, *vén aszont* (ötször). A *karácsonj* és az *ennjhanj* bizonytalan olvasatú a jelölés módja miatt, ennek ellenére úgy gondolom, valószínűleg az *n*-ezés eseteiről lehet szó. Egyetlen név kapcsán merült fel az *ny*-ezés: *Paplanyos Panka*, de az *l* palatalizált változata is csak három példában mutatható ki: *Károlyi Pélya*, *kár vallyását*, *pallyásán*. A *Károlyi Pélya* adat esetében gyanakodhatunk elírásra is, mert egyetlen példában szerepel a *Péla* palatalizált formában, 27-szer *l*-es alakban olvasható a név. *A*-zás csak a *mast* esetében, *o*-zás pedig a *szovára* szóban mutatható ki. Az *é*-zés szintén kevesebb nyelvjárási formában adatolható: *bérnék*, *égirtek*, *emlétt* (háromszor), *kézzott*, *kévánt*, *kévül*. Ahogy már utaltam rá, az *égirtek* lehet elírás eredménye is. Megjegyzendő továbbá, hogy nemcsak *i*-vel, hanem *i*-vel váltakozó *é*-zés is tetten érhető írnokunknál: *haragszék*, *ki ésmérte*, *ki latszék*, *le kölleté*, *másodék*, *megh szűnék*, *ménémő*, *szerént*, *urajémék*, *visza tekéntvén*. Az igeik esetén azonban nem teljesen egyértelmű, hogy *i* ~ *é* váltakozásról, vagy az elbeszélő múlt kifejezéséről van-

e szó, hiszen a szövegkörnyezet alapján utóbbi is elképzelhető. A *ménémő* kapcsán szintén felmerül a tévesztés lehetősége.

| Jelenség                                | Nyelvjárási adatok száma | Mai sztenderdnek megfelelő adatok száma |
|---|--------------------------|---|
| Depalatalizáció: $ny > n$               | 26                       | 35                                      |
| Palatalizáció: $n > ny$                 | 1                        | 4                                       |
| Palatalizáció: $l > ly$                 | 3                        | 31                                      |
| Magánhangzó nyíltabbá válása: $o > a$   | 1                        | 5                                       |
| Magánhangzó zártabbá válása: $a > o$    | 1                        | 2                                       |
| Magánhangzó nyíltabbá válása: $i/i > é$ | 18                       | 30                                      |

**4. táblázat.** A több nyelvjárási régióhoz is kapcsolható és a mai sztenderdnek megfelelő adatok számszerű megoszlása Széplakynál

### 4.3 Kerny János

Kerny esetében szintén nem beszélhetünk a nyugat-dunántúli/közép-dunántúli–kisalföldi régióban jellemző jegyek nagyszámú előfordulásáról. Az *i*-zés csak a következő adatban mutatható ki: *enyim*, amely azonban Szeged környékén is használatos. Az *l*-ezést ezekben a szavakban érhetjük tetten: *csináltak* 'csináljatok', *felleb*, *gombolaghba*, *ne busul* 'búsulj', *negállá*, *sörtélesseinek*, *utolán*, *valla*. A főnévi igenév egy szóban mutat *-nya* formát: *innya* (kétszer). Az *á* utáni *o*-zás sem tekinthető Kernynél kiugró számúnak: *ássotok*, *barázdánok*, *gyermek ágyos*, *leánkájának*, *pallásro*, *szerszámo*, *tarsolkodni*. Az explozív *ty*-zés egyetlen szóhoz kapcsolódóan mutatható ki: *gyaptyuban* (kétszer).

| Jelenség  | Nyugat-dunántúli/közép-dunántúli–kisalföldi régióra jellemző adatok száma | Szeged környéki régióra jellemző adatok száma |
|---|---|---|
| Erős <i>i</i> -zés  | 1   | 0   |
| Erős <i>l</i> -ezés   | 8   | 1   |
| Illeszkedéssel <i>ö</i> -zés                                  | 0   | 10  |
| Főnévi igenév képzőjének - <i>nyi</i> és - <i>nya</i> formája | 2   | 20  |
| <i>A</i> utáni <i>o</i> -zás                                  | 7   | 29  |
| Explozív <i>gy</i> -zés                                       | 0   | 3   |
| Explozív <i>ty</i> -zés                                       | 2   | 1   |

**5. táblázat.** A nyugat-/közép-dunántúli–kisalföldi és Szeged környéki nyelvjárásra jellemző adatok számszerű megoszlása Kernynél

Kerny esetében az *n*-ezés 15 esetben mutatható ki, igaz, 10 példa a *leány* depalatalizálódott formájához kapcsolódik: *el sinlődet volna, kéntelenétegett, kéntelenétegtünk, ki hánná, leánka* (háromszor), *leánkája* (háromszor), *leánkájának, leánkáját* (kétszer), *leánkával, termékenségnek*. Az *n* palatalizációja csupán egy név kapcsán merült fel: *Simony Mihály*, míg az *l* palatalizált változata háromszor fordult elő a *mivelly* szót érintően. Az *a*-zás Kernynél sem mutatható ki sok esetben: *étsakákan, ihassan, mast* (hátszor), *megh gyogulan* 'meggyógyuljon', *petákatt*; az *a* zártabbá válása hatszor látható: *gyogyétona, gyogyétona megh, gyonakodianak, Morkában* 'markában', *morkokban, ujobban*. A legtöbb nyelvjárási forma az *é*-zés kapcsán merül föl Kerny esetében, bár megjegyzendő, hogy a 18 példából 17 az *-ít* képzőhöz kapcsolódik: *be szorétotta, emletet, emlétett, gyogyetana megh, gyogyétok, gyogyétona* (kétszer), *kéntelenétegett, kéntelenétegtünk, megh fogom gyogyétani, megh gyogyétotta, megh gyogyétotta volna, megh gyogyétottam, megh nehezétette, megh nem szorétotta, méghlen, szorétván; i*-vel váltakozó *é*-zés csak két esetben fordult elő: *cseleksétek, szerént*.

| Jelenség  | Nyelvjárási adatok száma | Mai sztenderdnek megfelelő adatok száma |
|---|--------------------------|---|
| Depalatalizáció: <i>ny</i> > <i>n</i>               | 15                       | 10                                      |
| Palatalizáció: <i>n</i> > <i>ny</i>                 | 1                        | 1                                       |
| Palatalizáció: <i>l</i> > <i>ly</i>                 | 3                        | 0                                       |
| Magánhangzó nyíltabbá válása: <i>o</i> > <i>a</i>   | 10                       | 42                                      |
| Magánhangzó zártabbá válása: <i>a</i> > <i>o</i>    | 6                        | 5                                       |
| Magánhangzó nyíltabbá válása: <i>i/i</i> > <i>é</i> | 20                       | 6                                       |

**6. táblázat.** A több nyelvjárási régióhoz is kapcsolható és a mai sztenderdnek megfelelő adatok számszerű megoszlása Kernynél

#### 4.4 Muhoray Mihály

A Muhoray lejegyzéseiben talált nyugat-dunántúli/közép-dunántúli–kisalföldi jelenségek számadatai szintén olyan megoszlást mutatnak, mint ahogy Széplakynál és Kernynél láthattuk. Az erős *i*-zés egy példában látható csupán: *igiretire*. Az *l*-ezésre 5 adatot találtam: *hüvelk*, *mihelest* (háromszor), *tavali*. Az irat végén szerepel ugyan még két *l*-ező adat (*Tőke Golója*), de ezt minden valószínűség szerint másik kéz írhatta, mert az előző kéz betűtípusától egyértelműen eltérő a lejegyzett 16 sornyi szöveg – így ezeket nem is vettem számításba az összesítésnél. A főnévi igenév *-nya* formája csak egy szóban mutatható ki Muhoraynál: *innya*.

| Jelenség  | Nyugat-dunántúli/közép-dunántúli–kisalföldi régióra jellemző adatok száma | Szeged környéki régióra jellemző adatok száma |
|---|---|---|
| Erős <i>i</i> -zés  | 1   | 0   |
| Erős <i>l</i> -ezés   | 7   | 1   |
| Illeszkedéses <i>ö</i> -zés                                 | 0   | 13  |
| Főnévi igenév képzőjének <i>-nyi</i> és <i>-nya</i> formája | 1   | 15  |
| <i>Á</i> utáni <i>o</i> -zás                                | 0   | 0   |
| Explozív <i>gy</i> -zés                                     | 0   | 6   |
| Explozív <i>ty</i> -zés                                     | 0   | 5   |

7. táblázat. A nyugat-/közép-dunántúli–kisalföldi és Szeged környéki nyelvjárásra jellemző adatok számszerű megoszlása Muhoray lejegyzéseiben

Muhoray esetében csak az *o*-zás és az *é*-zés kapcsán mondható el, hogy többször előfordul, mint a sztenderdnek megfelelő adatok száma. Az *ny* depalatalizált formában csak egyszer mutatható ki: *reménkedet*; míg az *n* palatalizált változata hat példában látszik: *fa kalánnyal*, *Istvány* (háromszor), *kenyegette*, *Mártony*. Az *l* palatalizációja szintén egyszer fordult elő: *mellyire*. Az *a* ellenében jelentkező *o*-zás négyszer tűnik fel: *jovaslábúl* 'javaslásából' [sic!], *ujobban* (háromszor). Ahogy említettem, az *é*-zés nagyobb számban fordul elő Muhoraynál, de nála is túlnyomó többségben (36-ból 34-szer) az *-ít* képző volt érintett: *állétot vala*, *bizonyettya*, *csillapétássára*, *el igazétotta*, *gyógyécsa megh*, *gyógyetani*, *gyógyétani*, *gyógyétasa*, *gyógyétásában*, *gyógyétásáért* (kétszer), *gyógyétására*, *gyógyetésban*, *gyógyétotta*, *gyogyetotta vala*, *gyógyétotta vala*, *gyógyetottak*, *gyógyetották vólna*, *igazétotta*, *kijis szabadétotta*, *megh gyógyetanám*, *megh gyógyétanám*, *megh gyógyéthatlak*, *megh gyógyéthattya*, *megh gyógyétotta* (kétszer), *megh gyógyetottam vólna*, *megh nem gyógyéthatom*, *megh nem gyógyéthattya*, *megh nem gyógyétom*, *megh nem gyógyettya* (háromszor), *megis gyógyétottalak vólna*, *méglen* (kétszer). Muhoray *i*-vel váltakozó *é*-ző adatai a következők: *megént*, *megtekentvén*, *szerént* (kétszer). A korábban már említett másik kéz az irat végén ezeket az *i*-ző adatokat szolgáltatta: *héta*, *sérna*, illetve van nála egy *é*-vel váltakozó *i*-zés is: *édejn* 'idején'. Utóbbi adatokat a táblázatos összesítésnél szintén nem vettem figyelembe.

| Jelenség                                | Nyelvjárási adatok száma | Mai sztenderdnek megfelelő adatok száma |
|---|--------------------------|---|
| Depalatalizáció: $ny > n$               | 1                        | 25                                      |
| Palatalizáció: $n > ny$                 | 6                        | 7                                       |
| Palatalizáció: $l > ly$                 | 1                        | 1                                       |
| Magánhangzó nyíltabbá válása: $o > a$   | 0                        | 27                                      |
| Magánhangzó zártabbá válása: $a > o$    | 4                        | 0                                       |
| Magánhangzó nyíltabbá válása: $i/i > é$ | 40                       | 16                                      |

**8. táblázat.** A több nyelvjárási régióhoz is kapcsolható és a mai sztenderdnek megfelelő adatok számszerű megoszlása Muhoraynál

## 5 Összefoglalás

Dolgozatom hipotézise az volt, hogy a nyelvjárási jegyek segítségével meg lehet határozni a szegedi boszorkányperek négy írójának anyanyelvjárását. Barcza János sok olyan adatot szolgáltatott, melyek alapján valószínűsíthető, hogy a nyugat-dunántúli, esetleg a közép-dunántúli–kisalföldi régió nyugati részéből érkezhetett Szegedre. Ugyanez a másik három lejegyző esetében már problematikusnak tekinthető: a tallózó gyűjtés alapján tett előzetes hipotézisem, miszerint Barczán kívül más jegyzők is a nyugat-dunántúli, esetleg más dunántúli régióból származhattak Szegedre, nem igazolható megnyugtatóan. Az egyéb nyelvjárási területekre jellemző jelenségek sem fordultak elő akkora számban, hogy biztosan igazolni lehessen a lejegyzők anyanyelvjárását. Ugyanakkor kiemelném, hogy Széplaky és Kerny esetében felmerül a lehetősége a dunántúli származásnak, hiszen – még ha rendkívül csekély számban is mutatható ki –, Széplakynál az *i*-zés és *l*-ezés, Kernynél az *ty*-zés és az *o*-zás erre mutathat. Természetesen az adatok a tanúk nyelvhasználata révén is megőrződhetnek, vagyis nem feltétlenül a lejegyzők anyanyelvjárásainak részeként kell őket értékelnünk, így ez a következtetés csak óvatosan kezelendő. Muhoray esetében az adatok alapján kimondható, hogy minden valószínűség szerint nem dunántúli származású volt, sőt, a rendkívül kevés nyelvjárási adat miatt őt tekinthetjük a legtudatosabb lejegyzőnek.

Noha az *ny* depalatalizált formája viszonylag többször előfordult a jegyzők esetében, olyan szavak voltak legtöbbször érintettek, mint az *asszony*, *leányka*, melyek több nyelvjárási régióban is depalatalizált alakban realizálódnak. Az *é*-zés kapcsán úgy vélem, az a megállapítás megállja a helyét, miszerint egyes elemek a deáki nyelvváltozat részét képezhették (*-ít* képző *-ét* formája vagy a *szerént* stb.). A többi esetben viszont nem lehet egyértelműen eldönteni, hogy az élőbeszédben meglévő egykori nyelvjárási használatra utalnak-e az *é*-ző szavak, vagy inkább arról van szó, hogy az *é*-zés az *i*-zés ellenében egyre nagyobb teret kapott, s az írások egyre több elemre analógiásan is kiterjesztették. A többi jelenség esetében (*o*-zás, *a*-zás stb.) sem tisztázható egyértelműen az iratok adottságai miatt (ezekre a 3.1-es alfejezetben utaltam), hogy a nyelvi jegyek a jegyzők vagy a tanúk nyelvhasználatának voltak-e részei. Az *ly*-ezés jelensége nem igazán illik a vizsgált jelenségek sorába, mégis mindegyik jegyzőnél feltűnt néhány adat. E. Abaffy (1965: 190) szerint a *holly*,

*gyámolya* stb. adatok „hiperurbanisztikus jellegűek, tehát a beszélt nyelvben éppen az ellenkező folyamatra, az erős *l*-ezésre mutatnak”. Így ezek az *ly*-ező adatok a nyugat-dunántúli *l*-ező jelenséget támaszthatják alá, de mivel mind az *l*-ező, mind az *ly*-ező adatok rendkívül csekély számban fordulnak elő, így meggyőzően nem érvelhetünk ezzel sem.

Összességében az is elmondható, hogy a különböző nyelvjárási jelenségek viszonylag kis számban fordultak elő Széplakynál, Kernynél és Muhoraynál, ami talán annak a következménye, hogy ezek a jegyzők jóval tudatosabbak voltak Barczánál, jobban kerültek azokat a jelenségeket, melyek nem illettek a kialakuló normába (*i*-zés, *l*-ezés stb.) és többször alkalmaztak olyan elemeket, melyek viszont kimondottan a regiszterbe tartozhattak (*é*-zés). Ugyanakkor ne feledjük: lehetséges, hogy amiatt érhető tetten jóval kevesebb nyelvjárási forma a másik három írónk esetében, mert az iratok száma jegyzőnként nem tekinthető nagyszámúnak. További vizsgálatot igényel még az is, hogy tisztázzuk a CCVII. 80-as irat végén található 16 sornyi szöveg másik lejegyző kezének anyanyelvjárását, ill. az esetlegesen hozzá köthető további iratok felkutatását.

## Irodalom

- Berente, A. 2017. Nyelvjárási jelenségek vizsgálata a szegedi boszorkányperek egyik írónakán irataiban. Esettanulmány. In: Forgács, T., Németh M., Sinkovics B. (szerk.) *A nyelvtörténeti kutatások újabb eredményei IX.* Szeged: Szegedi Tudományegyetem, Magyar Nyelvészeti Tanszék. 7–22.
- Brandl, G., Tóth G., P. 2016. *Szegedi boszorkányperek 1726–1744.* Budapest: Balassi.
- E. Abaffy, E. 1965. *Sopron megye nyelve a XVI. században.* Budapest: Akadémiai.
- Forgács, T. 1993–1994. Zárt korpuszok és pótkompetencia. *Néprajz és Nyelvtudomány*, 35: 17–23.
- Gréczi-Zsoldos, E. 2007. *Nógrád vármegye nyelve a XVII. században.* Salgótarján: Nógrád Megyei Levéltár.
- Gréczi-Zsoldos, E. 2013. A történeti szociolingvisztikáról. In: Kontra, M., Németh M., Sinkovics B. (szerk.) *Elmélet és empiria a szociolingvisztikában.* Budapest: Gondolat. 173–187.
- Imre, S. 1971. *A mai magyar nyelvjárások rendszere.* Budapest: Akadémiai.
- Kálmán, B. 1966. *Nyelvjárásaink.* Budapest: Nemzeti Tankönyvkiadó.
- Juhász, D. 2003: A magyar nyelvjárások területi egységei. In: Kiss, J. (szerk.) *Magyar dialektológia.* Budapest: Osiris. 262–324.
- MNyA. = Deme, L., Imre, S. (szerk.) 1968–1977. *A magyar nyelvjárások atlasza I–VI.* Budapest: Akadémiai.
- Németh, M. 2004. *Nyelvjárás, beszélt nyelv és spontán sztenderdizációs törekvések a XVIII. századi szegedi írónki nyelvváltozatban.* Budapest: Akadémiai.
- Németh, M. 2008. *Nyelvi változás és váltakozás társadalmi és műveltségi tényezők tükrében.* Szeged: Juhász Gyula Felsőoktatási Kiadó.
- Nevalainen, T., Raumolin-Brunberg, H. 1996. *Sociolinguistics and Language History – Studies based on the Corpus of Early English correspondence.* Amsterdam–Atlanta: Rodopi.
- Papp, L. 1961. *Nyelvjárás és nyelvi norma XVI. századi deákjaink gyakorlatában.* Budapest: Akadémiai.
- Vass, E. 1985: A város önkormányzata és gazdálkodása. In: Farkas, J. (szerk.) *Szeged története: 1686–1849.* Szeged: Somogyi Könyvtár. 517–98.
- Végh, J. 1971. A nyelvjárás jellemzése. In: Juhász, A. (szerk.) *Tápé története és néprajza.* Tápé: Tápé Község Tanácsa. 813–837.



## Források

A források levéltári jelzetének közlésében Brandl–Tóth G. (2016) megoldását követem.

- CCVII. 44. irat: Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. 44.
- CCVII. 65. Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. 65.
- CCVII. 67. Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. 67.
- CCVII. 70. Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. 70.
- CCVII. 72. Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. 72.
- CCVII. 74. Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. 74.
- CCVII. 75. Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. 75.
- CCVII. 76. Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. 76.
- CCVII. 77. Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. 77.
- CCVII. 79. Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. 79.
- CCVII. 80. Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. 80.
- Farkasné Nedelko Margit elleni tanúvallomások: Magyar Nemzeti Levéltár Csongrád Megyei Levéltára, Szeged, IV. A 1021 b, 66. Szeged Város Titkos Levéltára, Boszorkánypercek Iratai, 2. db. S. N.

## Doménspecifikus korpusz építése és validálása

Dodé Réka

ELTE BTK Nyelvtudományi Doktori Iskola  
kovacs.reka@nytud.mta.hu

**Kivonat:** Egy terminuskivonatoló fejlesztésének első lépése az adott domént lefedő korpusz építése, minél relevánsabb és minél több anyagból. Elegendő nagyságú korpusz kézzel történő gyűjtése azonban időigényes. Jelen tanulmány a disszertációhoz szükséges korpusz építésének módszertani vizsgálata, amelyben az ökoinnováció doménhez tartozó szövegeket gyűjtök és értékelek, oly módon, hogy hasonlóságot mérek a letöltött anyagok között. A szövegek letöltése egy filtereket alkalmazó programkóddal történt. A szkript magyarra történő átültetése után szükség volt egy kezdő URL-lista, illetve egy mintákat tartalmazó fájl összeállítására. A kezdő, manuálisan összeállított lista alapján a program letöltötte a szövegeket, amelyeket az előre megadott minták alapján szűrt. A 2000 szövegből a validáláshoz 20 darab szöveget választottam. Referenciaszövegeknek a legtöbb mintát tartalmazó 4 szöveg lett kijelölve. A szövegek előfeldolgozását követően a JSI Similarity programmal számoltam hasonlóságot (vektorok közötti koszinusz-hasonlóság). Nem parametrikus tesztet használtam annak megállapítására, hogy a hasonlósági értékek eltérnek-e az előre megállapított 0,5-es küszöbtől. Egyik szöveg esetében sem találtam szignifikáns eltérést a 0,5-es értéktől. A szövegek szövszáma és hasonlósága között enyhe negatív kapcsolatot találtam, amelyet egy kiugró szószámmal rendelkező szöveg okozott.

### 1 Bevezetés

Napjainkban az interneten fellelhető tartalmak tömegesen elérhetők, és némi ráfordítással milliárdos nagyságrendű korpuszok állíthatók elő. Ezek előnyei, hogy számos nyelvi jelenség megfigyelhető bennük, hátrányuk viszont, hogy sokféle szöveget tartalmaznak, így nem tudunk például ezeken olyan ún. doménspecifikus osztályozókat tanítani és építeni, amelyek csak az adott típusú szövegre jellemzőek.

Jelen tanulmány a doktori disszertációhoz tartozó előzetes kutatás, mivel értekezésem a magyar szövegeken történő terminuskivonatolással foglalkozik, és a terminuskivonatoló alkalmazás fejlesztésének első lépéseként szükség van egy olyan korpuszra, amely lehetőleg minél relevánsabb és minél több anyagot tartalmaz a vizsgált doménből. (Nagy 2012).

A disszertációm az ökoinnováció tárgykörrel foglalkozik, a domén szövegeiből áll majd a használt doménspecifikus korpusz. Az ökoinnováció mint fogalom nem feltétlenül egyértelmű, így nézzük, hogyan definiálja az ökoinnovációt az Európai Bizottság *Az ökoinnovációs cselekvési tervben* (Eco-AP 2011: 3).

„Az ökoinnováció az innováció minden olyan formája, amelynek eredménye vagy célja a fenntartható fejlődés irányába történő jelentős és igazolható előrelépés a környezeti hatások csökkentése, a környezetterheléssel szembeni ellenálló képesség növelése, vagy a természeti erőforrások hatékonyabb és felelősségteljesebb felhasználásának megvalósítása révén” (Eco-AP 2011: 3).

A választott domén korszerű technológiákkal (hulladékkezelés, alternatív energiák stb.) kapcsolatos szövegeket, anyagokat foglal magában, így egyrészt folyamatosan változik, másrészt jelenleg még nem áll rendelkezésre nagy mennyiségű magyar nyelvű írásos anyag a témában, ami még inkább megnehezíti a kutatást. Mivel azonban innovatív és kurrens témáról van szó, a későbbiekben rendkívül hasznossá válhatnak a kutatás eredményei.

A tanulmány két fő részből áll: az első felében (2. fejezet) a korpusz építését részletezem, azon belül is a speciális mintákat használó keresőmotor működését, a második részben pedig a korpusz validálásának módszertanát és eredményeit (3. fejezet). A validáláshoz a korpusz egy részét használtam, amely így a 2000 szöveg 1%-ából, azaz 20 véletlenszerűen kiválasztott szövegből, illetve 4 referenciaszövegből áll. Ezek között a szövegek között mértem hasonlóságot a JSI Similarity Service<sup>1</sup> segítségével. Az eredményeket és következtetéseket tartalmazó részben (4. fejezet) részletesen bemutatam a kapott eredményeket, melyeket statisztikailag elemeztem.

A kutatáshoz két hipotézist állítottam fel.

- (1) A letöltött szövegek elérik a 0,5-es hasonlósági értéket.
- (2) Minél kevesebb szövegszóból áll a szöveg, annál kisebb lesz a vektorok közti hasonlóság értéke az előre kijelölt referenciaszöveghez képest.

## 2 A korpusz építése

Ha nem az a célunk szövegek letöltésekor, hogy a gyűjtés különösebb tematikus megkötéssel történjen, a letöltés könnyedén kivitelezhető. Egy bizonyos doménhez tartozó korpusz építéséhez azonban már különböző módszerek alkalmazására van szükség. Ezekhez a módszerekhez pedig már szükség van valamilyen korpuszvezérelt eredmény, eszköz, forrás felhasználására, mint például a WordNet, egy előre meghatározott taxonómia vagy egy n-gram statisztika (vö. Remus–Biemann 2016; Safran et al. 2012; Chakrabarti et al. 1999), amely kevésbé támogatott nyelvek esetében nem minden esetben áll rendelkezésre. Az irányított webkeresés (focused crawling, topical crawling, directed crawling) egy olyan webes letöltési folyamat, amely irányított módon specifikus témakörre fókuszál (Remus–Biemann 2016: 3607).

Jelen doménspecifikus korpusz építése egy irányított keresőmotor segítségével történt, amelyet Gregory Greffenstette és Lawrence Muchemi mutattak be 2016-ban egy nemzetközi konferencián (Greffenstette–Muchemi 2016). Greffenstette-ék keresőszkriptje angol nyelvre volt megírva, így azt először némileg módosítani kellett. Ez csupán annak a lépésnek az eltávolítását jelentette, amelyben a program angol szavakat keresett a szövegekben (quickEnglish.awk). A magyarban ezt a lépést amiatt nem tartottuk különösképpen fontosnak, mivel a kulcsszavak listája 48 elemből állt, ezért kisebb volt rá az esély, hogy idegen nyelvű oldalak is bekerülnek a letöltött szövegek közé. A későbbi kutatásokban azonban tervezem alkalmazni.

<sup>1</sup> Elérhető: <http://aidemo.ijs.si/xling/wikipedia.html>. Letöltve: 2017. július 14.

A szkript shell programozási nyelvben van megírva, és az awk (szövegfeldolgozó és programozási nyelv), illetve fgrep (szövegfeldolgozó) parancsokkal dolgozza fel az előre megadott szöveges mintákat: az URL-eket és a kulcsszavakat. Az awk parancs a UNIX-rendszerek része, így nem szükséges külön installálni. Az awk és az fgrep nagyon hatékony parancsok, amelyeket szöveges fájlok feldolgozására, például minták keresésére használnak. A szkriptben jelen futtatáskor maximum 2000 oldal letöltése volt beállítva paraméterként. A kód egy része azt a célt szolgálta, hogy egy oldal csak egyszer kerülhessen bele a letöltött szövegek közé – amennyiben megtalálható volt benne a mintafájl legalább egy mintája.

A futtatáshoz szükség volt egy 40 elemből álló kezdő URL-listára, amelyet kézzel állítottam össze – a böngészőben az „ökoinnovációra” kapott találatokból kiindulva, illetve egy mintákat tartalmazó fájlra, amelyek gyakorlatilag kulcsszavak, kulcskifejezések és azok variánsai voltak a weboldalakból kiindulva. Összesen 48 kifejezést tartalmaz, amelyek magukba foglalják a kisbetűs, nagybetűs, kötőjeles, kötőjel nélküli variánsokat, a gépelési hibákat (rövid, hosszú magánhangzó) és néhány eltérő szótővariánsot. Ezek felvételére a karakterek pontos egyezése miatt volt szükség (pl. *fenntartható energia*, *Fenntartható energia*, *ökoinnováció*, *öko-innováció*, *napenergia*, *napenergiá*). Az eredeti Greffenstette–Muchemi-féle korpusz-összeállításban a mintafájl csupán 2 elemű volt, a domén kis és nagybetűs variánsa. A szkript a letöltött kezdő weboldalakon található szövegek, illetve az ott található URL-ek letöltése után a mintafájlok alapján elkülönítette a jó szövegeket a nem megfelelő szövegektől annak alapján, hogy megtalált-e benne legalább egy kulcskifejezést a felsoroltak közül. A 2000 letöltött szöveg hossza 2 625 164 szövegszó lett.

### 3 A részkorpusz validálása

A kutatás második felében az összeállított korpuszt validáltam. Azt a tanulmány első felében láthattuk, hogy a letöltés és a szövegek kiválasztása egy mintákat tartalmazó fájl alapján történt, mégis azt feltételeztem, hogy a validálásra szükség lehet ahhoz, hogy ténylegesen úgy kezelhessük az összeállított korpuszunkat, mint az ökoinnovációs domént reprezentatívan lefedő szövegek halmazát, amelyben a szövegek hasonlóak, tehát ugyanahhoz a doménhez tartoznak. A következőkben a validálás módszerét fejtem ki.

#### 3.1 A részkorpusz összeállítása

Mind a 2000 szöveg összehasonlítására az idő hiányában nem volt lehetőségem, így 20 szöveget választottam véletlenszerűen (randomgenerátor), csupán azzal a feltétellel, hogy a szöveg szószáma haladja meg az 500-at. Ezt azért tartottam fontosnak, hogy az összehasonlításnál kellő szó álljon rendelkezésre, illetve mert azt feltételeztem, hogy a kevés szövegszóból álló fájlok olyan kezdőoldalak, amelyeken a szövegek inkább csak menüpontok, illetve egyéb nem releváns pontok, mint Kapcsolat / Elérhetőség, Szervezetről / Rólunk stb.

Az összehasonlítandó részkorpusz összeállítása után kerültek kiválasztásra a referenciaszövegek. A 2000 szövegből összesen négyet választottam referenciaszövegnek annak alapján, hogy egyrészt nem tartoznak a már kiválasztott szövegek közé, illetve ezekben fordult elő a legtöbb kulcskifejezés.

Mind a részkorpusz szövegeit, mind a referenciaszövegeket tokenizáltam, lemmatizáltam, és kiszűrtem belőlük a gyakori nem tartalmas szavakat (stopszósűrítés). A tokenizálást és lemmatizálást a 2016-ban elkészült *e-magyar* elemzőláncban lévő eszközökkel, az *emToken* és *emLem* eszközökkel végeztem (Mittelholcz 2017, Váradi et al. 2017). Az így kapott – stopszavak nélküli, szótövesített – szövegeken számoltattam aztán hasonlóságot. A 20 random kiválasztott szöveg átlag hossza stopszavak nélkül 1428 szövegszó, a szövegek hosszának szórása 1054,49. A 4 referenciaszöveg átlag hossza 4936, a szövegek hosszának szórása 3062,63.

### 3.2 A hasonlóság mérése

A hasonlítás során minden szöveget összevettem egymással, és a kapott eredményeket elemeztem. A hasonlítás a JSI Similarity Service webszolgáltatással történt, amely a 2012–2014 zajló XLike projekt keretében készült. Az XLike projekt célja egy olyan nyelvtechnológiai infrastruktúra készítése volt, amely hatékonyan nyer ki információkat különböző nyelvű szövegekből, és a nyelveken átívelő adatokkal segíti a különböző eszközök fejlesztését pl. hasonló cikkek ajánlását különböző nyelveken (García-Cuesta et al. 2014: 10). Ennek az infrastruktúrának az egyik eleme a JSI Similarity Service, amely (akár eltérő nyelvű) szövegek hasonlóságát állapítja meg 0 és 1 közé eső mértékben.

#### 3.2.1 A korpusz- és dokumentumhasonlítás elméleti háttere

Nem új keletű igény, hogy korpuszokat, dokumentumokat képesek legyünk objektíven összehasonlítani. Adam Kilgarriff 2001-es tanulmányának célja olyan különféle módszerek bemutatása, amelyekkel korpuszokat lehet egymással összehasonlítani, minél objektívebb módon (Kilgarriff 2001). Általában elmondható, hogy az összes módszer a szövegek szavainak előfordulásával számol különböző módokon. Kilgarriff először a statisztikai módszereket ismerteti: khi-négyzet-próba, t-teszt, kölcsönös információs (MI), log-likelihood, Fischer-egzakt-teszt, TF-IDF, melyek közül a Mann–Whitney rangsorolási tesztet emeli ki, majd a valószínűségi eloszlás három módszerét: a Poisson-eloszlást, a binomiális eloszlást és a normál eloszlást. (Ezek a szavak gyakoriságával és azzal a ténnyel számolnak, hogy a tartalmas szavak halmozottan fordulnak elő, szemben a grammatikai szavakkal.) Végül az emberi interpretációs módszerek közül mutat be néhányat (előre meghatározott fogalmak, dimenziók nyelvi megnyilvánulásai). Ezt követően a korpuszhasonlóság, korpuszhomogenitás számolására javasol megoldást. Ehhez elkészíti a „Known-Similarity Corporát”, a korpuszhasonlóság mérésére szolgáló gold standardot, majd az 500 leggyakoribb szón Spearman-féle rangkorrelációt, khi-négyzet-próbát és keresztentropiát számol a hasonlósági értékhez. Bemutatja, hogy a módszerek közül a khi-négyzet-próba teljesített a legjobban.

A szövegek, dokumentumok hasonlóságának számolása az általam használt módszerrel egészen a disztribúciós szemantika elméletéig, annak háttere pedig Saussure-ig vezethető vissza, aki megfigyelte, hogy a szavak jelentésének kulcsa a szavak funkciójának különbsége (Saussure 1916/1983). Ennek két típusa van, a szintagmatikus és a paradigmikus. A szintagmatikus kapcsolat egy szó helyzetére vonatkozik, azon entitások kapcsolatára, amelyek együtt fordulnak elő a szövegben

(szekvenciális kombinációk). Ilyen például egy mondat a szövegben. A paradigmikus kapcsolat ezzel szemben a szó helyettesítésére vonatkozik, azon entitások kapcsolatára, amelyek ugyanabban a kontextusban fordulnak elő, de nem ugyanakkor (pl. tej, kávé, tea, szörp). A disztribúciós hipotézis Harris nevéhez fűződik. A hipotézis szerint a szavak, amelyek hasonló kontextusban jelennek meg, szemantikailag kapcsolódnak egymáshoz (Harris 1968, 1970). A disztribúciós hipotézis a korpusznyelvészet megjelenésével tudott teret hódítani, mivel modellje az empirikus (korpuszok által szolgáltatott) adatoknak egy geometrikus, vektoralapú interpretációján alapul (Sahlgren 2008). A vektorok és az azokból álló mátrixok képesek dokumentumokat reprezentálni a szavak előfordulási gyakorisága szerint.

### 3.2.2 A hasonlóságot mérő eszköz

Az XLike dokumentumindexelője, a JSI Similarity Service fogalomalapú módszereket használta a dokumentumok összehasonlítására. E módszer a dokumentumokat egy többdimenziós térbe transzformálja oly módon, hogy a dokumentumok kifejezéseinek súlyai, fontossága (term weight) által reprezentált vektorokat egy ún. illesztett konceptuális vektorra transzformálja a többnyelvű konceptuális térben, és azokat veti össze egymással (vektortér-modellezés). Ezek a vektorok a fogalmak egyfajta leírásai különböző nyelveken. Ezen belül a fogalmak meghatározásához különböző elméleteket használtak: K-középpontú klaszteranalízist, látens/rejtett szemantikai indexelést, explicit szemantikai analízist és kanonikus korrelációelemzést. A vektorok hasonlóságát a koszinusztávolsággal számolják, amely a két vektor által bezárt szög koszinusza (Rettinger et al. 2012).

Az 1. ábrán látható 0,88-os hasonlósági érték két génmódosítással kapcsolatos szöveghez tartozik. Ezek tartalmilag csak némileg tértek el egymástól. A következő részben foglalkozom a 0,5-es küszöbszint kijelölésével.

Similarity is:  
0.880357

Report

Dmoz cat Hungarian Hungarian Dmoz cat

Words that add the most to the similarity  
**növény élelmiszer szervezet mezőgazdaság eu vagy ember is termék oly**

Words that add the most to the similarity  
**növény élelmiszer hogy vagy eu is ember az ország leh**

Summary of categories:  
**agriculture science horticulture business forestry**

Summary of categories:  
**science agriculture plants home horticulture**

Top/Science/Agriculture/Crop\_Plants/  
Top/Science/Agriculture/Horticulture/  
Top/Science/Agriculture/Field\_Crops/  
Top/Business/Agriculture\_and\_Forestry/Biologicals/  
Top/Home/Gardening/Soil\_and\_Additives/  
Drogbá Inspires Ivory Coast to 2-1 Win Over Japan

Top/Science/Agriculture/Crop\_Plants/  
Top/Science/Agriculture/Horticulture/  
Top/Science/Agriculture/Field\_Crops/  
Top/Business/Agriculture\_and\_Forestry/Horticulture/  
Top/Shopping/Home\_and\_Garden/Plants/  
世界杯: 科特迪瓦下半场连进两球胜日本

Az európaiak nagyrészt a génmanipuláció-mentes mezőgazdaság hívei, ennek ellenére az EU a tagállamokban engedélyezte az emberi és állati fogyasztásra szánt génmódosított termékek kereskedelmét, illetve a repace, a kukorica és a szója termesztését is.  
Jelenleg nem engedélyezett Magyarországon a genetikailag módosított növények termesztése, és ilyen élelmiszerek előállítása sem folyik, azonban egyre több híresztelést hallunk, miszerint különböző génmódosított (GMO) élelmiszerek lépik át a magyar határt és kerülnek az asztalunkra.  
Heves vitákat generál  
Számos szakember vitázik azon, hogy a genetikailag módosított élelmiszereknek milyen hatásai lehetnek az emberi szervezetre. A megosztó szakmai vélemények között akad olyan, ami előrevetíti az esetleges egészségkárosító hatásukat és olyan is, ami szerint nem kell számolni ilyen jellegű veszéllyel. Az viszont biztos, hogy már önmagában a környezeti károk és az élővilág

A génmódosított élelmiszerekre vonatkozó előírások rendkívül szigorúak az EU-ban, génmódosított növényeket is csak kockázatelemzést követően lehet termesztetni. Három évnyi vita után az EU Tanács jóváhagyta azt a javaslatot, amely nagyobb rugalmasságot biztosítana a tagállamoknak abban, hogy területükön szabályozzák a GMO-növények termesztését. Az EP plenárisa január 13-án kedden délelőtt vitázik majd szavaz a témáról. A vitát és a szavazást honlapunkon is élőben lehet követni. Háttér.  
Lehet-e genetikailag módosított növényeket termesztetni az EU-ban?  
Igen, amennyiben uniós szinten az Európai Élelmiszerbiztonsági Hatóság (EFSA) engedélyezte azt. Az engedélyezést követően az uniós országok egy **védelemdélre** hivatkozva tilthatják csak be a génmódosított terméket a területükön. A döntést indokolni kell, és bizonyítani, hogy a növény ártalmas lehet a környezetre vagy az emberekre.

1. ábra. A génmódosításról szóló szövegek hasonlósága a JSI programban

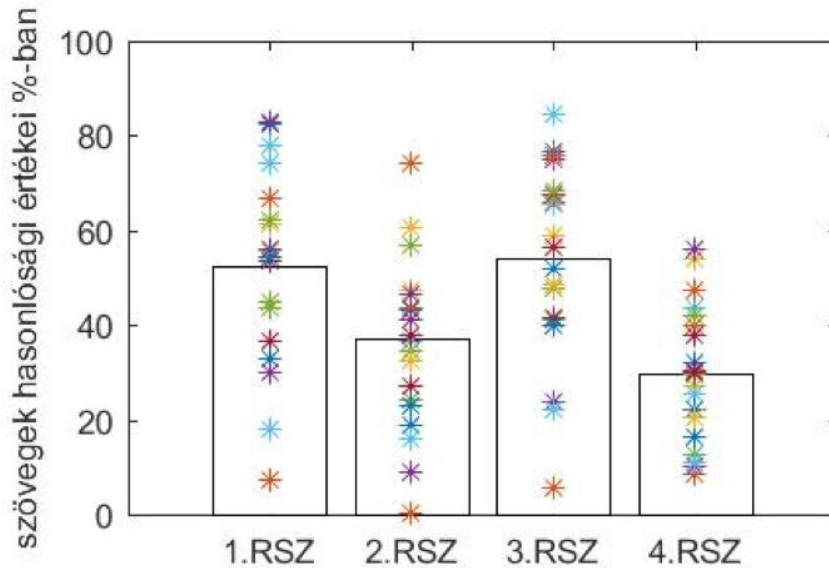
## 4 Eredmények és következtetések

A következőkben az eredményeket ismertetem a hipotézisek alapján, ezt követően pedig az eredmények alapján levonok néhány következtetést a kutatással kapcsolatban.

### 4.1 Első hipotézis

A letöltött szövegek elérik a 0,5-es hasonlósági értéket.

A küszöbérték meghatározása előzetes tapasztalatok alapján történt. Míg azok a szövegek, amelyek egymásnak fordításai voltak, 0,92–0,95-os hasonlósági értéket értek el, a tematikusan megegyező szövegek (lásd 1. ábra) 0,85–0,88, a tematikusan eltérő szövegek 0,25–0,3-es értéket. Ezeknek a tapasztalatoknak a birtokában állítottam fel a 0,5-es küszöbértéket, számolva a szövegek eltérő hosszával és az előforduló zajjal (nem releváns szövegek megjelenése). Az eredmények a 2. ábrán és az 1. táblázatban láthatók.



2. ábra. A szövegek hasonlósági értékei referenciaszövegenként százalékban

A 2. ábrán látható az összes szöveghez tartozó hasonlósági érték referenciakorpuszonként. (Az egyszerűség kedvéért a 0 és 1 közötti értékek helyett 0 és 100 közötti értékeket használok.) Az ábrán látható az átlagos érték is, illetve látható a szórás mértéke is. Az 1. táblázat ezeket számszerűsítve mutatja.

| Referenciaszöveg | 1     | 2     | 3     | 4     |
|------------------|-------|-------|-------|-------|
| Átlag hasonlóság | 0,524 | 0,373 | 0,541 | 0,297 |
| Szórás           | 0,198 | 0,192 | 0,201 | 0,14  |

1. táblázat. Átlag és szórás referenciaszövegenként.

Az átlagokat tekintve látható, hogy az 1-es és a 3-as szövegek esetében átlagosan magasabb a hasonlósági érték, azonban statisztikailag számolva ez nem mondható el. Mivel az értékek nem normál eloszlásúak voltak, ezért non-parametrikus tesztet (az egymintás t-próba nem parametrikus megfelelőjét) használtam annak megállapítására, hogy a hasonlósági értékek eltérnek-e az előre megállapított 0,5-es küszöbtől. Egyik szöveg esetében sem találtam szignifikáns eltérést az 0,5-es értéktől ( $p_{1szöveg} = 0,68$ ;  $p_{2szöveg} = 0,36$ ;  $p_{3szöveg} = 0,8$ ;  $p_{4szöveg} = 0,22$ ).

A referenciaszövegek között is mértem hasonlóságot. A eredményeket a 2. táblázatban láthatók.

| Referenciaszövegek | 1-2   | 1-3   | 1-4   | 2-3   | 2-4   | 3-4   |
|--------------------|-------|-------|-------|-------|-------|-------|
| Hasonlóság         | 0,585 | 0,844 | 0,443 | 0,772 | 0,568 | 0,505 |

2. táblázat. Referenciaszövegek hasonlósága egymáshoz mérve



A 3. táblázat a szövegek hasonlóságainak eltérését mutatja referenciaszövegenként. Ennek megállapítására a páros t-próba nem parametrikus formáját használtam.

| Referenciaszövegek |    | p-érték         | Szignifikáns<br>(<0,05*, <0,01**, <0,001***) |
|--------------------|----|-----------------|--|
| 1.                 | 2. | 0,021           | *  |
| 1.                 | 3. | 0,695           |  |
| 1.                 | 4. | <0,001 (0,0006) | ***  |
| 2.                 | 3. | 0,011           | *  |
| 2.                 | 4. | 0,223           |  |
| 3.                 | 4. | <0,001 (0,0004) | ***  |

**3. táblázat.** Referenciaszövegek hasonlóságának eltérése a szövegek hasonlóságának függvényében

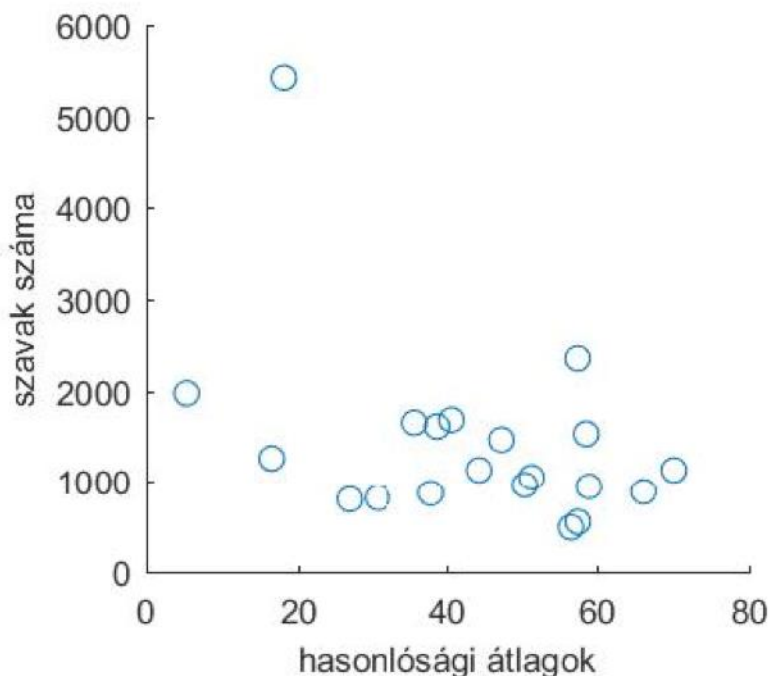
A táblázatból látható, hogy az 1-es, a 4-es, illetve a 3-as és a 4-es referenciaszövegek között erős szignifikáns, míg az 1-es és 2-es, továbbá a 2-es és a 3-as szövegek között szignifikáns eltérés van.

Az első hipotézis részben teljesült, mivel egyik referenciaszöveghez mért átlag hasonlósági érték sem lett magasabb, mint 0,5.

#### 4.2 Második hipotézis

A második hipotézisem, hogy minél kevesebb szövegszóból áll a szöveg, annál kevesebb lesz a hasonlósági értéke a referenciaszöveghez képest.

A második hipotézis tehát a szövegek szószáma és a hasonlóság mértékének kapcsolatára vonatkozott. Ennek megállapítására korrelációs analízist használtam. Ebből az látszott, hogy egy enyhe tendenciózus ( $0,05 < p < 0,1$ ) negatív kapcsolat van közöttük ( $r_{\text{Pearson}} = -0,41$ ,  $p = 0,071$ ), amely ellentétes a várttal, tehát minél több szóból áll a szöveg annál kevésbé hasonló, viszont ahogy a 3. ábrán látható, ezt az eltérést egyetlen kiugró szószámmal rendelkező szöveg okozta. Ennek a szövegnek az ignorálásával az értékek a következőképpen alakulnak:  $r_{\text{Pearson}} = -0,25$ ,  $p = 0,30$ .



3. ábra. A szövegek hasonlósági értékei referenciaszövegenként százalékban

Mind ezek alapján elmondható tehát, hogy a második hipotézis nem teljesült, mivel a szövegek mérete és az eredmény között ugyan mutatkozott egy gyenge negatív kapcsolat, azt azonban egyetlen szöveg okozta.

## 5 Összefoglalás és kitekintés

A kutatás a doktori disszertációhoz kapcsolódó pilot kutatás volt, melyben egy ökoinnovációhoz tartozó korpuszt állítottam össze egy kezdő URL-lista és egy mintákat tartalmazó fájl alapján. Ezt követően a korpusz egy részkorpuszán végeztem összehasonlítást a szövegek között, melynek során arra voltam kíváncsi, hogy az egyébként irányított módon letöltött szövegek valóban hasonlóak-e, azaz valóban az ökoinnováció tárgykörhöz tartoznak. Az első hipotézis nem igazolódott be, de összességében elmondható, hogy a letöltéshez használt szkript robusztus, egyszerűen átvihető más nyelvekre, és előnye más irányított webkeresővel szemben, hogy nincs szükség a szövegek előzetes elemzésére, illetve más módszerek, eszközök, források felhasználására, csupán egy átgondolt kulcskifejezés-lista összeállítására. A lista összeállításakor számolnunk kell a minták karakterszintű egyezésével, tehát a kulcsszavak variánsait (kisbetű/nagybetű, eltérő helyesírás, eltérő szintaktikai szerkezetek) is érdemes felvenni, illetve átgondolni a túl általános vagy épp túl speciális kulcsszavak megadását.

A kutatás arra is rámutat, hogy mindezek mellett is szükség lehet a korpusz validálásra. A JSI Similarity Service – bár nyelvek közötti hasonlóság számolására készült – egy nyelven belül is használható. A vektortérmodell mint szöveg-összehasonlítási módszer jó irány, mivel bár a szavak előfordulási gyakoriságát használja, a vektorrepresentációk miatt képes egyéb szemantikai információkat is kezelni (pl. szinonimitás). A későbbiekben érdekes volna további szöveg-összehasonlítási módszereket alkalmazni, mint például a khi-négyzet-próba.

A szövegek és a referenciaszövegek is eltéréseket mutattak, ami az ökoinnováció domén ismeretében annak tudható be, hogy rendkívül széles skálán mozognak az aldomének. Láthatjuk az ökoinnováció meghatározásából, hogy a megújuló energia aldomén mellett az újrahasznosítás is az ökoinnováció területe alá tartozik, miközben azok intuitíve és a szókincsben is nagy különbségeket mutatnak. Így tehát a későbbiekben megfontolandó, hogy figyelembe kell venni az ökoinnováció aldoménjeit, és esetleg szűkíteni a vizsgált területeket egy-egy aldoménre. Mindenesetre a további statisztikai vizsgálatok a minta növelését igénylik.

További érdekes kutatási módszer lehetne a szövegek összehasonlítására az emberi annotátorok bevonása, akiknek azt kellene eldönteniük, hogy a szöveg milyen mértékben tartozik a doménhez, illetve miért.

## Irodalom

- Chakrabarti, S., Van den Berg M., Dom B. 1999. Focused crawling: a new approach to topic-specific Web resource discovery. *Computer Networks*, 31(11): 1623–1640.
- De Saussure, F. 1916/1983. *Course in general Linguistics*. London: G. Duckworth.
- Európai Bizottság 2011. *Innováció a fenntartható jövőért – Az ökoinnovációs cselekvési terv (Eco-AP)*. Brüsszel, 2011.12.15. COM(2011) 899 végleges. [http://kormyezettechnologia.kormany.hu/dow\\_nload/1/a3/40000/EcoAP.pdf](http://kormyezettechnologia.kormany.hu/dow_nload/1/a3/40000/EcoAP.pdf). Letöltve: 2017. február 3.
- García-Cuesta, E., Caparros A., Fortuna B., Carreras X., Zhang L., Li Z., Rettinger A. 2014. *Final toolkit architecture specification*. D6.1.2 XLike project. Elérhető: <http://www.xlike.org/wp-content/uploads/2012/03/D6.1.2-Final-Toolkit-architecture-specification.pdf>. Letöltve: 2017. február 12.
- Grefenstette, G., Muchemi, L. 2016. Determining the Characteristic Vocabulary for a Specialized Dictionary using Word2vec and a Directed Crawler. In: Kernerman, I., Kosem I., Krek S., Trap-Jensen L. (szerk.) *Lexicographic Resources for Human Language Technology GLOBALEX 2016 Workshop Proceedings*. Portorož, 24 May 2016. Elérhető: [http://ailab.ijs.si/globalex/files/2016/06/LREC2016Workshop-GLOBALEX\\_Proceedings-v2.pdf](http://ailab.ijs.si/globalex/files/2016/06/LREC2016Workshop-GLOBALEX_Proceedings-v2.pdf). Letöltve: 2016. december. 3.
- Harris, Z. 1968. *Mathematical structures of language*. New York: Interscience Publishers.
- Harris, Z. 1970. Distributional structure. In: Harris, Z. (szerk.) *Papers in Structural and Transformational Linguistics*. Formal Linguistics Series. Dordrecht: Springer Netherlands. 775–794.
- Kilgarriff, A. 2001. Comparing Corpora. *International Journal of Corpus Linguistics*, 6(1): 97–133.
- Mittelholcz I. 2017. emToken: Unicode-képes tokenizáló magyar nyelvre. In: Vincze, V. (szerk.) *XIII. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY 2017*. Szeged: JATEPress. 61–70.
- Nagy Á. 2012. *Terminológiakivonatolás francia nyelvű szabadalmi leírásokból szabály alapú és statisztikai módszerek segítségével*. PhD-értekezés. Szegedi Tudományegyetem Bölcsészettudományi Kar, Nyelvtudományi Doktori Iskola. Elérhető: [http://doktori.bibl.u-szeged.hu/1768/1/disszertacio\\_NagyAgoston\\_egyben.pdf](http://doktori.bibl.u-szeged.hu/1768/1/disszertacio_NagyAgoston_egyben.pdf). Letöltve: 2016. május. 15.
- Remus, S., Biemann, Ch. 2016. Domain-Specific Corpus Expansion with Focused Webcrawling. In: Calzolari, N., Choukri K., Declerck T., Goggi S., Grobelnik M., Maegaard B., Mariani J., Mazo H., Moreno A., Odijk J., Piperidis S. (szerk.) *Proceedings of the Tenth International*

- Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, 23–28 May 2016. Párizs: European Language Resources Association (ELRA). 3607–3611.
- Rettinger, A., Zhang L., Rupnik J., Muhič A. 2012. *Cross-lingual document linking prototype*. D4.1.1 XLike project. Elérhető:  
<http://cordis.europa.eu/docs/projects/cnect/2/288342/080/deliverables/001-D411Crosslingualdocumentlinkingprototype.pdf>. Letöltve: 2017. február 12.
- Safran, M. S., Althagafi A., Che D. 2012. Improving Relevance Prediction for Focused Web Crawlers. In: 2012 IEEE/ACIS 11th International Conference on Computer and Information Science. Los Alamitos, California: IEEE Computer Society. 161–167.
- Sahlgren, M. 2008. The Distributional Hypothesis. *Rivista di Linguistica* 20(1): 33–53.
- Váradi T., Simon E., Sass B., Geröcs M., Mittelholcz I., Novák A., Indig B., Prószték G., Farkas R., Vincze V. 2017. Az e-magyar digitális nyelvfeldolgozó rendszer. In: Vincze, V. (szerk.) *XIII. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY 2017*. Szeged: JATEPress. 49–61.

## Hány VAN nincs? A létige – zéró váltakozás korpuszvezérelt vizsgálata

Dömötör Andrea

PPKE BTK Nyelvtudományi Doktori Iskola  
domotor.andrea@itk.ppke.hu

**Kivonat:** A kutatás korpuszvezérelt módszerrel vizsgálja, hogy a hagyományos értelemben vett zéró kopulán kívül milyen szintaktikai és szemantikai feltételekhez kötött a létige elhagyhatósága. A korpuszadatok szerint a létige lokatív mondatokban is lehet zéró alakú, különösen ha mutató névmás mellett áll. Ez arra utal, hogy a mutató névmásoknak – az angolhoz hasonlóan – közük van a létezés kifejezéséhez. Tagadó jelentésű egzisztenciális és birtokos mondatokból a *nincs* is elhagyható a tagadó névmások (*semmi, sehol*) mellől. A kutatás ezen tapasztalatai alapján a függőségi nyelvtanon alapuló automatikus mondatelemzők modelljének építéséhez a funkciófej-megközelítés (Simkó–Vincze 2016) tűnik jól alkalmazhatónak, a létige tartalmassága ugyanis nem feltétlenül meghatározó tényező az elhagyhatóság kérdésében. Az elemző eszközöket a formális nyelvleírásból ismert tipikus kopulákon kívül más lehetséges zéró alakú létigék jelenlétére is fel kell készíteni.

### 1 Bevezetés

A zéró kopula jelensége számos nyelvben ismert, de nincs univerzális elv arra, hogy a kopula mikor lehet testetlen. A magyarban az úgynevezett Third Person Parameter (Stassen 1994) érvényesül, azaz a jelen idő harmadik személy esetén nem jelenik meg testesen a létige. A kopula fogalmának értelmezésében azonban eltér a magyar és az angolszász szakirodalom. A magyar leíró nyelvészet hagyományosan csak a nominatívusi névszói predikátum esetén beszél kopuláról (Kádár 2011), ezzel szemben a nemzetközi, például az angol (pl. Heggie 1988) vagy a spanyol (pl. Maienborn 2005) irodalom ide sorolja a létige + PP szerkezeteket is.

Megfigyelhető, hogy a létige-zéró váltakozás valójában a magyarban sem korlátozódik csupán a nominatívusi névszói predikátumokra. Az ige hiánya lokatív állítmányú mondatokra is jellemző, például: *Hol a könyved?, Kulcs a lábtörő alatt.*

A kutatás azt vizsgálja, hogy a hagyományos értelemben vett zéró kopulán kívül milyen szintaktikai és szemantikai feltételekhez kötött a létige elhagyhatósága. A vizsgált létigetípusokat az 1. táblázat tartalmazza.

| Típus           | Példa                                 |
|-----------------|---------------------------------------|
| kopula          | <i>Az előadás érdekes volt.</i>       |
| lokatív         | <i>A macska az udvaron van.</i>       |
| egzisztenciális | <i>Van egy macska az udvaron.</i>     |
| birtokos        | <i>A szomszédnak zöld kertje van.</i> |

1. táblázat. A dolgozatban vizsgált létigetípusok

Az előfeltevés szerint a lokatív argumentumot tartalmazó létigés mondatok viselkedése a névszói predikátumosokhoz hasonló lesz, azaz a Third Person Parameter érvényesül, és csak akkor hagyható el a létige, ha az alany specifikus (azaz a mondat nem egzisztenciális jelentésű).

## 2 Motiváció

A kutatás eredményeit egy függőségi nyelvtanon alapuló automatikus mondatelemző modelljének tervezéséhez szeretnénk felhasználni. A tervezett elemző (Prószéky–Indig 2015) pszicholingvisztikai indíttatású, azaz a gépi szövegfeldolgozást az emberi értéshez próbálja hasonlóná tenni. Működés során az elemző szigorúan balról jobbra halad, a szöveg beolvasása szavanként (szóközönként vagy írásjelenként) történik. Az éppen elemzendő egységet az elemző különböző szálai (morfológiai elemző, különböző grammatikai jelenségeket azonosító szájak, korpuszgyakorisági szájak, anaforafeloldó szájak, fókuszazonosító szál stb.) egy időben, párhuzamosan vizsgálják, így egymás eredményeit is javíthatják, pontosíthatják. Minden szó beolvasása után első feldolgozási szálként elindul a morfológiai elemzés, amely létrehozza a szó lemmáját, és megadja a további elemzéshez szükséges morfoszintaktikai jegyeket. Ezek lehetnek kereslet típusúak, azaz meghatározhatják, hogy milyen jegyekre tartanak igényt, vagy kínálat formájában kielégíthetik más szájak keresletigényét. Egy finit ige például a vonzatkeret-szótárból indít el esetvégződésre vonatkozó keresleti szájakat, amelyekkel aztán a megfelelő esetragos főnevek (NP-k) kapcsolódni tudnak. A morfológiai és szintaktikai relációkat leíró szájak mellett párhuzamosan egyéb információk (statisztikai információk, korpuszgyakoriságok, ontológiák, világismeret) is részt vesznek az elemzésben. Az elemzés kimenete a függőségi nyelvtanokra emlékeztető irányítottgráf-reprezentáció, amely a megnyilatkozás által jelölt nyelvi szituáció minden fontos elemét igyekszik azonosítani.

A számítógépes nyelvészek többféle módon próbálták a *van* létigét függőségi nyelvtani keretben kezelni. Simkó és Vincze (2016) három lehetséges elmélet eredményeit hasonlítja össze a Szeged Korpusz adatain. A funkciófej-elemzés a mondatban a funkciószavakat tekinti fejnek. Ennek megfelelően ha a mondatban nincs ragozott ige, akkor egy üres igealakot kell beszúrni a szerkezetbe. Ezzel szemben a tartalmasfej-megközelítés a jelentéses egységeket tekinti a mondat fő elemeinek. Ebben az értelmezésben csak a „tartalmas” létige lesz fej, kopula esetén a névszói predikátum a mondat feje, és a létige ehhez kapcsolódik (ha testes). A harmadik lehetséges módszer a komplex címkés elemzés, amely lényegében megegyezik a funkciófej-elemzéssel, csak a zéró kopula esetén tér el attól. Ha a kopula nem jelenik meg testesen a felszíni

szerkezetben, akkor a mondat feje a névszói predikátum lesz, és a hiányzó ige ennek címkéjén és a belőle kiinduló éleken lesz jelölve.

Simkó és Vincze (2016) kísérleteiben a tartalmasfe-elemzés érte el a legjobb eredményt, azaz a gépi tanulásra alapozott automatikus elemző ezzel a címkézési módszerrel tudta a legnagyobb pontossággal felismerni a zéró kopulát. Egyúttal a szerzők hangsúlyozzák az elméleti keret átgondolásának jelentőségét. Jelen esetben fontos kérdés, hogy mely létigéket szeretnénk tartalmas egységnek tekinteni, és hogy pontosan milyen esetekben kell felkészülni a zéró létigék meg(nem)jelenésére.

### 3 Elmélet és terminológia

Amint a bevezetőben is említettem, a magyar formális nyelvelírás szerint „(az elliptikus szerkezeteket leszámítva) csak az azonosító és predikatív mondatok VAN-ja lehet zéró (hangalakú), azaz csak ezek alkotnak nominális mondat szerkezetet” (Kádár 2011: 7) Ugyanakkor szintén a bevezetőben láthattunk olyan lokatív példákat is, amelyek szerint ez nem igaz. Úgy tűnik, a létige más esetekben is lehet opcionálisan zéró hangalakú. Mivel a kutatás korpuszvezérelt, nem állítok előzetes hipotéziseket arra vonatkozóan, hogy melyek ezek az esetek.

A kérdés vizsgálata azonban felvet egy taxonómiai problémát, nevezetesen azt a sokat vitatott kérdést, hogy hány létige van a magyarban, vagy pontosítva: hányféle funkcióval bírhat a *van* szótő. Jelen dolgozat Clark (1978) alapján (részletesen I. 5. fejezet) az I. táblázatban is ismertetett négy fő típust különbözteti meg. Kopula alatt a nominális mondat segédigéjét értem. Az ezen kívüli funkcióban használt létigéket a lokatív, egzisztenciális és birtokos csoportokba osztom fel. Fontos megjegyezni, hogy egzisztenciális mondatnak is lehet helyhatározói bővítménye, lokatív mondatról azonban csak specifikus alany esetén beszélünk.

További terminológiai megjegyzés még, hogy létige, illetve VAN alatt a *van*, *nincs(en)*, *sincs(en)* szótövek bármely alakját érthetjük. Kádár (2011: 7) alapján nem tipikus zéró létigének kell tehát tekinteni minden olyan esetet, ahol a VAN nem kopula, de a felszíni szerkezetben nem jelenik meg hangzó alakban. Erre fókuszálunk ebben a tanulmányban.

### 4 Anyag és módszer

A kutatás módszere korpuszvezérelt, azaz nagy mennyiségű szöveges adat vizsgálatán alapszik, nincsenek kiinduló hipotézisei, minden következtetést az adatokból próbál levonni. A felhasznált adatbázis az MNSz. (Oravecz–Váradi–Sass 2014) korpusz volt, amely különböző műfajú, morfológiailag annotált szövegeket tartalmaz. Jelenlegi, 2.0.4-es verziója körülbelül 1,04 milliárd szövegszóból áll. Az adatbázishoz tartozó keresőfelületen lehetőség van szóalak, szótő és szófajcímke szerint is keresni, valamint a kapott találatokra további szűrőket is be lehet állítani. A korpuszon alapuló kutatások nagy nehézsége, hogy a morfológiai címkék gépi eszközzel készültek, így sok esetben hibásak, nem megbízhatóak (például a  *vagy* létigét sokszor kötőszónak, sőt: az *álltal* igét főnévnek címkézi), így az adatgyűjtést sajnos nem lehetett teljesen automatizálni.

Első lépésben olyan mondatokat kerestem, amelyekben nincs finit ige. Ez a korpusz keresőfelületén reguláris kifejezéssel megoldható ugyan, ám a kapott eredmények alig

voltak használhatók, a találatok nagy része ugyanis nem valódi szöveget tartalmazott, hanem HTML-fájlokból származott. Ezért inkább egy 100 000 mondatos random adatból Python-script segítségével szűrtem ki azokat a ragozott ige nélküli mondatokat, amelyekben van nominativus. Így 17 040 találatot kaptam, azonban korántsem minden esetben zéró létige okozta az ige hiányát. A keresést sokszor címek, az egyes műfajokra (pl. országgyűlési naplók, színművek rendezési instrukciói) jellemző nominális stílus, ellipszis, hiányos mondatok vagy éppen hibás címkék zavarták. Emiatt elkerülhetetlen volt a találatok kézi ellenőrzése, és így jelenleg 1000 mondatnyi megbízható adat áll rendelkezésre.

#### 4.1 Kiinduló eredmények

A kiinduló adatok mindegyikében zéró hangalakú létige található, melyek közül 154 (15,4%) nem hagyományos értelemben vett kopula. Bár a minta nem nagy, annyi azért jól látszik, hogy a jelenség semmiképpen sem mondható marginálisnak. Ezenkívül a véletlenszerűen kapott 154 adat arra is jó volt, hogy irányt mutasson a további keresésekhez.

A kiinduló eredmények mondat típus szerinti megoszlását a 2. táblázat tartalmazza.

| Típus            | Altípus          | Gyakoriság |
|------------------|------------------|------------|
| Lokatív          | itt/ott/hol      | 27         |
|                  | egyéb            | 28         |
|                  | irodalmi         | 63         |
| Egyszintenciális | semmi/sehol/soha | 15         |
|                  | egyéb            | 3          |
| Birtokos         | semmi/mi         | 8          |
|                  | egyéb (vége)     | 5          |
| Egyéb            |                  | 5          |

2. táblázat. A kiinduló korpuszadatok

Ezekből az látható, hogy a nem tipikus zéró létigék az irodalmi szövegek leírásaiban a legjellemzőbbek (pl. *Az árokparton cseresznyefasor*). Jelen kutatás ezzel a típussal nem kíván foglalkozni. Érdekes azonban, hogy a „köznyelvi” adatok között is találunk 31 egyszintenciális vagy birtokos, illetve 55 lokatív mondatot. Ezeket közelebbről vizsgálva az is észrevehető, hogy előbbieket több mint kétharmada (23 darab) valamilyen tagadó vagy kérdő névmást (*sehol, semmi, soha, mi*) tartalmaz, például: *Címekre semmi szükségem*. A birtokos esetén érdekes még, hogy ezen az altípuson kívül csak a *vége* fordul elő testes létige nélküli mondatokban (pl. *A teremtésnek vége*). A lokatív mondatokra szintén nagy arányban (az adatok nagyjából felében) jellemző a névmás (*itt, ott, hol*) jelenléte, például: *Ott az a nagy gumikerék*. Az „Egyéb” típusban a *minek* és a *miért* kezdetű kérdések fordultak elő (pl. *De akkor miért ez a nagy termelés?*).

Ezek alapján további korpuszbeli keresésre és vizsgálatra érdemesnek tűntek a *semmi* + főnév és az *itt/ott/hol* + NP szerkezetek. A lokatívok esetén ez ugyan számszerűen csak az összes eset felét fedi le, a kiinduló találatok másik fele azonban nehezen körülírható, és sokszor műfajspecifikus. A sajtónyelvben például gyakoriak a



hasonló példák: *Intenzív osztályon a Bielefeld labdarúgója.* A kiinduló kis adathalmazból tehát úgy tűnik, hogy a köznyelvi zéró lokatív létigék jellemzően az említett névmásokkal együtt fordulnak elő.

#### 4.2 További keresések

A mutató névmást tartalmazó lokatív mondatokat a következő alakú reguláris kifejezésekkel kerestem:

```
[word="[Hh]ol"] [word="az"]? [msd="DET|SZN.NOM"] [msd="MN.NOM"] *
[msd="FN.(PS[et]\d.)?NOM"]
```

Erre [1] bármelyik példája illeszkedik. Hasonló kereséseket végeztem az *itt*-re és az *ott*-ra is. A kapott találatokból egy Pythonban írt szűrővel zártam ki azokat, amelyek a keresőkifejezésre ugyan szintén illeszkednek, de tartalmazznak valamilyen (igei vagy névszói) állítmányt. Ilyenek [2] példái.

- [1] a. *Hol a macska?*  
 b. *Hol a fehér macskád?*  
 c. *Hol az a büdös macska?*
- [2] a. *Miért lóg ott az a bádogcsillag?*  
 b. *Ott a víz feneketlen.*

A szűrés után az egyes névmásokra a 3. táblázat szerinti találatszámokat kapjuk.

| Típus | Határozott névelős találatok száma | Határozatlan névelős találatok száma | Összesen |
|-------|------------------------------------|--------------------------------------|----------|
| hol   | 6758                               | 250                                  | 7008     |
| ott   | 6783                               | 563                                  | 7346     |
| itt   | 26574                              | 2181                                 | 28755    |

3. táblázat. Találatok az *itt/ott/hol* típusokra

Ezek a számok önmagukban még nem mondanak sokat, annyi látszik biztosan, hogy a határozatlan névelő előfordulása viszonylag ritka. A többes számú NP-kre is végeztem keresést, erre azonban egyetlen találat volt: *Ha angolul van, hol az adatvédelmi beállítások?*

Az egzisztenciálisok és birtokosok keresőkifejezései a következők voltak:

```
[lemma="[Ss]emmi"] [msd="FN.NOM"]
[lemma="[Ss]emmi"] [msd="FN.PS[et]\d.NOM"]
```

Emellett a korpusz szűrő funkcióját alkalmaztam a finit igét tartalmazó találatok szűrésére. Így 14 444 zéró létigét tartalmazó egzisztenciális, és 17 331 birtokos mondatot kaptam.

## 5 Az eredmények értelmezése

A korpuszból kapott eredményekkel kapcsolatban a legfontosabb kérdés, hogy: mihez viszonyítsuk a számokat, mennyire számít soknak a hét- vagy a tizenkétezer találat? Ennek eldöntésére Clark (1978) zérólétige-tipológiáját alkalmaztam egyetlen nyelven belül, esetünkben a magyar nyelvre.

Clark (1978) szerint a létigés mondatoknak van egy univerzális hierarchiája, amely megmutatja, hogy mennyire valószínű az ige nélküli szerkesztésmód az adott mondattípusban. Az általa felállított sorrend a következő:

- predikatív
- azonosító
- birtoklásmondat2 (be-possessive)
- lokatív
- egzisztenciális
- birtoklásmondat1 (have-possessive)

Az első három mondattípus az 1. táblázat kopula típusának felel meg. A be-possessive alatt az *Ez a macska a szomszédomé* típusú kopulás mondatokat értjük. A hierarchia elsősorban tipológiai megfigyeléseken alapszik, de egy nyelven belül is jelen lehet. Úgy tűnik, a magyarban bizonyos feltételek mellett a hierarchia egy pontjáig kötelező a zéró kopula, utána pedig opcionális. Ez nem egyedülálló eset, például az oroszban is minden jelen idejű predikatív, azonosító és lokatív mondatban testetlen a létige, de az ige nélküli szerkesztésmód opcionálisan a többi mondattípusban is előfordulhat. (I. Kádár 2011:34)

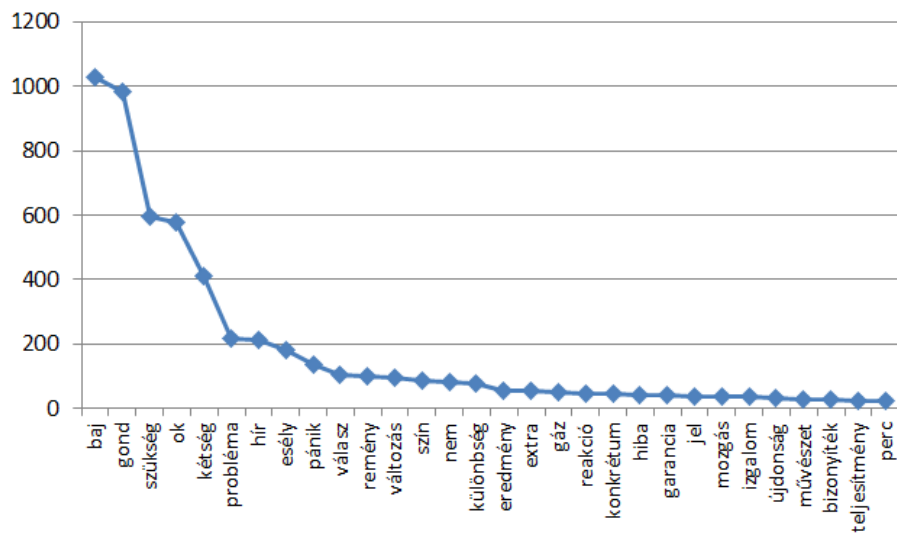
A kutatás alapvető, számítógépes nyelvészeti motivációjához is éppen annak megállapítására van szükség, hogy az egyes mondattípusokban mennyire valószínű a zéró létigés szerkesztésmód. Ez a korpusz segítségével úgy vizsgálható, ha a 4.2-ben ismertetett keresőkifejezésekbe beillesztjük a hiányzó létigét, és az így kapott találatok számát vetjük össze az azonos szerkezetű, ige nélküli mondatok számával. Ezen keresések eredményeit a 4. táblázat tartalmazza.

| Típus           | Testes létigék száma | Zéró létigék száma | Zéró létigék aránya |
|-----------------|----------------------|--------------------|---------------------|
| kopula          | n. a.                | n. a.              | ~100%               |
| hol             | 5062                 | 7008               | 58%                 |
| ott             | 9336                 | 7346               | 44%                 |
| itt             | 10788                | 28755              | 73%                 |
| lokatív össz.   | 25186                | 43109              | 63%                 |
| egzisztenciális | 8710                 | 12213              | 58%                 |
| birtokos        | 13079                | 17331              | 57%                 |

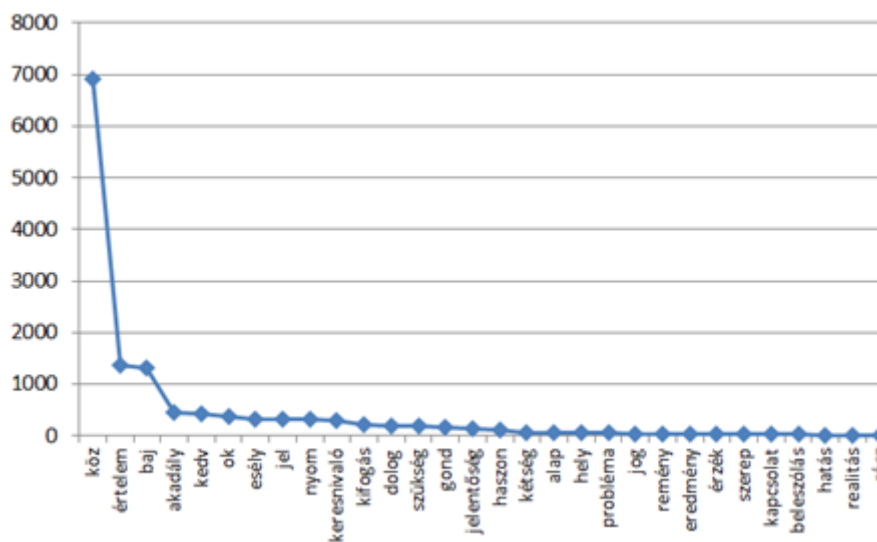
4. táblázat. A testes és zéró létigék eloszlása a vizsgált mondattípusokban

A zéró létige arányát természetesen úgy kell érteni, hogy az adott típusban a megfelelő feltételek mellett (jelen idő, harmadik személy) milyen arányban testetlen a létige. A lokatív, az egzisztenciális és a birtokos mondatok esetén ezek az arányok a Clark-hierarchiának megfelelő sorrendben követik egymást, de a különbségek nem olyan markánsak. Azt is meg kell jegyezni, hogy a vizsgálat nem vonatkozott az összes

lokatív, egzisztenciális és birtokos mondatra, csak azok egy-egy sajátos változatára. Ezeknél a mondat típusoknál tehát a hagyományos kopulához képest jóval több a megszorítás. Az eredmények mégis meglepőek, úgy tűnik ugyanis, hogy a névmások mellől gyakrabban hiányzik a testes létige ezekben az esetekben, mint ahányszor jelen van. Különösen az egzisztenciális és a birtokos mondatok nagyarányú igeelhagyása feltűnő. A jelenség okának felderítésére elvégeztem ezen mondatok alanyainak gyakorisági vizsgálatát, vagyis azt mértem, mely alanyok esetén milyen gyakorisággal jellemző a nincs létige elhagyása. Az eredmények grafikonos szemléltetését az 1. és 2. ábrák tartalmazzák.



1. ábra. Az egzisztenciális mondatok alanyainak gyakorisági megoszlása (30 leggyakoribb szótó)



2. ábra. A birtokos mondatok alanyainak gyakorisági megoszlása (30 leggyakoribb szótó)

A 4. táblázatban bemutatott értékekkel ellentétben az alanyok gyakorisági vizsgálata már mutat látványos különbséget az egzisztenciális és a birtokos mondatok között. A vizsgált zéró létigét tartalmazó egzisztenciális mondatokban összesen 2496-féle szótó fordul elő az alany helyén, a birtokos mondatoknál ezzel szemben mindössze 418, holott a birtokosok eleve majdnem harmincezerrel többen voltak az egzisztenciálisoknál. Amint a 2. ábra görbéjén is látszik, a birtokos mondatoknál az összes gyakoriság nagy részét néhány nagyon gyakori elem teszi ki, az összes előfordulás 90,5%-át lefedi a 15 leggyakoribb szótó (*köz, értelem, baj, akadály, kedv, ok, esély, jel, nyom, keresnivaló, kifogás, dolog, szükség, gond, jelentőség*). Ezzel szemben az egzisztenciálisok első 1500 leggyakoribb szótóve (amelyeknek jelentős része már a legritkább, 1 gyakoriságúak közül kerül ki) is csak az összes eset 89,9%-ának felel meg. Úgy tűnik tehát, hogy az egzisztenciális *semmi* + főnév szerkezet jóval produktívabb a birtokosnál. Mindemellett mindkét esetről elmondható, hogy ezeket a szerkezeteket (vagy legalábbis közülük a gyakoriakat) valószínűleg érdemesebb lenne a szófajcímekre alapozott szabályok helyett lexikálisan, kollokációként kezelni.

## 6 Következtetések

A korpuszadatok szerint az *itt, ott, hol* névmások mellett a létige elhagyása gyakori, de a hagyományos kopulával ellentétben csak egyes számban lehetséges. Amint [3] mutatja, a testes létige nélküli lokatív mondatokban az alany kötelezően névelős, és az adatok szerint a névelő az esetek kb. 93%-ában határozott.

- [1] a. Itt \*(van) Németország.  
 b. Itt \*(van) János kutyája.  
 c. Itt a pénzed.

Bár a korpuszvizsgálat a szórendekre nem tért ki, a mutató névmást tartalmazó lokatív mondatokról általánosságban elmondható az is, hogy a létige csak semleges szórend mellett lehet zéró:

- [2] a. Ott (van) a kedvenc képem.  
 b. A kedvenc képem ott \*(van).

Érdekes (az adatoktól független) észrevétel még, hogy a mutató névmás beillesztésével más helyhatározós, létige nélküli mondatok is „megjavíthatók”:

- [3] a. Az asztalon?(van) a kávé.  
 b. Az asztalon ott (van) a kávé.

Sőt, az is előfordul, hogy az *ott* valójában nem is helyre utal a mondatban, hanem a létige egzisztenciális jelentését erősíti. Egy korpuszbeli példa:

- [4] De nagyobb fejtörésnek ott volt nekik a ló.

Az egzisztenciális jelentés és a lokatív névmások kapcsolata más nyelvekben is ismert jelenség, például az angol *there* és a kínai *you* grammatikalizálódott formái is létezését fejeznek ki. (Hoekstra–Mulder 1990) Az adatok alapján, úgy tűnik, a magyarban is megfigyelhető ez az összefüggés. Ennek részletesebb vizsgálata érdekes elméleti kutatási téma lehet.

Az egzisztenciális és birtokos létigék jellemzően tagadó mondatokban lehetnek zéró alakúak. Ehhez szükséges valamilyen tagadó névmás (pl. *semmi*, *sehol*) jelenléte. A testes létigét nem tartalmazó birtokos mondatokban alanyként megjelenhető névszók köre lexikálisan behatároltnak látszik; úgy tűnik, a zérós szerkesztésmód ennél a típusnál nem annyira produktív.

## 7 Alkalmazás

A kutatás eredményei alapján visszatérhetünk a 2. fejezetben feltett kérdésekre: mi lehet a leghatékonyabb módja a zéró létigék számítógépes elemzésének. Ha a tartalmasfej-megközelítésnél maradunk, akkor azt kell újragondolni, hogy tartalmasnak tekintjük-e a lokatív mondatok létigéjét. Ha nem, mi lesz ezeknek a mondatoknak a feje? Ha igen, mit kezdünk a testes létigét nem tartalmazó lokatív mondatokkal? Még nagyobb probléma az egzisztenciális és birtokos létigék kezelése, hiszen ezek tartalmassága aligha vitatható, mégis láthattuk, hogy sok esetben lehetnek zéró hangalakúak. A dolgozatban vizsgált jelenségeket figyelembe véve tehát egyértelműsége miatt előnyösebbnek tűnik a funkciófej-elemzés.

Ebben a keretben az állítmány, azaz a mondat fejének keresése egy pszicholingvisztikai indíttatású elemzőben, tehát balról jobbra történő beolvasással, a következőképpen képzelhető el az *itt/ott/hol* + NP típusú mondatokban:

Input: *ott/itt/hol* + (egy- számú) NP  
 Ha előtte nem volt „alkalmasabb” állítmány → illessz be egy  
 potenciális zérót: *ott/itt/hol* + (van)? + NP  
 Csatold hozzá a szükséges éleket (határozók, alany)!

Olvasd végig a mondatot, és folytasd az elemzést!  
 Ha találtál másik állítmányt → az addig a zéróhoz kapcsolt éleket  
 (alany, határozók) ő kapja meg! A zéró törlődik.

Példa egy elemzésre:

Input eleje: „Ott egy hangya...”  
 Kezdeti elemzés: Ott (van)? egy hangya...  
 Input vége a.: „... a fűben.”  
 Elemzés a.: a mondat feje a (van).  
 Input vége b.: „... mászik.”  
 Elemzés b.: a mondat feje a mászik.

A zéró létige „előre jóslásának” jelentősége abban áll, hogy a kutatás motivációjául szolgáló, a 2. fejezetben ismertetett elemzőeszköz valós időben, szavanként elemez, így számára nagyon hasznos az ige esetleges hiányának minél korábbi jelzése, még ha ez később „backtrack”-kel is járhat. Az adatok szerint azonban a kérdéses jelenség elég gyakori ahhoz, hogy az elemző jó okkal feltételezzen potenciális zéró létigét az elemzés előbbiekben látott korai pontján.

A vizsgált egzisztenciális és birtokos típusok automatikus elemzésében a *semmi* + főnév szerkezeteket érdemes egy egységként kezelni, és a leggyakoribb ilyen egységeket akár lexikálisan betanítani a gépi eszköznek mint olyan egzisztenciális vagy birtokos jelentésű kollokációkat, amelyek nem feltétlenül kívánják meg a *nincs* létige jelenlétét.

## 8 Összegzés

A zéró létige a magyarban nem csak a hagyományos értelemben vett névszói állítmányt érintő jelenség. Különböző feltételek mellett bármelyik típusú létige opcionálisan elhagyható jelen idő harmadik személyben, azaz a bevezetőben tett hipotézissel ellenkezően az egzisztenciális jelentésű mondatok létigéje is. A Clark-hierarchiával egybehangzóan azonban elmondható, hogy az egzisztenciális és birtokos mondatok esetén a legmegszorítottabb az ígétlen szerkesztésmód lehetősége: ezekben a létige többnyire csak akkor lehet zéró, ha a mondatban tagadó névmás szerepel.

A lokatív mondatok esetén különösen gyakori jelenség a létige elhagyása, ha a mondatban mutató névmás található. Ez arra utal, hogy van valamilyen szemantikai összefüggés a névmások és a létezés jelentéstartalma között.

Mivel a zéró alakot felvehető létigék esetén úgy tűnik, nem mindig lehet egyértelműen állást foglalni a tartalmasság kérdésében, a függőségi elemzésekben érdemesebb lehet a funkciófej-megközelítést követni. Mind az *itt/ott/hol* + NP, mind a *semi* + főnév szerkezetek előfordulása testes létige nélküli mondatokban elég gyakorinak bizonyult ahhoz, hogy ezeket egy elemző programnak valamilyen módon tanítsuk. Erre a 7. fejezetben egy pszicholingvisztikai indíttatású elemző működési elveit követve láttunk példát.

## Hivatkozások

Clark, E. V. 1978. Locationals: Existential, Locative and Possessive Constructions. In: Greenberg, J. H. (szerk.) *Universals of Human Language IV. Syntax*. Stanford: Stanford University Press. 85–126.

- Heggie, L. 1988. *The Syntax of Copular Structures*. PhD-értekezés. Los Angeles: University of Southern California.
- Hoekstra, T., Mulder, R. 1990. Unergatives as copular verbs: Locational and existential predication. *The Linguistic Review*, 7: 1–79.
- Kádár, E. 2011. *A kopula és a nominális mondatok a magyarban*. Budapest: Akadémiai Kiadó.
- Maienborn, C. 2005. A Discourse-based Account of Spanish *ser/estar*. *Linguistics*, 43(1): 155–180.
- Oravecz, Cs., Váradi, T., Sass, B. 2014. The Hungarian Gigaword Corpus. In: Calzolari, N., Choukri K., Declerck T., Loftsson H., Maegaard B., Mariani J., Moreno A., Odijk J., Piperidis S. (szerk.) *Proceedings of Ninth International Conference on Language Resources and Evaluation (LREC 2014)*. Elérhető: [http://www.lrec-conf.org/proceedings/lrec2014/pdf/681\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/681_Paper.pdf). Letöltve: 2017. szeptember 26.
- Prószéky, G., Indig, B. 2015. Magyar szövegek pszicholingvisztikai indíttatású elemzése. *Hungarian Journal of Applied Linguistics*, 2015(1–2): 29–44.
- Simkó, K., Vincze, V. 2016. *Egyszer „van”, hol nem „van”: A létige kezelése függőségi nyelvtanokban*. In: Tanács, A., Varga V., Vincze V. (szerk.) XII. Magyar Számítógépes Nyelvészeti Konferencia (Szeged, 2016. január 21–22.). Szeged: JATEPress. 243–250.
- Stassen, L. 1994. Typology Versus Mythology: The Case of the Zero-Copula. *Nordic Journal of Linguistics*, 17(2): 105–126.

## A beszélői szubjektívitas vizsgálata szentiment- és emóciókorpuszokon<sup>1</sup>

Drávucz Fanni<sup>1</sup>, Szabó Martina Katalin<sup>2, 3, 4</sup>

<sup>1</sup> ELTE BTK Nyelvtudományi Doktori Iskola

<sup>2</sup> Szegedi Tudományegyetem, Szláv Intézet, Orosz Filológiai Tanszék

<sup>3</sup> Precognox Informatikai Kft.

<sup>4</sup> MTA TK „Lendület” RECENS Kutatócsoport

dravucz.fanni@gmail.com

szabo.martina@lit.u-szeged.hu

**Kivonat:** A dolgozatban a nyelvi szubjektívitas és az emóció- és szentimenttartalmak összefüggéseit vizsgáljuk számítógépes nyelvészeti eszközök és módszerek segítségével. A munka elméleti alapvetése az, hogy a bizonytalanságot jelölő elemek a beszélői szubjektívitas indikátorai lehetnek, ezért az elemzéshez a nyelvi bizonytalanságjelölők szólistáit alkalmazzuk. A munka során a bizonytalanságjelölő elemek szótáraival emóció- és szentimentkorpuszokat elemeztünk, és a különböző típusú szemantikai tartalmak gyakorisági összefüggéseit vizsgáljuk. Bár a jelenségekkel a nyelvtechnológiában a szentiment- és az emócióelemzés részben foglalkozik, nincs tudomásunk olyan dolgozatról, amely az emóciókat és a szentimenttartalmakat a nyelvi szubjektívítással összefüggésben, a nyelvtechnológia szempontjából göröcső alá venné. A kutatásunk további haszna, hogy a nyelvi bizonytalanság detekciójára tesz kísérletet, amely feladat az automatikus értékelélemzés kardinális problémája.

### 1 Bevezetés

A jelen dolgozatban a nyelvi szubjektívitas és az emóció- és szentimenttartalmak összefüggéseit vizsgáljuk számítógépes nyelvészeti eszközök és módszerek segítségével. A számítógépes nyelvészetben a szentimentek alatt a szerzői attitűdöt tükröző nyelvi elemeket (pl. „a főnököm remek ember”; Péter 1991), míg az emóciók alatt a szöveg szintjén tetten érhető érzelmeket értjük (pl. „nagyon örültem a főnökömmel”; Szabó et al. 2016a), melyeket a háttérben húzódó kognitív értékelő, illetve emotív funkciók különböztetnek meg.

---

<sup>1</sup> A jelen kutatást Az Emberi Erőforrások Minisztériuma Új Nemzeti Kiválóság Programja, az MTA Társadalomtudományi Kutatóközpont „Lendület” RECENS Kutatócsoportja támogatta, valamint a European Research Council (ERC) az Európai Unió „Horizon 2020” nevű, 648693-es számú kutatási és fejlesztési programja keretében. Az elemzések alapjául szolgáló egyes korpuszok, szótárak és adatok rendelkezésre bocsátásáért a szerzők köszönik Vincze Veronika segítségét.



Korábbi dolgozatunkban (Drávucz–Szabó–Vincze 2017) szentimentszótárakat alkalmazva emóciókorpust, és emóciószótárakat alkalmazva szentimentkorpust elemeztünk. Megállapítottuk, hogy a szentimentkorpuszunkban annotált értékelő tartalmaknak csupán a negyede volt jelezhető az emóciószótárainkkal. Úgy véljük, ez az eredmény egybevág azzal a megközelítési móddal, miszerint a szentimenteknek csupán egy része tekinthető inkább szubjektív (pl. személyes, kontextusfüggő) ítéletnek, és az inkább objektív (pl. személytelen, tényszerű) ítéletek nem tartalmaznak emotív (ily módon emóciószótárral detektálható) elemeket. Mivel az emóciószótárral elemzett szentimentkorpusz (vö. Szabó–Vincze 2015, Szabó et al. 2016a) termékvélemény-szövegeket tartalmaz, valószínűsíthető, hogy a tesztelt termékek értékelői értéktételek megfogalmazására, és nem saját érzelmi viszonyulásuk kifejezésére törekedtek bennük. E tapasztalat alapján úgy döntöttünk, megvizsgáljuk, vajon a szentimentek, az emóciók, valamint a beszélői szubjektivitás milyen összefüggései tárhatóak fel automatikus eszközökkel.

Az interakcióban megjelenő érzelmek, valamint a szubjektív tartalmak kifejezését a pszichológia tudománya részleteiben vizsgálja (vö. pl. Ekman–Friesen 1969, Ekman 2007), azonban a nyelvi elemek emotív illetve szubjektív tartalmának megismerésével a nyelvtudomány területén viszonylag kevés kutatás foglalkozik.

A jelenségekkel a nyelvtechnológiában részben a szentiment- és az emócióelemzés foglalkozik (vö. pl. Liu 2012, Mulcrone 2012, Szabó et al. 2016a, Szabó–Vincze–Morvay 2016), ugyanakkor nincs tudásunk olyan dolgozatról, amely az emóciókat és a szentimenttartalmakat a nyelvi szubjektivitással összefüggésben, a nyelvtechnológia szempontjából venné górcső alá. A kutatásunk további haszna, hogy a nyelvi bizonytalanság detekciójára tesz kísérletet, amely feladat – amint arról később részletesen szólunk (l. fentebb) – az automatikus értékelélemzés kardinális problémája.

## 2 Elméleti háttér

A jelen fejezetben tisztázzuk a kutatásunkban használt legfontosabb fogalmakat, illetve rámutatunk néhány, a terminológiát érintő fontosabb problémára.

A számítógépes értékelélemzés, másképpen *szentimentelemzés* (*sentimentanalysis*) a számítógépes nyelvészet egy részfeladata, amely arra irányul, hogy az értékelő tartalmakat megtalálja a szövegekben, meghatározza ezeknek az értékeknek a típusát, azaz pozitív vagy negatív voltát, valamint, hogy megállapítsa az értékelés tárgyát, tehát azt, hogy az értékelés mire irányul (vö. Liu 2012). A számítógépes érzelemelemzés, másképpen *emócióelemzés* (*emotiondetection* vagy *emotionrecognition*) a szentimentelemzéstől eltérően tipikusan nem az adott dolog nyelvi értékelése, hanem a szövegekben megbúvó – adatközlő által megélt – emóciótartalom kinyerését célozza. Ily módon, bár a két tartalomelemzési feladat, sőt az azokon keresztül vizsgálható kognitív jelenségek is átfedést mutatnak, azok tárgya és célja nem azonos egymással (vö. Drávucz–Szabó–Vincze 2017).

A beszélői szubjektivitást a szentimentelemzés oldaláról megközelítve a következőket kell megemlítenünk: A nemzetközi gyakorlatban a szentimentelemzés feladatát gyakran a szubjektivitáselemzés fogalmával jelölik, e két terminust tulajdonképpen egymás ekvivalenseként használják (vö. pl. Wilson et al. 2005, Liu 2010). Ahogyan Liu (2010) érvel, a szöveges információk két alapvető kategóriára oszthatók: tényyszerűekre és véleményekre. Ez utóbbiakat nevezi a szerző szubjektív

kifejezéseknek, amelyek a megnyilatkozó értékelését vagy érzését tartalmazzák valamely entitás, esemény, vagy azok valamely sajátos vonatkozásában. Mindezek alapján a szerző nem tekinti a szentimentelemzés feladatához tartozónak a tényszerű közlések kivonatolását. Liuhoz (2010) hasonlóan, Wilson és munkatársainak szentimentelemző rendszere (Wilson et al. 2005) is a szubjektív mondatok azonosítását célozza, és dolgozatukban a feladatot szubjektivitáselemzőként említik.

Magunk a magyar nyelvű szövegek szentimentelemzésében nem törekszünk e két információtípus szétválasztására (vö. Szabó 2016a). Meglátásunk szerint a szentimentelemzés eredményének felhasználása szempontjából valójában nincs jelentősége annak, hogy feltárjuk, vajon egy adott értékelő kifejezés inkább tényszerű megállapítás, vagy inkább az adott megnyilatkozó szubjektív megítélése-e. Ebből a szempontból saját elemzési alapelvünk Péter (1991) osztályozási megoldásával áll összhangban, amelyet a nyelvi értékelés vonatkozásában tesz. A szerző megkülönbözteti az értékelés „racionális” és „emocionális” típusát egymástól (vö. Péter 1991: 46), és az előbbire *a főnököm remek ember*, az utóbbira *a habbeton rossz hővezető* mondatokat hozza példaként. Magunk tehát mindkét információtípus kinyerését célozzuk az automatikus értékeléselemzésünk során (vö. Szabó 2016).

A valódi jelentőséggel bíró sajtóságot, és egyben problémát a szentimentelemzésben azt jelenti, hogy hogyan tudjuk kiszűrni azokat a nyelvi tartalmakat, amelyek esetében a megnyilatkozó a közölt információt illetően bizonytalan. A bizonytalanságot jelölő kifejezések automatikus azonosítása napjaink nyelvtechnológiai kutatásainak egyik fontos problémaköre (vö. Vincze 2014a: 99). A szövegek automatikus tartalmi elemzése során ugyanis csupán akkor tudunk megfelelő elemzési eredményt produkálni, ha az elemzőrendszerünk képes a tényszerűen közölt és a bizonytalan nyelvi tartalmakat elkülöníteni egymástól. A szentimentelemzés feladatában ez azt jelenti, hogy elsősorban azokat az értékelő tartalmakat kívánjuk kinyerni, amelyeket a megnyilatkozó tényként közöl. Tekintsük az alábbi példákat!

- [1] A hangminőség jó.
- [2] a. A hangminőség valószínűleg jó.  
b. Jó a hangminőség?  
c. Nem tudom, hogy a hangminőség jó-e.  
d. A hangminőség jó lehet.  
e. Minden bizonnyal jó a hangminőség.

Az [1] alatti, faktív olvasatú példával ellentétben a [2] alattiak nem faktív olvasatúak. Mindegyikük tartalmaz ugyanis valamilyen olyan nyelvi eszközt, amely lehetetlenné teszi az értékelést megfogalmazó szövegrész faktív olvasatát. Másképpen: megakadályozza, hogy az adott szentimentet a megfogalmazó által tényként közölt információként fogadjuk el (vö. Szabó–Vincze 2015: 221–222). Belátható tehát, hogy az automatikus értékeléselemzésben csupán az [1] alatti példa tartalmát volna szerencsés tényszerű értékítéletként feldolgozni, a [2] alatti példák tartalmát nem.

A probléma megoldásával az angol nyelv vonatkozásában több dolgozat is foglalkozik (vö. Farkas et al. 2010, Szarvas 2012). A magyar nyelvet illetően egyetlen kézzel annotált bizonytalansági korpuszról van tudomásunk, valamint az e korpusz feldolgozása alapján született első eredményekről a nyelvi bizonytalanságot jelölő elemek automatikus felismerését illetően (vö. Vincze 2014a) (az ebben a projektben alkalmazott osztályozási megoldásról részletesebben l. lentebb, [3]).

Amint arról a dolgozat bevezetőjében már szóltunk (l. fentebb), korábbi munkánkban (Drávucz–Szabó–Vincze 2017) szentimentszótárak segítségével

emóciókorporust, és emóciószótárak segítségével szentimentkorporust elemeztünk. Megállapítottuk, hogy a kapott eredmények egybevágóak a megállapítással, miszerint a szentimenteknek egy része objektívként klasszifikálható, amellyel összefüggésben azok nem tartalmaznak önreflexív elemeket. Ez a vizsgálati eredmény adta az alapötletét a jelen kutatásnak, amelyben a beszélői szubjektivitás, az emóciók, valamint a szentimentek összefüggéseit igyekszünk feltárni.

A szubjektivitás vizsgálatához a bizonytalanságot jelölő nyelvi elemeket választottuk. Úgy véltük ugyanis, hogy a bizonytalanságot jelölő elemek a beszélői szubjektivitás indikátorai lehetnek, ezért a bizonytalanság és a szentiment- és emóciótartalmak összefüggéseinek megismerésével valamelyest betekintést nyerhetünk a szubjektivitás és a szentiment- és emóciótartalmak összefüggéseibe is.

### 3 A munka módszere és eszközei

A nyelvi bizonytalanság automatikus felismerésére a természetesnyelv-feldolgozás (NLP) eszköztárából a szótárillesztés módszerét alkalmaztuk, amely például a gépi tanulás, a szintaktikai elemzésen alapuló mintaillesztés mellett egyszerűbb információkinyerési módszer. A kutatás során egy bizonytalanságjelölő elemeket tartalmazó szótárt (vö. Vincze 2014b) illesztettünk olyan magyar nyelvű korpuszokra, melyeket korábban kézzel annotáltunk az emóciók, illetve a szentimentek szintjén (vö. Szabó et al. 2016, Szabó–Vincze 2016).

Ahogy arra Vincze (2014) felhívja a figyelmet, a nyelvi bizonytalanságot hagyományosan a mondat szemantikájához szokták kötni, azonban vannak olyan bizonytalanságot jelző nyelvi elemek is, amelyek a megnyilatkozás kontextusában –diskurzusbeli tényezőknek köszönhetően – válnak bizonytalanságjelölővé. Ezzel összefüggésben az alábbi példák nem azonos okból kifolyólag tekinthetők tartalomelemzési szempontból bizonytalannak:

- [3] a. Lehet, hogy esik az eső  
 b. Számos kutató szerint az ezüstkolloid belsőleg sokkal hatásosabb, mint a mesterséges antibiotikumok

Amíg a [3a] alatti példában azonosítani tudunk egy, a közölt tartalom faktivitását törlő nyelvi elemet, addig a [3b] alatti példa esetében ilyen elemet nem találunk; a bizonytalanságot az okozza, hogy semmilyen konkrét, ellenőrizhető információt nem kapunk a közölt tartalom forrására, hitelességére vonatkozóan. Azt látjuk tehát, hogy amíg egyes, bizonytalanságot kifejező mondatokban a bizonytalanság ténye valószínűleg automatikusan is könnyen azonosítható elemekhez kötődik, addig más esetekben az csupán a pragmatika szintjén érhető tetten, ami nyilvánvalóan az automatikus elemzés számára is komoly kihívást jelent (erről még a későbbiekben, az elemzés kapcsán is lesz szó, l. lentebb).

Szarvas és munkatársai (Szarvas et al. 2012), valamint Vincze (2013, 2014) dolgozataikban a szemantikai bizonytalanságnak négy, a diskurzusszintű bizonytalanságnak három fajtájával foglalkoznak részletesen. Magunk a jelen kutatásban erre az osztályozási rendszerre támaszkodunk. Az egyes típusokat az alábbiakban röviden, példák segítségével prezentáljuk.

#### I. Szemantikai bizonytalanság:

- a. **Episztemikus (epistemic)**: Egy proposíció episztemikusan bizonytalannak számít, ha a világtudásunk alapján nem tudjuk eldönteni az adott pillanatban, hogy igaz-e vagy hamis. Pl. *Lehet, hogy esik.*
- b. **Feltételes mondatok (condition)**: Az episztemikus bizonytalanságnál ismertetett sajátság jellemző az ún. hipotetikus bizonytalanságra is, amelyeket közé soroljuk a feltételes mondatokat (**condition**), pl. *Ha esik, itthon maradunk.*
- c. **Vizsgálati bizonytalanság (investigation)**: Nincs konkrét, ellenőrizhető információ a közölt tartalom forrására, hitelességére vonatkozóan, az például vizsgálattal lehet megállapítható. Különösen tudományos cikkekben gyakori, hiszen a kutatási kérdést gyakran a vizsgálati bizonytalanság nyelvi eszközeivel fogalmazzák meg a szerzők. Pl. *A felvétel manipuláltságáról vizsgálatot folytattak.*
- d. **Doxasztikus(doxastic)**: A hiedelmekkel összefüggő, ebből adódó bizonytalanság. Pl. *Azt hiszi, hogy a Föld lapos.*

**II. Diskurzusszintű bizonytalanság** (az itt ismertetett típusokat – magyar nyelvű terminológia híján – angol nyelven nevezzük meg):

- a. **weasel**: az információtartalomhoz nincs egyértelműen forrás rendelve (nem derül ki, kihez köthető az adott információ), vagy pedig hiányzik a közlésből egy fontos és releváns információrészlet, amely szükséges lenne. Pl. *Egyesek szerint inkább megszállást kellene mondani.*
- b. **hedge**: homályossá teszik bizonyos mennyiségek vagy minőségek jelentését. Pl. *A belga lakosság kb. 10%-a él Brüsszelben.*
- c. **peacock**: bizonyítatlan (vagy bizonyíthatatlan) értékeléseket, minősítéseket vagy túlzásokat fejeznek ki. Pl. *Apafi négy évet keserves tatár fogságban töltött.*

A bizonytalanságjelölők szólistái, amelyeket a jelen kutatásban alkalmaztunk:

| Típus           | Elemzés    |
|-----------------|------------|
| condition       | 6          |
| doxastic        | 11         |
| epistemic       | 13         |
| investigation   | 7          |
| hedge           | 37         |
| weasel          | 26         |
| peacock         | 22         |
| <b>Összesen</b> | <b>122</b> |

**1. táblázat.** A bizonytalanságjelölők elemszáma a szótárban, típusonként

Az emóciókorporusz szöveganyagát tévés és mozi témájú blogoldalokról származó, különböző terjedelmű és szerzőségű kritikákból, hírekből, valamint kommentekből állítottuk össze (vö. Szabó et al. 2016). A szentimentkorporuszunkat termékvélemény-szövegek alkotják (vö. Szabó–Vincze 2016). A korpuszokról néhány alapvető információt az alábbi táblázatban közlünk.

|                 | Szentimentkorpusz  | Emóciókorpusz   |
|-----------------|--|---|
| <b>Méret</b>    | 17 000 mondat<br>250 000 token   | 794 mondat<br>7205 token  |
| <b>Címkézés</b> | bináris: pozitív/negatív   | többszörös: 7 emóció  |
| <b>Domain</b>   | termékvélemények   | blog és komment   |
| <b>Forrás</b>   | divany.hu  | tévés és filmes blogok  |
| <b>Példák</b>   | <i>A nem kávéfüggőknek bejött<br/>Omlós, finom, van benne valami<br/>nagyszerű.<br/>Ez a konkrét zacskó senkinek nem<br/>ízlett, túl száraz, túl pörkölt és<br/>túl sárga.</i> | <i>Tesóm rámrúgta az ajtót már fél<br/>kilenckor nem örültem annyira...<br/>Közben persze egy csomó olyan<br/>cuccot találtam, amit fogalmam<br/>sincs, mikor táraztam be... :D<br/>Olyan csoda cuccokat találtam.<br/>:)</i> |

2. táblázat. A felhasznált korpuszok alapvető adatai

A vizsgálatot a következőképpen végeztük el: először a bizonytalanságot jelölő elemek hét listáját illesztettük a szentiment- és az emóciókorpuszra, majd a kapott eredmények alapján megvizsgáltuk a különböző típusú bizonytalanságjelölők, valamint a különböző szentimentértékek és emóciók összefüggéseit.

## 4 Eredmények, következtetések

A vizsgálat legfontosabb feltételezése az volt, hogy a nyelvi bizonytalanság szoros összefüggést mutat a beszélői szubjektivitással, ily módon az emóciófragmentumokban magasabb frekvenciával fognak szerepelni, mint a szentimentfragmentumokban. Az elemzési eredményeket a jelen fejezetben ismertetjük.

### 4.1 Az emóciókorpusz elemzése a bizonytalanságjelölők szólistáival

Elsőként az emóciókorpuszban a valamely emóció címkéjével ellátott fragmentumok és bizonytalanságjelölő elemek illeszkedésének a gyakoriságát mutatjuk be.

Az emóciókorpuszunk összesen 794 fragmentumot és 7205 tokent (központozással) tartalmaznak. Ebből összesen 330 fragmentumban találtunk olyan kifejezést, amely bizonytalanság meglétére utal. Ez az összes token 4,6%-a, illetve az összes fragmentum 41%-a. A 3. táblázat közli a vizsgálat részletes eredményeit. A táblázatban a százalékos értékek azt mutatják, hogy a különböző típusú bizonytalanságjelölők hogyan oszlanak meg a különböző emóciófragmentumok között.

| Típus            | Össz. | Emóciófragmentum |            |            |         |            |            |            |
|------------------|-------|------------------|------------|------------|---------|------------|------------|------------|
|                  |       | düh              | feszültség | undor      | félelem | öröm       | bánat      | meglepetés |
| condi-<br>tional | 28    | 2<br>7,1%        | 4<br>14,3% | 4<br>14,3% | 0       | 6<br>21,4% | 8<br>28,6% | 4<br>14,3% |

|               |            |              |              |             |             |              |              |              |
|---------------|------------|--------------|--------------|-------------|-------------|--------------|--------------|--------------|
| doxastic      | 48         | 4            | 8            | 0           | 4           | 16           | 8            | 8            |
|               |            | 8,3%         | 16,7%        |             | 8,3%        | 33,3%        | 16,7%        | 16,7%        |
| epistemic     | 16         | 2            | 2            | 0           | 2           | 2            | 6            | 2            |
|               |            | 12,5%        | 12,5%        |             | 12,5%       | 12,5%        | 37,5%        | 12,5%        |
| investigation | 2          | 0            | 0            | 0           | 0           | 2            | 0            | 0            |
|               |            |              |              |             |             | 100%         |              |              |
| hedge         | 94         | 12           | 22           | 2           | 2           | 22           | 28           | 6            |
|               |            | 12,8%        | 23,4%        | 2,1%        | 2,12%       | 23,4%        | 29,8%        | 6,4%         |
| weasel        | 110        | 14           | 18           | 2           | 2           | 38           | 22           | 14           |
|               |            | 12,7%        | 16,4%        | 1,8%        | 1,8%        | 34,5%        | 20,0%        | 12,7%        |
| peacock       | 32         | 2            | 4            | 2           | 2           | 14           | 6            | 2            |
|               |            | 6,3%         | 12,5%        | 6,3%        | 6,3%        | 43,8%        | 18,8%        | 6,3%         |
| <b>Össz.</b>  | <b>330</b> | <b>36</b>    | <b>58</b>    | <b>10</b>   | <b>12</b>   | <b>100</b>   | <b>78</b>    | <b>36</b>    |
|               |            | <b>10,9%</b> | <b>17,6%</b> | <b>3,0%</b> | <b>3,6%</b> | <b>30,3%</b> | <b>23,6%</b> | <b>10,9%</b> |

**3. táblázat.** A nyelvi bizonytalanság különböző típusú elemeinek gyakorisága a 7 emóciónál (felül a szemantikai, alul a diskurzusszintű bizonytalanságjelölők)

A 3. táblázat adatai alapján a következő legfontosabb megállapításokat tehetjük: Az emóciófragmentumok közül a legkevesebb bizonytalanságjelölő az undor, a legtöbb pedig az öröm érzelemnél jelent meg. A legtöbb bizonytalanságjelölő elem az öröm mellett a bánat érzelmek kategóriában jelent meg: majdnem minden harmadik bizonytalanságjelölő elemet az öröm, minden negyediket a bánat emóciótaggal annotált fragmentumban azonosítottunk. A bizonytalanságjelölők a tipikusan impulzív érzelmek (düh, undor, meglepődés) esetében jelentek meg ritkábban.

Ha a különböző érzelmek polaritását nézzük, tehát azt, hogy pozitív vagy negatív-e az adott emóció, akkor a bizonytalanságjelölők megoszlását tekintve a negatív polaritású emóciók vannak túlsúlyban.

A fejezet további részében áttekintjük a különböző típusú bizonytalanságjelölő elemek összefüggéseit a különböző emóció típusokkal.

A bizonytalanságjelölők két fő típusa közül a diskurzusszintű bizonytalanságra utaló elemek fordultak elő nagyobb arányban.

A szemantikai bizonytalanságjelölők közül a doxasztikus, azaz a hiedelmekkel összefüggő típus fordult elő az emóciófragmentumokban a leggyakrabban, és az emóció típusok közül a legnagyobb frekvenciával az öröm esetében szerepelt, pl. (A példák esetében az emotív elemet vastagon szedjük, a bizonytalanságjelölőt pedig aláhúzással jelöljük.)

[4] Kate még szerintem üvöltözne tovább

Módszerünk episztemikus típusú bizonytalanságjelölőket a bánat emóció fragmentumaiban azonosított kiemelkedően gyakran, pl.

[5] Laci Vilibának **panasz** **codik** a félresikerült majdnem lánykérés miatt.

A vizsgálati bizonytalanságra utaló elem összesen két találata jól mutatja, hogy a jelenség a tudományos szövegek jellemzője, és az emóciókorpuszban feldolgozott szövegtípusra nem jellemző.

A diskurzusszintű bizonytalanságjelölő elemekről elmondható, hogy a weasel szerepelt a leggyakrabban az emóciófragmentumokban, és őt a hedge típusú bizonytalanságjelölő követte. A weasel típusú, azaz a forrás bizonytalanságát jelölő elemek legmagasabb számban az öröm emóciónál jelentkeztek, pl.

[6] egy kis szerelmes évődésbe kezdenek

Bizonytalanságjelölőket a félelemnél és az undornál ritkán azonosítottunk. A hedge típusú – mennyiségre és minőségre utaló bizonytalanságjelölő – elemek a bánat emóciónál voltak a leggyakoribbak, pl.

[7] sajnos összességében kevés olyan emlékezetes, kreatív ötletet integrált a forgatókönyvbe, amitől most egy közepesnél jobb műfajparódiáról beszélhetnénk

A peacock típusú, bizonyíthatatlan értékelést, minősítést vagy túlzást jelölő elemek az öröm emóciónál mutattak kiemelkedően gyakori előfordulást, pl.

[8] az egyik legnagyobb élményből maradsz ki.

#### 4.2 A szentimentkorpusz elemzése a bizonytalanságjelölők szólistáival

A szentimentkorpuszösszesen 251 202 tokent tartalmaz a 15 675 fragmentumban, melyek polaritása közel kiegyenlített: 8465 (54%) negatív, valamint 7210 (46%) pozitív polaritású. A szentimentkorpuszon módszerünk a bizonytalanságjelölők szótárával összesen 4553 esetben állapított meg bizonytalanságot, ez a tokenek 1,8%-a, illetve a fragmentumok 29%-a (l. részletesen a 4. táblázatot).

| Típus           | Szentimentfragmentum         |                              | Össz.                        |
|-----------------|------------------------------|------------------------------|------------------------------|
|                 | negatív                      | pozitív                      |                              |
| conditional     | 163                          | 125                          | 288                          |
| doxastic        | 262                          | 160                          | 422                          |
| epistemic       | 154                          | 77                           | 231                          |
| investigation   | 0                            | 0                            | 0                            |
| hedge           | 1 521                        | 1 207                        | 2 728                        |
| weasel          | 212                          | 96                           | 308                          |
| peacock         | 254                          | 322                          | 576                          |
| <b>Összesen</b> | <b>2 566</b><br><b>30,3%</b> | <b>1 987</b><br><b>27,6%</b> | <b>4 553</b><br><b>29,0%</b> |

**4. táblázat.** A nyelvi bizonytalanság különböző típusú elemeinek gyakorisága a szentimentfragmentumokban (felül a szemantikai, alul a diskurzusszintű bizonytalanságjelölők)

A különböző típusú bizonytalanságjelölők valamelyest eltérő megoszlást mutatnak a pozitív és a negatív szentimentfragmentumokban, és összességében jellemzően a negatív fragmentumokban gyakoribbak.

Amennyiben egyesével megvizsgáljuk a bizonytalanságjelölők különböző típusainak az előfordulását, a következő megállapításokat tehetjük: A

bizonytalanságjelölők két fő típusa közül – akárcsak az emóciókorporusz esetében – itt is a diskurzusszintű bizonytalanságra utaló elemek fordultak elő nagyobb arányban. Ami az egyes típusokat illeti, meg kell említeni, hogy nem találtunk a vizsgálati bizonytalanságra utaló elemet. Nyilvánvaló, hogy ez a jelenség nem jellemző a termékvélemény-szövegekre, amelyekből a szentimentkorporuszt építettük. Mindhárom további, szemantikai típus a negatív fragmentumokban volt a gyakoribb.

A diskurzusszintű bizonytalanságjelölő elemek közül – meglepő módon az emóciókorporuson tapasztaltaktól teljesen eltérően – a legmagasabb arányban a hedge típusú, mennyiségre és minőségre utaló bizonytalanságjelölő elem szerepelt. (A szentimenteket vastagon szedtük, a bizonytalanságjelölőt aláhúzással jelöltük.)

[10] a csomagolása elég **kecsegtető**

Ez a típus az emóciókorporusz fragmentumaiban a legkevésbé frekvenciát diskurzusszintű bizonytalanságjelölő volt. Mindemellett a weasel típusú elemek fordultak elő legritkábban a diskurzusszintűek közül, amelyek pedig az emóciókorporuszban a legnagyobb gyakorisággal szerepeltek ebből a típusból, például:

[11] Ez mégis valami **drágább** kategóriás lehet az olcsók között.

## 5 Összegzés

A dolgozatban a szubjektivitás, valamint az emóciók és a szentimentek összefüggéseit vizsgáltuk úgy, hogy a nyelvi bizonytalanságjelölő elemek szótáraival emóció- és szentimentkorporuszokat elemeztünk, és különböző típusú szemantikai tartalmak gyakorisági összefüggéseit kutattuk.

A vizsgálat eredményei alapján a következő legfontosabb megállapításokat tehetjük: Amíg az emóciókorporuszunkban az összes fragmentum 41%-ában (összesen 330 fragmentumban) találtunk olyan kifejezést, amely bizonytalanság meglétére utal, addig a szentimentkorporuszunkban valamivel kevesebben, a fragmentumoknak csupán a 36%-ában (összesen 5 600 fragmentumban). Más adatokkal, amíg az emóciókorporuszban az összes token 4,6%-át, addig a szentimentkorporuszban ugyanennek csupán az 1,8%-át azonosítottuk bizonytalanságjelölőként. Mindez egybehangzónak látszik kezdeti hipotézisünkkel, miszerint a bizonytalanságjelölők szorosabb kapcsolatban állnak az érzelmekkel, mint a nyelvi értékeléssel, és ez gyakorisági megoszlásukban is megmutatkozik. Amennyiben helytálló az a feltételezésünk, hogy a bizonytalanságjelölők a nyelvi szubjektivitás egyfajta indikátorai, úgy a fentebbi vizsgálati eredmények összhangban állnak azzal a feltételezésünkkel is, miszerint a beszélői szubjektivitás is inkább az emóciókkal, mint a szentimentekkel mutat összefüggést.

Mind az emóció, mind a szentimentkorporusz elemzésekor azt láttuk, hogy a bizonytalanságjelölők a negatív polaritású emóciókban voltak túlsúlyban. Ez utóbbi tapasztalat véleményünk szerint azt igazolja, hogy a negatív véleményünket árnyaltabban, kevésbé direkt módon fejezzük ki.

Az emóciókorporuszban a legtöbb bizonytalanságjelölő elem az öröm és a bánat érzelmek kategóriában jelent meg. A bizonytalanságjelölők ritkábban jelentek meg a tipikusan impulzív érzelmek (düh, undor, meglepődés) esetében. A szemantikai bizonytalanságjelölők közül az episztemikus és hipotetikus típus a bánat, míg a doxasztikus az öröm emóció esetében volt a leggyakoribb. Ugyancsak az



emóciókorporuszban a diskurzusszintű bizonytalanságjelölő elemek közül a weasel és a peacock típusúak az öröm emóciónál, a hedge típusúak pedig a bánat emóciónál a leggyakoribbak.

A vizsgálati bizonytalanságra utaló elemek mindkét korpuszunkban rendkívül ritkák voltak. Nyilvánvaló, hogy ez a nyelvi eszközkészlet nem jellemző a két korpuszban feldolgozott szövegtípusok egyikére sem.

A diskurzusszintű bizonytalanságjelölő elemek közül a szentimentkorporuszban – az emóciókorporuszon tapasztaltaktól eltérően – a legmagasabb arányban a hedge típusú (mennyiségre és minőségre utaló) bizonytalanságjelölő elem szerepelt, amely az emóciókorporuszban összességében is a második leggyakoribb típus volt. Mindemellett a weasel típusú elemek fordultak elő legritkábban a diskurzusszintűek közül, amelyek pedig az emóciókorporuszban a legnagyobb gyakorisággal szerepeltek ebből a típusból. A szentimentkorporuszban a diskurzusszintű bizonytalanságjelölő elemek közül a peacock a pozitív, a hedge és a weasel a negatív fragmentumokban volt gyakoribb.

A kutatás további tervezett lépéseként indokoltnak tartjuk jelen tapasztalatok alapján a bizonytalanságjelölő szótár, illetve az illesztési módszer továbbfejlesztését. Ehhez többek között más, egyéb szövegtípusokat is szerepeltető korpuszok elemzése szükséges. A bizonytalanság, érzelmek és szentimentek viszonyát a nyelvben a bizonytalanságkorporusz (Vincze 2014a, 2014b) emóció- és a szentimentszótárakkal való elemzésével tervezzük tovább vizsgálni.

## Irodalom

- Drávucz, F., Szabó M. K., Vincze V. 2017. Szentiment- és emóciószótárak eredményességének mérése emóció- és szentimentkorporuszokon. In: Vincze, V. (szerk.) *XIII. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY 2017*. Szeged: JATEPress. 228–239.
- Ekman, P., Friesen, W. V. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1: 49–98.
- Ekman, P. 2007. *Emotions Revealed*. New York: Times Books. Elérhető: <https://zscalarts.files.wordpress.com/2014/01/emotions-revealed-by-paul-ekman1.pdf>. Letöltve: 2017. október 6.
- Farkas, R., Vincze V., Móra Gv., Csirik J., Szarvas Gv. 2010. The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. In: *Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL-2010)*. Association for Computational Linguistics: Uppsala. 1–12.
- Liu, B. 2010. Sentiment analysis and subjectivity. In: Indurkha, N., Damerau, F. J. (szerk.) *Handbook of Natural Language Processing*. Second Edition. Chapman and Hall/CRC. 627–666. Elérhető: <https://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf>. Letöltve: 2017. október 6.
- Liu, B. 2012. *Sentiment Analysis and Opinion Mining*. Kézirat. Elérhető: <http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>. Letöltve: 2017. október 6.
- Mulcrone, K. 2012. Detecting Emotion in Text. Elhangzott: UMM CSci Senior Seminar Conference, University of Minnesota, Morris, 2012. április 28.
- Péter, M. 1991. *A nyelvi érzelmkifejezés eszközei és módjai*. Budapest: Tankönyvkiadó Vállalat.
- Szabó, M. K., Vincze, V. 2015. Egy magyar nyelvű szentimentkorporusz létrehozásának tapasztalatai. In: Tanács, A., Varga V., Vincze V. (szerk.) *XI. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY2015*. Szeged: JATEPress. 219–226.
- Szabó M. K., Vincze V., Simkó K., Varga V., Hangya V. 2016a. A Hungarian Sentiment Corpus Manually Annotated at Aspect Level. In: Calzolari, N., Choukri K., Declerck T., Gogoi 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). 2873–2878.

- Szabó M. K., Vincze V., Morvay G. 2016b. Magyar nyelvű szövegek emócióelemzésének elméleti nyelvészeti és nyelvtechnológiai problémái. In: Reményi, A. Á., Sárdi Cs., Tóth Zs. (szerk.) *Távlatok a mai magyar alkalmazott nyelvészetben*. Budapest: Tinta. 282–292.
- Szarvas, Gy., Vincze V., Farkas R., Móra Gy., Gurevych I. 2012. Cross-genre and cross-domain detection of semantic uncertainty. *Computational Linguistics*, 38: 335–367.
- Vincze, V. 2013. Weasels, hedges and peacocks: Discourse-level uncertainty in wikipedia articles. In: *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. Nagoya: Asian Federation of Natural Language Processing. 383–391.
- Vincze, V. 2014a. Bizonytalanságot jelölő kifejezések azonosítása magyar nyelvű szövegekben. In: Tanács, A., Varga V., Vincze V. (szerk.) *X. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY2014*. Szeged: JATEPress. 99–108.
- Vincze, V. 2014b. Uncertainty Detection in Hungarian Texts. In: *Proceedings of COLING 2014*. Dublin. 1844–1853. Elérhető: <http://www.aclweb.org/anthology/C/C14/C14-1174.pdf>. Letöltve: 2017. október 6.
- Wilson, T., Hoffmann P., Somasundaran S., Kessler J., Wiebe J., Choi Y. et al. 2005. Opinion Finder: A system for subjectivity analysis. In: *Proceedings of HLT/EMNLP on Interactive Demonstrations*. Stroudsburg, PA: Association for Computational Linguistics. Elérhető: <http://people.cs.pitt.edu/~swapna/papers/OpinionFinder-extendedabstract.pdf>. Letöltve: 2017. október 6.

## The Generic Structure of Chinese MA Students' Thesis Acknowledgments: A Comparative Study across Contexts

Jing Feng

Language Pedagogy PhD programme, Eötvös Loránd University, Budapest  
seekjessica@hotmail.com

**Abstract:** In the past two decades, the generic structures of acknowledgments in theses and dissertations have been studied in various contexts (Giannoni 2002; Hyland 2004; Al-Ali 2010; Yang 2012; Jalilifar–Mohammad 2014), and it has been revealed that the genre of acknowledgments has been shaped by the influence of disciplinary, social, cultural and contextual variables. However, there is relatively little literature on the cross-context analysis. On the basis of Hyland's (2004) move model, the present study involves the genre-based comparative analysis of 20 MA theses acknowledgments written by Chinese students, 10 of them studying for their degree in Chinese universities and the other 10 in American universities. Focusing on the generic structures and linguistic features of these acknowledgments, this study supports the idea that the Chinese students in the two contexts dominantly applied the obligatory thanking move and optional reflecting move, without the announcing move. However, the slight divergences occurring in the construction of moves/step, choices of addressees and expressions of gratitude could possibly be explained by contextual factors including academic conventions and contextual values.

### 1 Introduction

Since the 1980s, the increasing application of genre theory to language research has established genre analysis as a special field in applied linguistics, and numerous publications on genre analysis have appeared in the academic domain. Under the influence of Swales' (1990) pioneering work on the CARS (Create A Research Space) model proposed for the analysis of the introductions of research articles, the macro-structure of various academic genres has been studied in the field of English for Academic purposes (EAP), including abstracts, introductions, discussions, results, conclusions, research paper titles etc. (Zhang–Jiang 2010). However, as to the genre of acknowledgments in the academic setting, scholars pay very little attention to it although it is of great scholarly importance in academic writing with a high frequency of appearance (Hyland 2003).

Described by Hyland (2003: 243) as a “Cinderella genre”, which is neither strictly academic nor entirely personal, the genre of acknowledgments in theses and dissertations is more than a simple list of addressees who deserve gratitude; it plays a crucial role in creating “both a professional and personal identity” and “affirming [the

writer's] commitment to values such as modesty, generosity, and gratitude which are prized by academic communities" (Hyland 2003: 264). For non-native English writers, it can be really difficult to express their gratitude adequately and appropriately in their theses and dissertations. According to Wen's (2004) study, the two most common problems affecting Chinese students' acknowledgments in theses and dissertations are: 1) over-general and abstract acknowledgments which lack explicit descriptions of assistance received; 2) wholly plagiarized or barely modified acknowledgments taken from other people's work, which might cause addressees' dissatisfaction. Besides, there is an increasing number of Chinese students studying abroad in order to acquire academic degrees, who are challenged by language proficiency and cultural and contextual differences when composing their theses and dissertations in English. Based on Cheng's (2012) research, a number of issues with considerable sociopragmatic relevance could be listed, including the following: how students in various cultural and social contexts express their gratitude, and how socio-cultural factors such as cultural differences, academic conventions and the use of English as an L1 or ESL or EFL in language contexts might affect the generic and linguistic features of acknowledgment writing.

So far the literature on the genre of acknowledgments in theses and dissertations has mainly focused on generic structure and the lexical-grammatical realization of acknowledgments written by one particular ethnic group in one social context by applying Hyland (2004)'s "three-tier model" (Hyland 2004; Hyland-Tse 2004; Al-Ali 2006, 2010; Lei 2006; Zhang-Jiang 2010). Other comparative studies from cross-cultural, cross-genre, or even cross-linguistic perspectives have also been conducted by researchers to compare acknowledgments in terms of their generic structures and linguistic features (Cheng 2012; Karakaş 2010; Kuhl-Rezaei 2014; Qin 2012; Mohammadi 2013; Jalilifar-Mohammadi 2014). However, there is still relatively little research involving the cross-contextual analysis of acknowledgments composed by writers who share similar cultural backgrounds but work in two different academic environments and language contexts. Consequently, three research questions will be explored in this present study:

- (1) What generic structure characterizes English thesis acknowledgements written by Chinese MA postgraduates studying in Chinese universities and American universities?
- (2) Are there any differences between Chinese MA students' construction of moves and steps and choices of linguistic elements in the two different contexts?
- (3) If there are, what contextual factors influence students' acknowledgement-writing habits?

On the basis of the research findings, possible pedagogical implications could be outlined to help Chinese students' acknowledgements writing.

## 2 Theoretical background

The concept of genre was introduced to linguistics in the 1970s and since then many scholars and researchers have defined genre in different ways. Swales (1990) stated that genres comprise "a class of communicative events", and that the members who take part in these events share the same communicative purposes. Taking into consideration psychological and sociological factors, Bhatia (1993) further pointed out that being

highly structured and conventionalized, genres are also constrained by the choice of certain lexis and moves, which are exploited by members of the discourse community, and that genres are connected with the organization of culture and social purposes around language. Genre analysis actually provides an explicit and systematic way to figure out how a particular genre is structured and distinguished in social and cultural contexts and finally achieves the goal of communicative purpose (Hyland 2004: 18).

Genre analysis can describe both linguistic features at a micro-level and generic structure at a macro-level to interpret how a particular genre is constructed and how its communicative purposes are realized through language uses. A great many studies on genre analysis at macro-level have been conducted to investigate moves and steps in academic discourses across different disciplines. The majority of the studies either adopted or modified the most influential CARS model, proposed by Swales (1999) to analyze research articles, to explore the move structure of various academic genres, including research paper titles, abstracts, introductions, discussions and conclusions (for example, Hyland 2000; Bunton 2002; Martin 2003; Swales 1990, 2004; Yang–Allison 2003, etc.). Micro-level genre analysis could emphasize linguistic features, such as tense and aspect, modality, adjectives and reporting verbs, in order to analyze the grammatical and stylistic aspects of particular genres (for instance, Vassileva 2001; Soler 2002; Flowerdew 2003, etc.).

Defined as a ‘Cinderella genre’ by Hyland (2003), as mentioned above, acknowledgements in academic settings have attracted the attention of some genre analysts. Giannoni (2002) first investigated acknowledgements by examining 100 samples in English and Italian research journals from six disciplines with reference to move structural pattern, to reveal how writers organize and express their gratitude. Two moves of acknowledgments were put forward: the *introductory move* (Framing) and the *main move* (Credit mapping). The greatest contribution to the generic study on acknowledgements was made by Hyland (2004), who investigated 240 MA and PhD theses and dissertations written by non-native English speaking students at Hong Kong universities. In order to explore how these non-native English writers express their gratitude in acknowledgment writing, he came up with a proposal for a three-tier structure, consisting of a main *Thanking move* framed by an optional *Reflecting move* and *Announcing move*, which has become a reference model for later similar research on acknowledgements in academic settings (Table 1). He also pointed out that moves and steps generally followed the same sequence, although there still might be many recursions of steps, especially of those expressing gratitude for academic and moral support. In a later study, Hyland and Tse (2004) further investigated the lexical-grammatical patterns that were employed to realize the component moves of the same acknowledgement samples. Hyland’s set of moves and steps can be summarized as follows:

| Moves/steps                          | Description  |
|--------------------------------------|--|
| 1. Reflecting Move (optional)        | Introspective comment on the writer's research experience            |
| 2. Thanking Move (obligatory)        | Mapping credit to individuals and institutions                       |
| 2.1 Presenting participants          | Introducing those to be thanked                                      |
| 2.2 Thanking for academic assistance | Thanks for intellectual support, ideas, analyses, and feedback, etc. |
| 2.3 Thanking for resources           | Thanks for data access, clerical, technical and financial support    |
| 2.4 Thanking for moral support       | Thanks for encouragement, friendship, sympathy, patience etc.        |
| 3. Announcing Move (optional)        | Public statement of responsibility and inspiration                   |
| 3.1 Accepting responsibility         | An assertion of authorial responsibility for flaws and errors        |
| 3.2 Dedicating the thesis            | A formal dedication of the thesis to one or more individuals         |

**Table 1.** Three-tier structure of acknowledgments (Hyland 2004)

Under the influence of Hyland (2004) and Hyland–Tse (2004)'s research, some of the existing literature on acknowledgements has focused on analyzing acknowledgements written by one particular ethnic group in one social context. In Arab settings, Al-Ali (2006, 2010) identified a new step, *Thanking Allah (God)*, by investigating 100 Arab PhD students' dissertation acknowledgements. By emphasizing the influence of religious beliefs and local academic and social conventions, he reached the conclusion that acknowledgements could reveal how writers perceive different selves in their interaction with their peers according to culture-specific preferred conventions. Some researchers have conducted cross-cultural studies by comparing acknowledgements in theses and dissertations written by native speakers (NS) and non-native speakers (NNS) of English in two different social contexts. In Taiwan, Cheng (2012) compared 60 thesis acknowledgements written in English by Taiwanese and North American MA students and found that the two groups vary in the arrangement of addressees and the choice of thanking strategies for various addressees, reflecting different cultural perceptions of the expression of gratitude. In a Turkish context, Karakaş (2010) compared native speakers of Turkish and of American English to analyze the pragmatic and discourse strategies used in thesis acknowledgment writing and stressed that the social and cultural background of writers still affects the length of acknowledgments, language choices and move structures. Lasaky (2011) focused on non-native (Iranian) and native English speakers' thesis acknowledgements and identified a similar *Thanking Allah (God)* to the one found in Al-Ali's research, due to the similar context. He also explained the absence of the *Reflecting move* and the *Announcing move* in terms of cultural difference.

Other comparative studies have been conducted either by combining cross-cultural and cross-genre studies to compare acknowledgements in academic writing, including theses and dissertations, research articles and textbooks written by English native and

non-native (Iranian) speakers (Kuhi–Rezaei 2014), or by combining cross-cultural and cross-linguistic studies to compare acknowledgements in theses and dissertations written in English and other languages such as Chinese and Persian (Qin 2012; Mohammadi 2013; Jalilifar–Mohammadi 2014). Cross-contextual analysis has also attracted the attention of several researchers who paid particular attention to contextual factors. Zhang–Jiang (2010) compared the similarities and differences of generic structure, gratitude expressions and modifiers used in MA and PhD thesis acknowledgements by analyzing 40 acknowledgments written by Chinese speakers at Chinese universities, with reference to Hyland’s (2004) Hong Kong students’ model. Yang (2012) further compared thesis acknowledgements written by Mainland Chinese speakers, Hong Kong Chinese speakers and Taiwanese Chinese speakers and found that a similar three-tier structure was employed by writers in all three contexts, although some divergence appeared in move construction and language choices. Yang (2013) continued the cross-context genre analysis of thesis acknowledgements by focusing on Taiwanese authors in EFL and ESL contexts and found that academic conventions, institutional preferences and the language context, together with socio-cultural factors, affected their construction of moves/steps and their choice of linguistic elements.

In summary, a large number of existing studies have explored a single ethnic group’s thesis acknowledgements, and compared acknowledgment written by native speakers and non-native speakers, and cross-cultural, cross-contextual or even cross-genre analysis have been conducted in the studies discussed above to confirm the social and cultural factors that influence the move structure and linguistic realization of acknowledgements. However, there is still little research involving the comparative study of thesis acknowledgements written by writers who come from the same cultural background and speak the same mother language, but study in different contexts. Especially for Chinese-speaking students from mainland China who choose to study in American universities for MA or PhD degrees, writing theses and dissertations in English has been a huge challenge, and different contextual factors such as English status and academic conventions could influence their way of expressing their gratitude in academic writing. The present pilot study aims to explore the generic structure of Chinese MA students’ thesis acknowledgements by comparing move structure and linguistic choices across two contexts, namely Chinese universities and American universities, and analyzing the contextual factors that may affect or trigger cross-contextual variations. In order to be able to do so, the following research questions are addressed in this study: 1) What generic structure characterizes English theses acknowledgements written by Chinese MA students in Chinese and American universities? 2) Are there any differences between Chinese MA students’ construction of moves and steps and choices of linguistic elements in two contexts? 3) If there are, what contextual factors influence students’ acknowledgement writing habits?

### 3 Methods

In order to answer these research questions, a small corpus was built and genre analysis was applied to explore the generic structure of acknowledgements, acknowledgees and expressions of gratitude. A sample analysis is also provided in the Appendix as an example to help with possible replication.

### 3.1 Corpus

The pilot study uses a corpus consisting of 10 MA thesis acknowledgements written by Chinese postgraduates majoring in Linguistics and Applied Linguistics in Chinese universities, and 10 written by Chinese postgraduates majoring in Education and Applied Linguistics in American universities. Although it is a small corpus, 10 samples in each group can be regarded as sufficient for a pilot study. In order to make a reasonable comparison, very similar disciplines were selected between Chinese and American universities. All the theses were written by MA students majoring in English Applied Linguistics or TESOL. MA Theses written at Chinese universities were downloaded electronically from a Chinese dissertation database<sup>i</sup> and MA theses written at American universities were downloaded electronically from the website of the University of Kansas.<sup>ii</sup> To ensure that the theses from the American university were written by Chinese mainland students, rather than Hong Kong or Taiwanese Chinese speakers, two strategies were adopted to confirm the writers' nationality: checking the name spelling, and reading the content of the abstracts and acknowledgements.

### 3.2 Procedures of analysis

Altogether 20 acknowledgements from Chinese universities (CU) and the American university (AU) were collected to establish two separate corpora and numbered CU001-010 and AU001-010. The generic structures of the two corpora were analyzed separately on the basis of Hyland (2004)'s model; another researcher was trained to co-code the sample and inter-coder agreement was checked to ensure the reliability of the analysis. 5 theses in each group were chosen at random and analyzed. The degree of agreement regarding moves reached 98.3% and regarding steps reached 86.5%. Based on the genre analysis, the generic structures of acknowledgements, mentions of acknowledgees and the frequency of expressions of gratitude were explored. A sample analysis of an acknowledgement which is included in the Appendix is intended to help demonstrate how the analysis was done by the coder in detail.

## 4 Data analysis and discussion

### 4.1 Generic structure

When the generic structure of 20 MA thesis acknowledgements was analyzed, the complexity of identification of certain steps immediately became apparent; as Zhang–Jiang (2010: 98) put it “the structures of different moves and steps seem to be vague and indefinite”. Some steps in the corpora seem to be integrated and overlapping. For example, when Chinese mainland students expressed their gratitude to their supervisors, both academic support and moral support were appreciated. Examples could be illustrated by samples from the two contexts.

<sup>i</sup> <http://www.cnki.net> Downloaded: 2017-10-07.

<sup>ii</sup> <http://kuscholarworks.ku.edu> Downloaded: 2017-10-07.



[1] First of all, I really would like to express my sincere gratitude to my supervisor associate professor Zhang Zhenhong, for her great help and support, constant encouragement, enlightening advice, patient instruction and meticulous revision of this paper, without which the present paper could never have reached its present form. (CU-006)

Another example of the complexity of identifying the steps was that the same sentence might cover Step 2.2 and Step 2.3 of the thanking move, including gratitude to the supervisor for academic support and for help with data collection and analysis.

[2] I was very grateful to my instructor, Mr. Xu Haiming, who encouraged and guided me to write this thesis, gave me unremitting support during my data collection and data analysis phrases and valuable suggestions when I wrote and revised this paper. (AU-003)

A thorough analysis of the generic structures revealed that both the *Reflecting Move* and the *Thanking Move* identified in Hyland (2004)'s model were found in the two sample groups, but without *Move 3 Announcing Move*. The occurrence of moves and steps in both corpora is explicitly shown in Table 2 and Table 3. The most noticeable feature of the overall generic structure is the dominance of *Move 2 Thanking Move*, the obligatory move identified by Hyland (2004), with an occurrence of 100%. The other moves are far less frequent, with only 10% for the *Reflecting Move* and complete absence of the *Announcing Move*. As to the particular steps, it is clearly seen that M2 S2 *Thanking for academic support* and M2 S4 *Thanking for moral support* are the most common steps, occurring in every text, followed by M2 S1 *Presenting participants*, which occurs in 60% of cases in both sample groups and M2 S3 *Thanking for resources* (50% for CU and 60% for AU). Therefore, it may be concluded that Chinese MA students in both contexts are still unlikely to announce and dedicate their academic works. Possible reasons for this could be Chinese mainland students' over-emphasis of the modesty maxim of the politeness principle, under the influence of Confucian culture and the MA students' self-concept as novice researchers whose work cannot be regarded as accomplishments worthy of being dedicated to anyone (Zhang–Jiang 2010). Chinese MA students studying in the two contexts followed similar structural patterns of acknowledgements, and different academic conventions in American universities did not obviously affect Chinese students' avoidance of the announcing move.

| Moves and Steps                     | CU        |            | AU        |            |
|-------------------------------------|-----------|------------|-----------|------------|
|                                     | Frequency | Percentage | Frequency | Percentage |
| M1 <i>Reflecting Move</i>           | 1         | 10%        | 1         | 10%        |
| M2 <i>Thanking Move</i>             |           |            |           |            |
| S1 Presenting participants          | 6         | 60%        | 6         | 60%        |
| S2 Thanking for academic assistance | 10        | 100%       | 10        | 100%       |
| S3 Thanking for resources           | 5         | 50%        | 6         | 60%        |
| S4 Thanking for Moral support       | 10        | 100%       | 10        | 100%       |
| M3 <i>Announcing Move</i>           |           |            |           |            |
| S1 Accepting responsibility         | 0         | 0.0%       | 0         | 0.0%       |
| S2 Dedicating the thesis            | 0         | 0.0%       | 0         | 0.0%       |

**Table 2.** The frequency of acknowledgment with each step in the two corpora

| Moves and Steps                     | CU         | AU         |
|-------------------------------------|------------|------------|
| M1 <i>Reflecting Move</i>           | 1 (1.8%)   | 1(1.7%)    |
| M2 <i>Thanking Move</i>             |            |            |
| S1 Presenting participants          | 7 (12.7%)  | 6(10.3%)   |
| S2 Thanking for academic assistance | 25 (45.5%) | 24 (41.4%) |
| S3 Thanking for resources           | 9 (16.4%)  | 12 (20.7%) |
| S4 Thanking for Moral support       | 13 (23.6%) | 15 (25.9%) |
| M3 <i>Announcing Move</i>           |            |            |
| S1 Accepting responsibility         | 0 (0.0%)   | 0 (0.0%)   |
| S2 Dedicating the thesis            | 0 (0.0%)   | 0 (0.0%)   |

**Table 3.** The distribution of acknowledgment with moves and steps in the two corpora

Despite the similarities between the two groups, and the fact that both groups basically followed the same structures, there are still a number of differences. To begin with, even *Move 1 Reflecting Move* occurred in acknowledgements across the two contexts with quite low frequency. This might be explained by the possibility that compared with PhD students, Chinese MA students don't get used to reflecting their study journey due to shorter period of research.

Secondly, the occurrence (percentage) of *M2 S1 presenting the participants* in the CU corpus (12.7%) is slightly higher than in the AU corpus (10.3%). Chinese students in their own country seem to be more inclined to enumerate all the acknowledgees first and then get down to the details, which possibly conforms to the traditional English writing style - moving from general to specific - taught in their undergraduate courses. Thirdly, the percentages of *M2S3 thanking for resources* (20.7%) and *M2 S4 Thanking for moral support* (25.9%) are higher in the AU corpus than in the CU corpus, (16.4% and 23.6 % respectively). It could be assumed that Chinese mainland students who studied abroad at American universities have confronted more difficulties in collecting data and obtaining resources during their research and consequently received more assistance than those who studied in their home country. What is more, it is also quite natural that Chinese students studying at American universities suffer more from homesickness and similar hardships during their stay abroad, and require more moral and spiritual support from their family members, partners and friends.

Thirdly, the sequence of the moves also could reflect some differences between the two contexts. For example, it is also interesting to notice that *Move 1* serves as an opening move in the Chinese context (CU007), but appears at the end of the text in the American one (AU-002).

[3] During the three-year of postgraduate life, I have gained and experienced a lot, which cannot forget the people who sincerely help and boundlessly support me in my college life. (CU-007)

[4] The challenging while rewarding study experience in the department will always be part of me. (AU-002)

Another example is the flexible occurrence of *M2 S1 presenting participants* in both contexts. *M2 S1* could be used as the prologue to signal the purpose of this move – recognition of help from those involved in theses – or as the conclusion or summary of thanks to individuals and institutions already mentioned. Examples appear in Chinese students' writing in both American and Chinese universities.

[5] I would like to take this opportunity to show my great appreciation and gratitude to the following people for their selfless assistance and precious suggestions. (C-007, prologue)

[6] Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the research. (AU-008, conclusion)

It can be concluded that all of the Chinese postgraduates' MA acknowledgements across two different contexts generally followed Hyland's (2004) model, except for *Move 3 Announcing Move*. Academic conventions in different tertiary educational institutions and cultural factors in different settings might also have an effect on how generic structure is tailored to the different contexts.

### 3.2 Acknowledgees

How writers interact with the people acknowledged in thesis acknowledgements could be influenced by culture-specific conventions, according to Al-Ali (2010). Even though they share a similar cultural background, Chinese MA postgraduate students in two different contexts might exhibit differences when acknowledging people's assistance.

| Acknowledgees                   | CU        |            | AU        |            |
|---------------------------------|-----------|------------|-----------|------------|
|                                 | Frequency | Percentage | Frequency | Percentage |
| Supervisor                      | 10        | 100%       | 10        | 100%       |
| Other teachers                  | 8         | 80%        | 6         | 60%        |
| School deans / administrators   | 1         | 10%        | 1         | 10%        |
| Committee members               | 0         | 0%         | 5         | 50%        |
| Research assistants             | 0         | 0%         | 1         | 10%        |
| Participants in data collection | 5         | 50%        | 3         | 30%        |
| Family                          | 10        | 100%       | 9         | 90%        |
| Friends                         | 4         | 40%        | 5         | 50%        |
| Peers/ classmates               | 3         | 30%        | 1         | 10%        |
| Colleagues                      | 4         | 40%        | 1         | 10%        |
| Writers published in the field  | 4         | 40%        | 0         | 0%         |

**Table 4.** The frequency of gratitude expressed toward different acknowledgees in the two corpora

Table 4 displays the distribution of the different categories of acknowledgees in the two different contexts. The most noticeable difference is that Chinese students in America thanked a broader range of people in their acknowledgements, such as committee members and research assistants, while Chinese students in their home country showed special appreciation to the writers of published works in the research fields. These divergences could be explained by the different contexts and academic conventions. Expressing gratitude to committee members in American universities is quite common because these professors have already contacted the writers and offered their assistance before the thesis defense. What is more, showing gratitude to those who have a direct influence on how the thesis might be received, such as committee members, could be a strategy intended to impress the committees and get these accepted easily. However, academic conventions in Chinese universities differ sharply from those in America. The composition of MA thesis defense committees is supposed to remain confidential until the day of the defense, so there is no chance for Chinese students in China to find out who their committee members will be and express their gratitude in advance. At the same time, 40% of Chinese students in their home country universities expressed their gratitude to the authors of papers and books published in the related research field, which might show their respect for these figures in academic fields and ensure their academic identity as novice researchers.

### 3.3 Expressions of gratitude

Based on Hyland and Tse's (2004) analysis, expressions used to show gratitude in thanking moves and steps in acknowledgment can be categorized into five types: nominalisation (e.g. *My sincere thanks go to...*), performative verbs (e.g. *I thank...*), adjectives (e.g., *I am grateful to...*), passive voice (e.g. *Appreciation is given to...*) and bare mention (e.g. *X was very helpful in..., without...,* ).

| Gratitude expressions | CU        |            | AU        |            |
|-----------------------|-----------|------------|-----------|------------|
|                       | Frequency | Percentage | Frequency | Percentage |
| nominalisation        | 37        | 53.6%      | 15        | 25.9%      |
| performative verb     | 6         | 8.7%       | 28        | 48.3%      |
| adjective             | 9         | 13.1%      | 8         | 13.8%      |
| passive voice         | 3         | 4.3%       | 1         | 1.7%       |
| bare mention          | 14        | 20.3%      | 6         | 10.3%      |
|                       | 69        | 100%       | 57        | 100%       |

**Table 5.** The frequency of gratitude expressions in the two corpora

There were some similarities and differences between the two groups of Chinese MA students, from Chinese and American universities. For instance, Chinese students in both contexts generally employed these expressions of gratitude, but the rank orders were quite different. Chinese students in American universities preferred to use “performative verb” patterns most often (48.3%) and “nominalisation” (24.1%), while Chinese students in China tended to use “nominalisation” (53.6%) and “bare mention” (20.3%). The higher frequency of “performative verbs” in the American context could possibly be explained in terms of the cultural and contextual effects. Under the influence of the American academic context, using verbs to express their gratitude in a more direct way could be the first choice for Chinese students. A similar finding was also reported by Qin (2009: 33), who pointed out that cultural difference influences the way in which writers express their gratitude. Directness was highly valued by western culture while indirectness was appreciated by Chinese students under the influence of Confucian culture. “Nominalisation” and “bare mention” could be regarded as more indirect expressions of gratitude and the more common use of these expressions suggests that Chinese students were more reserved when expressing their gratitude and emotions. Another possible cultural explanation for the low frequency of “passive tense” could be the characteristics of the Chinese language, in which the passive voice is always presented covertly, with the possible result that Chinese students are inclined to avoid the passive voice in their English acknowledgments (Zhang–Jiang 2010: 106). In summary, experience of studying in American universities, academic conventions and other contextual factors do affect the expression of gratitude, but the use of language in acknowledgments is still influenced by the writer's cultural background and mother language, and the way in which these factors interact as they influence language use deserves further exploration.

## 4 Conclusion

This pilot study explored 20 MA theses in the field of English Applied Linguistics written by postgraduates from mainland China in Chinese universities and American universities and compared the generic structure and linguistic features across the two different contexts. It is concluded that even though the writers in the two corpora share the same cultural background and mother language, slight divergences still exist in the construction of moves and the linguistic realization of these moves. Chinese MA students in both contexts adopt the *Reflecting Move* and the *Thanking Move* in their academic writing, but eschew the *Announcing Move*, because they still hesitate to announce and dedicate their academic works in both contexts. However, slight differences also exist in Chinese students' acknowledgements between contexts. As to flexible positions of the move sequences, Move 1 *Reflection Move* and Move 2 Step 1 *Presenting participants* could occur at different locations in both contexts. Only Chinese students in American universities thanked committee members and research assistants, while only Chinese students in their home country expressed appreciation to the writers of published works in the research field. Students in the American context preferred to express gratitude in a direct way by using performative verbs while those in Chinese universities used more indirect ways to express their gratitude by using nominalisation and bare mention.

Cultural influence and mother language may be the factors influencing the generic features of Chinese postgraduates' acknowledgement writing, but many contextual factors such as academic conventions and contextualized values should be considered when analyzing these divergences. Possible pedagogical implications of the findings are that EAP teachers in China could offer training enhancing genre knowledge relating to acknowledgements and instruct students how to purposely and effectively express their gratitude by taking different contexts into consideration.

Additionally, more follow-up studies could be conducted. Further investigations on generic structure and linguistic features of acknowledgements could be carried out, using interviews to explore how and why Chinese postgraduate students construct moves and choose particular expressions in their acknowledgment writing. Besides the American context, or that of any other native English-speaking countries, countries belonging to the expanding circles could also be investigated to see whether these different contexts could bring similar influence to bear on acknowledgements written by Chinese MA students. In addition, similar comparative studies on MA students' thesis acknowledgements across contexts could also be extended to analyze non-native speakers of English from other countries. Yet another option would be to select PhD dissertations as samples.

## Appendix

---

|   |   |
|---|---|
| <p><b>Move 2</b> Thanking Move<br/> <b>Step 1</b> Presenting participants<br/> <b>Acknowledgees:</b><br/> (all the participants)<br/> <b>Gratitude expressions:</b><br/> (Nominalisation: <i>gratitude</i>)</p>   | <p>Acknowledgements</p> <p><i>I wish to express my sincere gratitude to the <u>many individuals</u> who have provided assistance and encouragement throughout my graduate work.</i></p>   |
| <p><b>Step 2</b> Thanking for academic assistance<br/> <b>Acknowledgees:</b><br/> (supervisor, committee members)<br/> <b>Gratitude expressions:</b><br/> (Bare mention: <i>without the support of.....</i>)<br/> (Performative verb: <i>appreciate, acknowledge</i>)<br/> (Nominalisation: <i>gratitude</i>)</p> | <p>This thesis would not have been possible <i>without the support of my <u>advisor</u></i>, Dr. Mary Morningstar, who has willingly spent the amount of time providing me detailed feedback and wise advice about this study. <i>I deeply appreciate</i> her guidance, advice, and instruction over the past two years. <i>I also wish to acknowledge</i> Dr. Knowlton and Dr. Skritic who are <u>committee members</u> for whom <i>I would like to express my gratitude</i> for their thoughtful consideration and corrections.</p> |
| <p><b>Step 3</b> Thanking for resources<br/> <b>Acknowledgees:</b><br/> (participants in data collection)<br/> <b>Gratitude expressions:</b><br/> (Nominalisation: <i>thanks</i>)</p>   | <p><i>A great deal of thanks also goes to 99 <u>doctoral students</u> who gave of their time to complete the survey.</i></p>  |
| <p><b>Step 4</b> Thanking for moral support<br/> <b>Acknowledgees:</b><br/> (family, especially the wife)<br/> <b>Gratitude expressions:</b><br/> (Nominalisation: <i>gratitude</i>)<br/> (Performative verb: <i>appreciate</i>)</p>  | <p>Finally, and most importantly, <i>I respectfully express my love and gratitude to <u>my family</u></i> who always trusted and encouraged me every step of the way through my master's program all the time. Especially, <i>my wife, I greatly appreciate</i> your special love and support from the bottom of my heart.</p>  |

---

Table 6. A sample analysis of acknowledgement

## References

- Al-Ali, M. N. 2006. *Conveying academic and social identity in graduate dissertation acknowledgments*. In: Proceedings of the 5th International Conference of European Association of Language for Specific Purposes, Zaragoza. 35–42.
- Al-Ali, M. N. 2010. Generic patterns and social-cultural resources in acknowledgements accompanying Arabic PhD dissertations. *Pragmatics* 20(1): 1–26.
- Bhatia, V. K. 1993. *Analysing genre: Language use in professional settings*. London: Longman.
- Bunton, D. 2002. Generic moves in Ph.D. thesis introductions. In J. Flowerdew (ed.) *Academic discourse*. London: Pearson Education Limited. 57–75.
- Cheng, S. W. 2012. A contrastive study of master thesis acknowledgements by Taiwanese and North American students. *Open Journal of Modern Linguistics*, 2(1): 8–17.
- Flowerdew, J. 2003. Signalling nouns in discourse. *English for Specific Purposes*, 22(4): 329–346.
- Giannoni, D. S. 2002. Worlds of gratitude: A contrastive study of acknowledgements texts in English and Italian research articles. *Applied Linguistics*, 23(1): 1–31.
- Hyland, K. 2000. *Disciplinary discourses: Social interactions in academic writing*. London: Pearson.
- Hyland, K. 2003. Dissertation Acknowledgements: The Anatomy of a Cinderella Genre. *Written Communication* 20(3): 242–268.
- Hyland, K. 2004. Graduates' gratitude: The generic structure of dissertation acknowledgements. *English for Specific Purposes*, 23(3): 303–324.
- Hyland, K., Tse, P. 2004. "I would like to thank my supervisor". Acknowledgements in graduate dissertations. *International Journal of Applied Linguistics*, 14(2): 259–275.
- Jalilifar, A., Mohammadi, M. J. 2014. Cross-Cultural Investigation into Generic Structure of Dissertation Acknowledgements in English and Persian: Reflections on Politeness Strategies. *Journal of Teaching Language Skills*, 6(1): 23–47.
- Karakaş, Ö. 2010. *A cross-cultural study on dissertation acknowledgments written in English by native speakers of Turkish and American English*. Unpublished master's thesis. Middle East Technical University, Ankara, Turkey.
- Kuhi, D., Rezaei, S. 2014. An Analysis of Generic Features of Acknowledgments in Academic Writing: Native Speakers of English vs. Non-Native (Iranian). *International Journal of English Linguistics*, 4(1): 19–43.
- Lasaky, F. 2011. A contrastive study of generic organization of doctoral dissertation acknowledgements written by native and non-native (Iranian) students in applied linguistics. *The Modern Journal of Applied Linguistics*, 3(2): 175–199.
- Lei, T. 2006. *Genre Analysis of acknowledgements in thesis and dissertation: A corpus-based study*. Unpublished master's thesis. Wuhan: Huazhong University of Science and Technology.
- Martin, M. 2003. A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes*, 22(1): 25–43.
- Mohammadi, M. J. (2013). Do Persian and English Dissertation Acknowledgments Accommodate Hyland's Model: a Cross-Linguistic study. *International Journal of Academic Research in Business and Social Sciences*, 3(5): 534–547.
- Qin, L. 2009. *A contrastive genre analysis of thesis and dissertation acknowledgements written in English and Chinese*. Unpublished master's thesis. Chongqing: Chongqing University.
- Qin, L. 2012. A study on writing acknowledgements of thesis and dissertation. Based on genre-based contrastive approach and cultural differences. *Journal of Shanxi Agricultural University (Social Science Edition)*: 11: 860–864.
- Soler, V. 2002. Analysing adjectives in scientific discourse: an exploratory study with educational applications for Spanish speakers at advanced university level. *English for Specific Purposes*, 21(2): 145–165.
- Swales, J. M. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Swales, J. M. 2004. *Research genres*. Cambridge: Cambridge University Press.
- Vassileva, I. 2001. Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes*, 20(1): 83–102.
- Wen, Q. F. 2004. *Applied linguistics: Research methods and thesis writing*. Beijing: Foreign Language Teaching and Research Press.



- Yang, R., Allison, D. 2003. Research articles in applied linguistics: Moving from Results to conclusions. *English for Specific Purposes*, 22: 365–385.
- Yang, W. 2012. Comparison of gratitude across context variations: A generic analysis of dissertation acknowledgements written by Taiwanese authors in EFL and ESL contexts. *International Journal of Applied Linguistics and English Literature*, 1(5): 130–146.
- Yang, W. 2013. Genre analysis of dissertation acknowledgements: A comparative study across contexts. *3L: Language, Linguistics, Literature The Southeast Asian Journal of English Language Studies*, 19(2): 27–40.
- Zhang, M., Jiang, Y. 2010. Dissertation Acknowledgements: Generic Structure and Linguistic Features. *Chinese Journal of Applied Linguistics*, 33(1): 94–109.

# Online-szótár-használati szokások vizsgálata Magyarországon

Gaál Péter

PTE BTK Nyelvtudományi Doktori Iskola  
gaal.peter.hun@gmail.com

**Kivonat:** A tanulmány egy két részből álló, magyarországi online-szótár-használati kutatás második fázisát mutatja be. A kutatás első szakaszában fordítók, idegennyelv-tanárok és idegennyelv szakos hallgatók körében kérdőíves felmérést végeztem a résztvevők online-szótár-használati szokásairól. A kapott eredményeket összegeztem és kiértékeltem, majd a második szakaszában a vizsgált célcsoport egy-egy képviselőjével mélyinterjúkat készítettem, hogy a kérdőíves eredményeket kibővítsam, pontosítsam. E tanulmány a három interjú tapasztalatait és a belőlük származó következtetéseket összegzi.

## 1 Bevezetés

### 1.1 A kutatás szakirodalmi előzményei

A szótárhasználati kutatások a lexikográfiában fontos vizsgálati területnek számítanak. Az elmúlt évek során – a felhasználó-központú megközelítés megerősödésének köszönhetően – egyre több az online szótárak használatával foglalkozó vizsgálat.

Az alábbiakban néhány fontosabb kutatást sorolok fel. Az ún. log fájlok elemzésének módszerét alkalmazta Laufer és Hill, melyben az elektronikus szótárak használatát és a szókincselsajátítási hatékonyság kapcsolatát vizsgálták angol szakos hallgatók körében (Laufer–Hill 2000, 2003). Szintén a logfájl-analízis módszerével élt De Schryver és Joffe, akik több mint 2500 szótárhasználó adatai alapján arra keresték a választ, hogy azok mely szavakra keresnek rá a leggyakrabban (Schryver–Joffe 2004). Sánchez Ramos 2005-ös vizsgálatában arra a következtetésre jutott, hogy a fordító szakos hallgatók nem igazán ismerik az elektronikus szótárakat (Sánchez Ramos 2005). Boonmoh és Nesi kb. 1200 középiskolás diák és 30 középiskolai tanár elektronikus szótárgép-használatát mérték fel kérdőívek segítségével (Boonmoh–Nesi 2008), a kutatást Boonmoh tovább folytatta 540 egyetemi hallgatóval (Boonmoh 2011). Bank három online szótár használhatóságát mérte fel 30 diák kérdőíves kérdésekre adott válaszai alapján (Bank 2010). Law és Li kérdőíves és interjú módszerrel 342 fordító szakos hallgató mobil telefonokon futtatható szótárakkal kapcsolatos ismereteit vizsgálta (Law–Li 2011). Dziemianko angol szakos hallgatók angol, egynyelvű szótárhasználatát (és a szótárhasználat hatékonyságát) vetette össze ugyanazon szótárak papír és elektronikus verzióinak viszonylatában (Dziemianko 2010, 2011, 2012).

Az utóbbi években az egyik legfontosabb, szótárhasználatra vonatkozó komplex, átfogó kutatást a mannheimi Institut für Deutsche Sprache végezte 2009 és 2011 között (IDS-kutatás: Müller-Spitzer–Koplenig–Töpel 2012, Müller-Spitzer 2014). A mannheimi projekt keretében kérdőíves felméréssel gyűjtöttek online-szótárhasználati adatokat, melyeket statisztikai módszerekkel dolgoztak fel. A nemzetközi kutatásokhoz képest a magyar szakirodalomban kevés a(z online) szótárak használatára vonatkozó kutatás (pl. Kontra 1997; Márkus–Szöllösy 2006; Földes 2005).

A mannheimi felmérés módszereit használtam fel arra, hogy vizsgálatot végezzek a magyarországi online-szótárhasználati szokásokról (OHM kutatás). A vizsgálatom két részből állt: az egyik részében a kérdőíves felmérést kisebb módosításokkal, magyarországi környezetben megismételtem, az eredményeket leíró és következtetési statisztikai módszerekkel kiértékeltem. A másik részben mélyinterjúkat készítettem a kérdőívet kitöltők három fő csoportjának (fordítók, idegennyelv-tanárok, idegen nyelv szakos hallgatók) egy-egy képviselőjével, ami azt a célt szolgálta, hogy a kérdőívekből származó eredményeket kibővítssem és árnyaljam, illetve (amennyire lehetséges) teszteljem. Az empirikus kutatás fő célja az volt, hogy adatokat szolgáltatson a magyarországi online-szótárhasználatról és az online szótárak használóival kapcsolatban. Tanulmányomban a három interjú feldolgozása, kiértékelése során szerzett tapasztalatokról számolok be.

## **1.2 Az interjúk előzményeit képező kérdőíves felmérés eredményei (OHM kutatás)**

A kérdőíves kutatás során az online kérdőíveket 294 válaszadó töltötte ki. Az IDS-kutatáshoz hasonlóan a kitöltők legnagyobb hányada fordítók, idegennyelv-tanárok és idegen nyelv szakos hallgatók közül került ki. A fő kutatási cél az volt, hogy információkat gyűjtsék arról, hogy a szótárhasználóknak milyen elvárásaik vannak az online szótárakkal kapcsolatban, mely szótárakat használják szívesen, melyek azok a szótári jellemzők, amelyeket a legfontosabbnak tartanak a szótárválasztás során (a részletes eredményeket lásd Gaál 2016, 2017). A használati gyakoriság tekintetében az online szótárak térnyerése szembevetendő. Az OHM kutatási adatai szerint a papírszótárhasználat kizárólagossága már csak nagyon kevesekre jellemző. Az online szótárak közül az egy- és kétnyelvű szótárak használata is rendkívül elterjedt, az egynyelvűeken belül pedig leginkább az általános-, és a szinonimaszótárakat használják a felhasználók. A szótárhasználat céljánál a szakmai célok dominálnak, de ez nincsen összefüggésben a szótárhasználatért cserébe fizetendő anyagi ellenszolgáltatással, hiszen a többség inkább „eltűri” az elektronikus hirdetések jelenlétét a szótároldalon, ha cserébe az összes szótári tartalomhoz ingyenesen hozzáfér. Az eszközhasználatról elmondható, hogy a legtöbben még mindig hagyományos számítógépeket (asztali számítógép, notebook) alkalmaznak, de már nagy arányt ér el azoknak a száma, akik a hordozható mobil eszközöket szótárhasználatra (is) igénybe veszik.

A főleg vagy kizárólag online szótárt használók aránya szinte minden korcsoportban (–21, 21–25, 26–30, 31–35, 36–45 év) 50% felett, vagy kicsivel alatta (46–55 év) volt. Az 55 év feletti korosztálynál viszont látványos az elmaradás: valamivel több, mint 10% tekinti csak elsődleges referenciaműnek az online szótárakat. Az egynyelvű online szótáraknál kimutatható, hogy a vizsgált populáció tagjai közül a fordítók nagyobb arányban fordulnak szakszótárakhoz, mint a nem fordítók.

A szótár árával kapcsolatos vizsgálatok nem igazolták egyértelműen, hogy azok, akiknél az online szótár kvázi munkaeszköznek számít (idegennyelv-tanárok, fordítók) nagyobb arányban lennének hajlandók fizetni az online szótárakért, mint a csoporton kívül esők. Összefüggés mutatkozott azonban abban a tekintetben, hogy a nem fordítókhoz képest a fordítókra kevésbé jellemző, hogy kis képernyős eszközökön futtatják az online szótárakat.

A szótárak szerkesztői háttere alapján kiemeltem a közösség által szerkesztett szótárakat, és megvizsgáltam, hogy a fiatalabb korosztályok megbízhatóbbnak tartják-e ezt a szótártípust az idősebb generációba tartozókhoz képest. A statisztikai számítások alapján ez a feltételezés nem nyert bizonyítást.

Az IDS-kutatással egybevágó módon a megkérdezettek az OHM kutatásban is a megbízhatóságot, áttekinthetőséget és naprakészséget a legfontosabbnak, a testreszabhatóságot és a multimédiát pedig a legkevésbé fontos jellemzőnek ítélték. A szerkesztői háttér alapján a szótárkiadó cégek és az intézmények/szervezetek által kiadott szótárak bizonyultak a leginkább megbízhatóknak.

## 2 A módszertan

### 2.1 Az interjúk és a kutatási kérdések

Az interjúk lebonyolításának algoritmizálásához Kvale (2005: 95) hétlépcsős modelljét vettem alapul. Az interjú a következő fázisokból állt:

1. Témamegjelölés: az interjúk témája a magyarországi online-szótár-használat a fordítók, idegennyelv-tanárok és idegen nyelv szakos hallgatók körében. Az interjúkészítés célja a kérdőíves felmérésből származó eredmények és következtetések tesztelése, illetve kibővítése volt.
2. Tervezés: az elméleti háttér az IDS-kutatás, és az OHM kutatás kérdőíves felmérésének eredményei adták. A tervezési fázishoz tartozott a kutatásban részt vevő személyek informálása a kutatás céljáról és anonimitásuk garantálásáról.
3. Interjúvolás: a félig strukturált interjúk vezérfonalát az OHM kutatás kérdőívének struktúráját követő kérdések és az IDS-kutatásában megjelenő, de az OHM kutatásból kimaradó egyéb pontok adták. Bizonyos kérdésekhez szemléltető anyagként illusztrációkat használtam. Az interjúkészítés során az interjúalanyokat megkértem, hogy ahol szükségét érzik, fejtsék ki a kérdéshez szorosan nem tartozó, de fontosnak ítélt gondolataikat is a szótárhasználati szokásaikkal kapcsolatban, annak érdekében, hogy további szubjektív információkhoz juthassak. Az interjúk időtartama átlagosan 35–40 percet ölelt fel.
4. Adatrögzítés: az interjúkról digitális hangfelvételt készítettem, majd a rögzített hanganyagot (begépelés útján) írott szöveggé alakítottam át.
5. Elemzés: az elemzés során először a három interjúalany azonos kérdéskörre adott válaszait hasonlítottam össze, majd azokat összevettem IDS- és OHM kutatásokból származó adatokkal.
6. Vizsgálat: az interjúkból és a kérdőíves felmérésekből származó, egymással párhuzamba állított eredményeket abból a szempontból vizsgáltam, hogy mennyire egybehangzóak vagy különbözőek.
7. Beszámoló: az összesített eredményeket és következtetéseket a fejezet végén összegeztem.

Kutatásomhoz az *egyedülálló, félig strukturált interjút* (vö. Dörnyei 2007: 134–136) választottam, mivel ez a típus felelt meg leginkább a céljaimnak. Előre elkészített kérdéseim a kérdőíves felmérésben feltett kérdéseimhez igazodtak, továbbá kiegészültek a nálam nem, de az eredeti, IDS-kutatásban szereplő további kérdésekkel, amelyek végül a kérdőívembe nem kerültek be. A kérdéshez egy esetben (a szócikkek megjelenítésére vonatkozóan), a könnyebb érthetőség érdekében szemléltető ábrákat is mutattam az interjúalanyoknak. Mindhárom interjúalany minimális különbségekkel és módosításokkal ugyanazokat a kérdéseket kapta.

Az interjúk lebonyolítása szóban történt, két alannal személyesen zajlott az interjú, a harmadikkal (a jelentős földrajzi távolság miatt) skype-os videobeszélgetésen keresztül történt a kommunikáció. Az interjúalanyok az anonimitás érdekében tanulmányomban nem valódi nevükön, hanem álnéven szerepelnek. Az interjúkról hangfelvétel készült, amelyek szövegét utólag jegyeztem le.

Az interjúkhoz kapcsolódó kutatási kérdéseim a következők voltak:

- (1) Milyen egyezések és különbségek mutatkoznak egymáshoz képest a három interjúalany által adott válaszokban?
- (2) Az interjúkból kapott információk megerősítik vagy cáfolják az IDS- és az OHM kutatás kérdőíves részéből származó eredményeket?
- (3) Az OHM kutatásban nem szereplő kérdéskörökre adott válaszok párhuzamba állíthatók-e az IDS-kutatásban kapott eredményekkel?

## 2.2 Az adatközlők

Az interjúban részt vevő három alany az OHM kutatás kérdőíves részében elsősorban megszólított csoportokból, tehát a fordítók, az idegennyelv-tanárok és az idegen nyelv szakos hallgatók köréből került ki. Az interjúalanyok a kérdőívet előzetesen nem töltötték ki, azok kérdéseit nem ismerték. Tanulmányomban az interjúalanyok nem valódi nevükön, hanem álnéven szerepelnek, így biztosítva anonimitásukat. A válaszadóknak a következő elnevezéseket adtam: Réka, Anita és Emese.

Réka 1978-ban született nő, angol szakos, egyetemi végzettségű. Fordítóként dolgozik egy fordítóipari cégnél, ahol a fordítók (akik főként szakfordítók és mérnökök) munkáját segíti a háttérből, feladatai közé tartozik a cégnél használt terminológiai adatbázis és fordítói memória karbantartása. Mellékállásban, magánúton angol–magyar és magyar–angol nyelvpárban vállal fordításokat, időnként angolnyelv-oktatást és nyelvvizsgára felkészítést is. Anyanyelve magyar, angolból „ERŐS C1 SZINTEN VAN”, spanyolból B2-es szinten, de ebből a nyelvből nem fordít, olasz és német nyelvtudásának szintjét körülbelül A2-es szinten határozta meg.

Anita 1970-ben született nő, tanító szakon végzett német specializációval, majd ezután elvégezte a német nyelv és irodalom tanári szakot. Pályafutását általános iskolai pedagógusként kezdte, majd szakközépiskolában németnyelv-tanárként folytatta. Ezután a céges nyelvoktatásban tevékenykedett, főleg szakmai nyelvet (pénzügyi, műszaki) oktatott, illetve egészségügyi szaknyelvi nyelvvizsgára felkészítő kurzusokat tartott egészségügyi főiskolára járó hallgatóknak. Jelenleg ismét középiskolában tanít német nyelvet, szakképző és szakközépiskolásoknak. Anyanyelve a magyar, németnyelv-tudását „C2 SZINT KÖRÜL”, angoltudását A2 szintűnek ítélte meg.

Emese 1992-ben született nő, anglisztika alapszakon végzett, kínai nyelv minor szakkal kiegészítve, jelenleg fordító-tolmács mesterszak megkezdése előtt áll. Nyelvtanulás céljából másfél évet töltött Kínában, ösztöndíjas formában. Anyanyelve

magyar, kínai és angol nyelvtudását egyaránt C1-es szintben, németnyelv-tudását B1 és B2-es szint között határozta meg.

### **3 Eredmények**

#### **3.1 A használt szótártípusok**

##### **3.1.1 Hagyományos vs. online szótár**

A három interjúalany mindegyike főleg online szótárakat használ, ami összhangban áll mind az IDS-, mind az OHM kutatás kérdőíves felmérésének eredményeivel, hiszen ott is ez volt a jellemző a válaszadók legnagyobb hányadára. Az interjúalanyoknál a hagyományos szótárak háttérbe szorulásában feltehetően az életkoruk is közrejátszik, hiszen a legidősebb is csak 46 éves közülük, így a technikai eszközöktől, az internettől való idegenkedés vagy hozzá nem értés nem befolyásolja online-szótár-használati szokásaikat.

##### **3.1.2 Kétnyelvű és egynyelvű online szótárak használata**

A válaszadók közül mindannyian használnak egy- és kétnyelvű szótárakat is. Egyik szótártípus aránya sem tér el nagymértékben a másiktól, ami az IDS- és az OHM kutatások kérdőíves válaszaiban is hasonló eredményt hozott.

##### **3.1.3 A használt egynyelvű szótártípusok**

Az általános szótárak használatának magas aránya az IDS és az OHM kérdőíves kutatásokban is megfigyelhető volt, és az interjúalanyoknál is ez volt az a szótártípus, amit minden válaszadó említett. További hasonlóság, hogy a kérdőívekben második leggyakoribb szótártípusnak jelölt szinonimaszótárakat is mindhárman említették.

Az OHM kutatásban vizsgáltam a fordítók szakszótár-használatának gyakoriságát, ahol az az eredmény jött ki, hogy ők jóval nagyobb arányban használják ezt a szótártípust, mint a nem fordítók. Ezt a jelenséget megerősítette, hogy az alanyok közül Réka (aki ugyancsak fordító), az interjú során szintén kiemelte, hogy az általános szótárak mellett nagyon gyakran használ szakszótárakat. Az interjú során kaptam példát arra a jelenségre is, amit Prószéky (2011: 8) említ cikkében, miszerint sok szakterületen valódi szakszótár nem is létezik, ezért azt sokszor egy szűk szakembergárdának vagy egy szakfordítónak a saját gyűjtése helyettesíti.

## 3.2 A szótárhasználat külső körülményei

### 3.2.1 A szótárhasználat célja

A szakmai célú online-szótár-használat fölénye mind az interjúk során, mind a kérdőívek eredményeiből jól látszik. Ennek háttérében az állhat, hogy a vizsgált célcsoport tagjainak munkájukból és/vagy tanulmányaikból kifolyólag muszáj online szótárakat használniuk.

Az öt vizsgált szótárhasználati tevékenységnél az IDS-kutatásból származó eredményekkel vetettem össze a kapott válaszokat. A „fordításhoz” kötődő szótárhasználathoz az IDS-kutatásban a többség (71%) a „gyakran” lehetőséget jelölte meg, ami egybevégt az interjúk során kapott információkkal, hiszen mindenki ezt az opciót választotta. A szövegalkotásnál a gyakran (47,1%) és a néha (40,4%) voltak a leggyakoribb válaszok, ami szintén hasonló az interjúkhoz (egy gyakran, és két néha válasz). Az IDS-kutatásban ugyanez volt a két legnépszerűbb válasz „javításhoz” lehetőségre: gyakran (39,6%), néha (37,1%), ami az OHM kutatásban két néha és egy gyakran válaszban realizálódott. Az „olvasáshoz” választási lehetőségre a néha (38,9%) és a ritkán (32,3%) voltak a legnépszerűbb visszajelzések, az interjúkban két ritkán és egy gyakran válasz született. Végül az „internetes böngészéshez” elnevezésű tevékenységnél a legtöbben (36,9%), a ritkán-t jelölték meg, szemben az interjúalanyok két néha és egy gyakran feleletével.

Összegezve tehát: a fordítás, szövegalkotás, javítás voltak azok a tevékenységek, ahol egyértelműen párhuzam vonható az interjúalanyok és az IDS-kutatásban részt vevők többségére jellemző szótárhasználati szokások között, és ez részben igaz az olvasásra is, egyedül az internetes böngészés során tűnik gyakoribbnak a szótárhasználat az interjúkban megszólalók körében.

### 3.2.2 A szótár ára

Akárcsak az IDS- és OHM kutatásokban, az interjúk során is az ingyenes szótárakat preferálták a válaszadók. További megerősítést nyert tehát, hogy alapvetően a szótárhasználók anyagi ellenszolgáltatás nélkül próbálnak hozzájutni az online szótári tartalmakhoz, még ha – bizonyos feltételek teljesülése esetén – teljesen nem is zárkoznak el a fizetős szótáraktól.

Prószéky (2011: 5) Közép-Kelet-Európára különösen jellemző jelenségnek látja azt a tendenciát, hogy „a fizetős internetes szolgáltatások használatának nem alakult ki igazi kultúrája”, és sokan úgy látják, hogy egy online szótár esetében „nem baj, ha nem igazán jó, csak ne kérjenek érte pénzt, mert annyira azért nem kell”. A kérdőíves felmérések és az interjúk tükrében igazat kell adni Prószéky megállapításának, azzal a kiegészítéssel élve, hogy az IDS-kutatásból kiindulva nem csak Kelet-Közép-Európára, hanem nyugatabbra sem jellemző az online szótárakért történő nagyobb fizetési hajlandóság.

### 3.2.3 Az online szótárak futtatási platformjai

Az IDS- és az OHM kutatásban megadott szótárfuttatási-plattformok mindegyikét említették az interjúk során is. Személyenként változó volt, hogy kinél melyik volt a leginkább preferált eszköz. Az azonban a mobil eszközök népszerűségét támasztja alá, hogy mindegyik interjúalany használ (kisebb vagy nagyobb mértékben) okostelefont vagy táblagépet. A kisképernyős eszköz-használat aránya az OHM kutatásban a nem fordítóknál jóval magasabb volt, mint a fordítóknál, ami egyaránt igaz volt az összes eszközt (73,53% és 58,44%) és a leggyakrabban használt eszközt (19,85% és 7,14%) tekintve is. A meginterjúvált válaszadók is megerősítették ezt a trendet, hiszen a nem fordító Anitánál és Emesénél a táblagép és az okostelefon, míg a fordító Rékánál a hagyományos asztali számítógép volt a szótárhasználatra leggyakrabban igénybe vett elektronikus platform.

## 3.3 Az online szótári jellemzők

### 3.3.1 Tíz szótári szempont rangsorolása fontosság alapján

Az interjúalanyok viszonyulását a tíz megadott online szótári jellemzőhöz meglehetősen nehéz összevetni a kérdőíves adatok összesítésével, hiszen egy-egy egyéni vélemény adott esetben szöges ellentétben állhat a nagy mennyiségű adatból levont következtetésekkel. A könnyebb összehasonlíthatóság érdekében a három interjúalany szótári jellemzőkre adott sorszámaikat összeadtam, és ennek alapján állítottam fel a hármójuk véleménye alapján kirajzolódó sorrendet, amely a következőképpen alakult: Megbízhatóság (1.), Sebesség (2–3.), Naprakészség (2–3.), Áttekinthetőség (4.), Testreszabhatóság (5.), Multimédiás tartalom (6–7.), Korpuszhoz kapcsolódó linkek (6–7.), Más szótárakhoz kapcsolódó linkek (7–8.), Hozzáférhetőség (9.), Böngészési javaslatok (10.).

A kérdőíves felmérésekben abszolút első helyre rangsorolt *megbízhatóság* az interjúk eredményeinek összesítésében is az első helyre került. A második legfontosabb *áttekinthetőséget* az interjúalanyok összességében azonban csak a negyedik pozícióba rangsorolták. A *sebesség* és a *naprakészség* az IDS- és OHM kérdőívek eredményeihez hasonló helyre került. Meglepetést jelent viszont a *testreszabhatóság*, hiszen az interjúkban viszonylag fontosnak (5. hely) ítélték, míg az IDS-kutatásban nyolcadik, az OHM kutatásban pedig az utolsó lett. Hasonlóan fontosabbnak gondolták az interjúalanyok a *multimédiát*, hiszen a 6–7. helyezést kapta, a kérdőíves felmérés 9–10. helyével szemben. A *korpuszhoz kapcsolódó linkek* és a *más szótárakhoz kapcsolódó linkek* helyezése (minimális eltéréssel) minden forrás szerint egyezett, a *hozzáférhetőség* azonban csak kilencedik lett az interjúalanyok minősítésében, míg a kérdőívekben a „középmezőnybe” (az 5. helyre) sorolták. Az interjúkban utolsó helyezett *böngészési javaslatok* a kérdőívekben is hátul (8–9. hely) végzett.



### **3.3.2 Az online szótárak megbízhatóságának megítélése a szerkesztői háttér alapján**

Az öt – szerkesztői háttér szerint megkülönböztetett – szótártípusból az interjúkban is a szótárkiadó cég és az intézmények által kiadott szótárakat említették mint a legmegbízhatóbb altípusokat, ebben tehát nem volt különbség az IDS- és OHM kérdőívekhez képest. A felhasználók által szolgáltatott információkat felhasználó szótárak esetében az a kép rajzolódott ki, hogy ezt a fajta referenciaművet nem igazán tartják számon külön kategóriaként, de ettől függetlenül megbízhatóbbnak vélik, mint a közösségi vagy magánszemély által szerkesztett szótárakat. Ez utóbbi két kategóriát – akárcsak a kérdőívek eredményeiben – utolsó helyre rangsorolták (a megbízhatóság aspektusából).

### **3.3.3 A szótári jellemzők részletesebb értékelése**

A megbízhatóság megítélésekor az IDS-kutatás válaszadói a valós nyelvhasználat megjelenését (ami korpuszon alapszik) várták el leginkább egy online szótártól, ami két interjúalany véleményével is egybevág. Anita viszont azt választotta a legfontosabbnak, hogy egy neves kiadó cég vagy intézmény üzemeltesse a szótárt, ami az IDS-kutatásban (8,2%-kal) messze a legnépszerűtlenebb opció volt.

A naprakészség tekintetében az OHM kutatásban a leginkább előre sorolt jellemvonás az aktuális nyelvészeti információk mielőbbi szótárba kerülése volt, ez azt jelenti, hogy megegyezik az IDS-kutatás első helyezettjével.

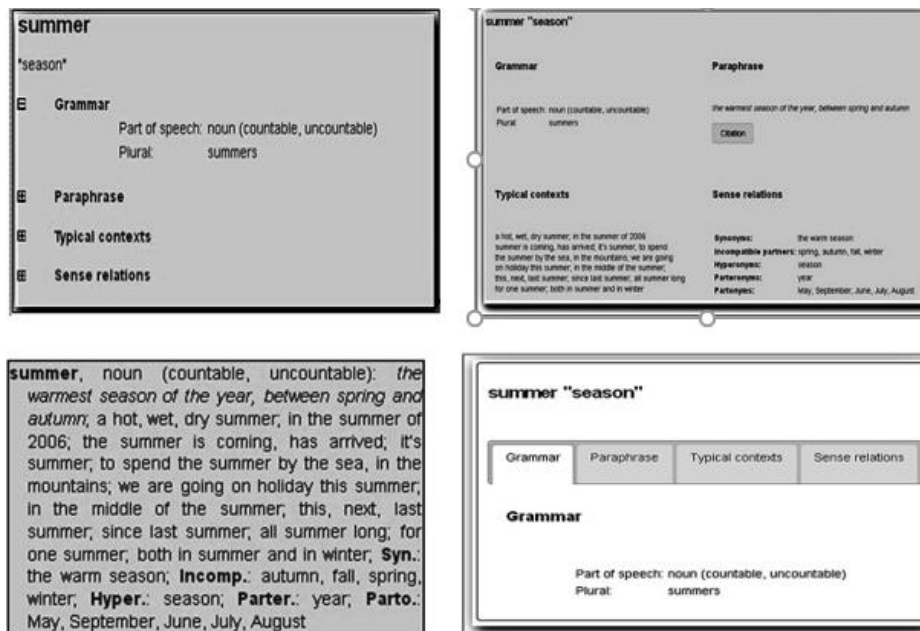
A hozzáférhetőségnél az OHM kutatásban teljesen eltérő vélemények születtek, az IDS-kutatásban első helyre rangsorolt állítást (Ne legyenek karbantartás miatti szerverleállások) egy interjúalany jelölte meg.

### **3.3.4 Az online szótárak innovatív jellemzői**

A kiejtést segítő hangfájlok talán a leggyakrabban felbukkanó multimédiás elemek az online szótárakban, és úgy tűnik, hogy a szótárhasználók élnek is ezzel a lehetőséggel a szótárhasználat során. Valamilyen szinten nyilván ennek is szerepe van abban, hogy az IDS és az OHM kutatásban is a leginkább hasznosnak ítélt kiegészítőként ítélték meg a lehetséges válaszok közül.

### **3.3.5 A szócikkek megjelenítésére vonatkozó kérdések**

A négy választható szócikkforma közül (lásd 1. ábra) az IDS- és az OHM kutatásban is egyértelműen a „Print view” nézetet sorolták az utolsó helyre, ami alapján az látszik, hogy az online szótárak használói nem szívesen látják viszont a hagyományos szótárak szócikkstruktúráját és megjelenését digitális formában, hanem ennél innovatívabb megoldásokra vágnak, mint amilyen például az IDS-kutatásban első helyen végzett „Tab view”. A napjainkban különféle egyéb szoftverekben is alkalmazott megoldás Anita számára is fontos volt; ugyanakkor Réka idegenkedett ettől a megoldástól. Az „Explorer view”-t minden válaszadó egyaránt a második legjobbnak ítélte, a „Panel view” pedig Réka és Emese számára is a legjobbnak bizonyult.



1. ábra. Az interjúalanyok által rangsorolt szócikktípusok (balról jobbra haladva: Explorer view, Panel view, Print view, Tab view; forrás: Müller-Spitzer 2014: 192)

#### 4. Összegzés

Tanulmányomban az Online-szótár-használat Magyarországon (OHM) nevű kutatásom interjúkon alapuló részét ismertettem. Az interjúk kapcsán megfogalmazott kutatási kérdésekre adható válaszaimat az alábbiakban összegzem.

(1) **Milyen egyezések és különbségek mutatkoznak egymáshoz képest a három interjúalany által adott válaszokban?** Néhány kérdés vonatkozásában az alanyoktól egyforma vagy nagyon hasonló válaszokat kaptam. Ebbe a körbe lehet sorolni az online szótárak használatának magas arányát, amelyek között egy- és kétnyelvűek is előfordulnak, ezek közül is az általános egynyelvű szótár az, amelyet mindenki igénybe vesz. A célok közül egyértelműen a szakmai célú szótárhasználat dominál, a tevékenységek közül pedig a fordítás az, amelyhez online szótárakat használnak. Az interjúalanyok közül senki nem fizetett még online szótári tartalomért, de adott esetben elképzelhetőnek tartják, hogy ez változni fog. A szótári jellemzők közül főleg a megbízhatóságot, a sebességet és a naprakészséget emelték ki mint legfontosabb szempontokat. A szótárkiadók és az intézmények által szerkesztett szótárak megbízhatóságát tartották a legmagasabbnak, a közösségi és magánszemély által szerkesztett szótárakat pedig a legkevésbé megbízhatónak. Az online szótárak innovatív tulajdonságai közül egybehangzóan a kiejtést segítő hangfájlokat preferálják leginkább.

A legszembetűnőbb különbségek az interjúalanyok válaszaiban a leggyakrabban használt szótárfuttatási platformokban (okostelefon, táblagép, asztali számítógép), néhány fő szótári jellemzőben (pl. áttekinthetőség, multimédiás tartalom), valamint azok részletesebb tárgyalásában (pl. hozzáférhetőség, naprakészség), illetve a szócikkek dizájnját illetően voltak.

**(2) Az interjúkból kapott információk megerősítik vagy cáfolják az IDS- és az OHM kutatás kérdőíves részéből származó eredményeket?** Az interjúalanyok válaszaik nagyjából összhangban álltak a kérdőíves felmérésekből származó eredményekkel. Ez elsőként az online-szótár-használat magas arányában és az alkalmazott szótárak típusainál jelentkezett. Az interjúkból származó információk megerősítették továbbá, hogy a fordítók gyakran használnak online szakszótárakat. A szótárhasználati célok között megerősítést nyert, hogy a vizsgált célcsoport tagjai főleg szakmai célokra alkalmazzák az online szótárakat. A szótár árával kapcsolatban kijelenthető, hogy a kérdőíves felmérések eredményeinél tapasztalható szótárakért történő fizetési hajlandóság (illetve annak hiánya) az interjúalanyoknál is megfigyelhető volt. Az eszközhasználatot illetően a nem fordítók nagyobb arányú kisképernyős eszköz-használata szintén visszaköszönt az interjúk során is. A tíz szótári jellemző közül a megbízhatóság, naprakészség, áttekinthetőség, sebesség voltak azok, amelyek hasonló fontossággal bírtak az interjúalanyoknál, mint a kérdőíveket kitöltőknél. Végül a szótárak megbízhatóságának szerkesztői háttér szerinti megítélése is a kérdőíves eredményekhez hasonlóan történt.

Az kérdőívekhez képest az interjúkban a megkérdezettek másképpen ítélték meg a hozzáférhetőség, testreszabhatóság és a multimédia fontosságának szerepét, de egyéb kirívóan eltérő vélemények a többi kérdésre vonatkozóan nem voltak.

**(3) Az OHM kutatásban nem szereplő kérdéskörökre adott válaszok párhuzamba állíthatók-e az IDS-kutatásban kapott eredményekkel?** Az OHM kutatás kérdőívében nem szereplő pontok közül a szótárhasználati tevékenységek közül a fordítás, szövegalkotás, és javítás voltak azok, ahol az interjúalanyok hasonló, az olvasás esetében részben hasonló, a böngészés során pedig eltérő gyakorisággal használtak online szótárakat. A szótári jellemzők részletesebb értékelésénél, a naprakészségnél azonos, a megbízhatóságnál részben azonos, a hozzáférhetőséget illetően azonban eltérő vélemények születtek az IDS-kutatás eredményének átlagaihoz képest. Az innovatív jellemzők rangsorolásánál az IDS-kutatáshoz hasonlóan elől végeztek a kiejtést segítő hangfájlok, és abban is egyformán nyilatkoztak a válaszadók, hogy a hagyományos szócikktípust szeretnék a legkevésbé viszontlátni egy online szótárban.

## Irodalom

- Bank, C. 2010. *Die Usability von Online-Wörterbüchern und elektronischen Sprachportalen*. Hildesheim: Universität Hildesheim.
- Boonmoh, A., Nesi, H. 2008. A survey of dictionary use by Thai university staff and students, with special reference to pocket electronic dictionaries. *Horizontes de Lingüística Aplicada*, 6(2): 79–90.
- Boonmoh, A. 2011. Students' knowledge of pocket electronic dictionaries: recommendations for the students. In: Akasu, K., Satoru, K. (szerk.) *ASIALEX2011 Proceedings Lexicography: Theoretical and practical perspectives*. Kiotó: Asian Association for Lexicography. 66–75.
- Dziemanko, A. 2010. Paper or electronic? The role of dictionary form in language reception, production and the retention of meaning and collocations. *International Journal of Lexicography* 23(3): 257–273.

- Dziemanko, A. 2011. Does dictionary form really matter? In: Akasu, K.– Satoru, U. (szerk.), *ASIALEX2011 Proceedings Lexicography: Theoretical and practical perspectives*. Kiotó: Asian Association for Lexicography. 92–101.
- Dziemanko, A. 2012. On the use(fulness) of paper and electronic dictionaries. In: Granger, S. – Paquot, M. (szerk.) *Electronic lexicography*. Oxford: Oxford University Press. 319–341.
- Dörnyei, Z. 2007. *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- De Schryver, G. M., Joffe, D. 2004. On How Electronic Dictionaries are Really Used. In: Williams, G., Vessier, S. (szerk.) *Proceedings of the Eleventh EURALEX International Congress, Lorient, France, July 6th–10th*. Lorient: Université de Bretagne Sud. 187–196.
- Földes, A. 2005. „tök jó, de nincsenek benne csúnya mondatok” – Egy WAP-alapú szótári rendszer üzemeltetésének tapasztalatai. In: Alexin, Z., Csendes, D. (szerk.) *III. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport. 155–166.
- Gaál, P. 2016. Online-szótár-használat Magyarországon (OHM) – egy kérdőíves szótárhasználati felmérés eredményei I. *Alkalmazott Nyelvtudomány*, 16(2). Elérhető: <http://alkalmazottnyelvtudomany.hu/wordpress/wp-content/uploads/GaalP.pdf>. Letöltve: 2017. január 2.
- Gaál, P. 2017. Online-szótár-használat Magyarországon (OHM) – egy kérdőíves szótárhasználati felmérés eredményei II. *Alkalmazott Nyelvtudomány*, 17(1). Elérhető: [http://alkalmazottnyelvtudomany.hu/wordpress/wp-content/uploads/Gaal\\_P.pdf](http://alkalmazottnyelvtudomany.hu/wordpress/wp-content/uploads/Gaal_P.pdf). Letöltve: 2017. szeptember 26.
- Kontra, M. 1997. Szótárhasználók a magyar szótárirás aranykorában. In: Kiss, G., Zaicz, G. (szerk.) *Szavak – Nevek – Szótárak. Írások Kiss Lajos 75. születésnapjára*. Budapest: MTA Nyelvtudományi Intézete. 228–232.
- Kvale, S. 2005. *Az interjú – Bevezetés a kvalitatív kutatás interjútechnikáiba*. Budapest: Jászöveg Műhely.
- Laufer, B., Hill, M. 2000. What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning & Technology*, 3(2): 58–76.
- Laufer, B., Hill, M. 2003. Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *IRAL-International Review of Applied Linguistics in Language Teaching*, 41(2): 87–106.
- Law, K., Li, K. 2011. Mobile Phone Dictionary: Friend or Foe? A User Attitude Survey of Hong Kong Translation Students. In: Akasu, K., Uchida, S. (szerk.) *ASIALEX2011 Proceedings Lexicography: Theoretical and practical perspectives*. Kiotó: Asian Association for Lexicography. 303–312.
- Márkus, K., Szöllősy, É. 2006. Angolul tanuló középiskolásaink szótárhasználati szokásairól. In: Magay T. (szerk.) *Szótárak és használók (Lexikográfiai füzetek 2.)*. Budapest: Akadémiai Kiadó. 95–116.
- Müller-Spitzer, C. (szerk.) 2014. *Using Online Dictionaries (Lexicographica Series Maior 145.)* Berlin–Boston: Walter de Gruyter.
- Müller-Spitzer, C., Koplénig A., Töpel A. 2012. Online dictionary use: Key findings from an empirical research project. In: Gragner, S., Paquot, M. (szerk.) *Electronic Lexicography*. Oxford: Oxford University Press. 425–458.
- Sánchez Ramos, M. M. 2005. Research on dictionary use by trainee translators. *Translation Journal*, 9(2). Elérhető: <http://translationjournal.net/journal/32dictuse.htm>. Letöltve: 2016. október 25.
- Prószéky, G. 2011. A szótári világ átalakulási tendenciái az internet megjelenésével. *Modern Nyelvoktatás*, 17(4): 3–11.

## Nyelvmegtartás és identitásépítés egy nemzetiségi horvát településen

Huszár Attiláné Fábriánkovits Erzsébet

ELTE BTK Nyelvtudományi Doktori Iskola  
huszar.attilane1@gmail.com

**Kivonat:** A dolgozat célja, hogy felmérje egy nyugat-dunántúli horvát nemzetiséggel rendelkező településen – Kópházán – tapasztalható nyelvmegtartó, identitásépítő törekvéseket, valamint azt, hogy a horvát nyelv presztízsének növekedése és az identitásépítő törekvések milyen mértékben befolyásolják a nyelvcsere és a nyelvmegőrzés egy időben jelenlévő folyamatát. A nyugat-dunántúli horvátok (grádistyei horvátok) helyzete speciális, hiszen közösségeik nemcsak Horvátországgal tartanak fenn kapcsolatokat, hanem az ausztriai grádistyei horvát régióval is, melytől a trianoni határ szakította el őket. Ennek megfelelően a vizsgált településen a horvát nyelv három változata figyelhető meg: a magánéletben használt nem standardizált archaikus horvát dialektus, a grádistyei horvát standardizált nyelvváltozat, valamint a Horvátországban beszélt standard horvát nyelv, amelyet a helyi általános iskolában is tanítanak, s amely a kulturális kapcsolatok hatására, valamint a nyelvi kontaktusok következményeként hatással van a beszélőközösség horvát nyelvhasználatára.

### 1 Bevezetés

A hazai horvát kisebbség a kutatók véleménye szerint hét, egyesek szerint tizenkét részben eltérő nyelvjárást beszélő, egymástól földrajzilag is elkülönülő népcsoportra oszlik (Soksevits 2013: 1). A grádistyei horvátok, a nemzetközi szakirodalomban Burgenland Croatian (Crystal 1998: 25), a trianoni határok által feldarabolt népcsoport kis része Magyarországhoz, valamint a mai Szlovákiához és Csehországhoz, nagyobb része Ausztriához került a békeszerződést követően. A grádistyei horvát települések a mai Győr-Moson-Sopron és Vas megye területén találhatók. Kutatásomat a népcsoport legnagyobb számú horvát identitású lakossággal rendelkező magyarországi településén (Szarka 2003: 290), a Sopron mellett elhelyezkedő Kópházán (horvát neve: *Koljnof*) végeztem.

Tanulmányom célja, hogy felmérjem a faluban tapasztalható nyelvmegtartó, identitásépítő törekvéseket, valamint azt, hogy a horvát nyelv presztízsének a politikai és gazdasági változások eredményeként bekövetkezett növekedése milyen mértékben befolyásolja a nyelvcsere és nyelvmegőrzés egy időben jelenlévő folyamatát (Bartha 1999: 130; Borbély 2001: 195). Radó Péter megfogalmazása szerint „(...) az anyanyelv státusa igen fontos abból a szempontból, hogy a kisebbségi csoporthoz való

tartozás mennyire lehet büszkeség forrása, mennyire segíti az azonosságtudat kialakítását és fenntartását” (Radó 1992: 141).

A településen élő horvátok (a helyi önkormányzat adatai alapján a falu 2000 fős lakosságának mára már csak mintegy fele vallja magát horvátnak) a horvát nyelv három változatával kerülnek kapcsolatba: a helyi grádistyei horvát archaikus dialektussal, a grádistyei standardizált nyelvváltozattal, amely a településre is eljutó burgenlandi (Ausztria) médiatermékek és a római katolikus egyház liturgikus nyelve, valamint egyénileg eltérő mértékben a standard horvát nyelvvel is. A nyelvsere generációs rétegzettséggel jellemzett folyamata előrehaladott állapotot mutat. A legidősebbek – a 70-80 évesek, valamint az annál idősebb néhány fő – kivételével, akiknek a domináns nyelve a magyar, s egyénileg eltérő mértékű a grádistyei nyelvváltozat és a standard horvát nyelv ismerete, illetve használata. „Kópháza a horvát lokális közösségek típusait tekintve kiegyensúlyozott etnikai magatartást mutató település, melyre a nemzetiségi és nemzetiségi önbevallás hasonlósága vagy csekély mértékben változó aránya jellemző” (Szarka 2003: 292).

### 1.1 A grádistyei horvátok

„A grádistyei horvátok nyelve jellemzően XVI. századi nyelvállapotot tükröz, nyelvjárásszigetet alkot, s a nyelvi közösségek egy évszázadok során kialakult regionális nyelvet beszélnek” (Szoták 2004: 43). Horváth Sándor kutatása rávilágít arra, hogy már a XVIII. századi grádistyei horvát forrásokban is felmerült a horvát nyelv „romlásának” kérdése, s az elhatárolódás vélelme a Horvátországban beszélt nyelvváltozatoktól (Horváth 2005: 181). Jelentős kutatók – köztük Zoltán András, valamint A. D. Duličenko – a grádistyei horvát nyelvet, melyhez a Kópházán beszélt horvát nyelvváltozat is tartozik, olyan elszigetelt mikronyelvnek határozzák meg, amelynek nincs anyaországa (Zoltán 2001: 139). Számos résztvevő megfigyelés tapasztalata alapján kimondható azonban, hogy bár a horvátországi és a grádistyei horvát beszélők kommunikációja sok esetben nem zökkenőmentes, a nyugat-magyarországi grádistyei horvát nyelvi közösségek a nyelvmegtartást célzó törekvések eredményeképpen számos szállal kötődnek nemcsak a közeli grádistyei horvát régióhoz, hanem Horvátországhoz is. Kimondható tehát, hogy a grádistyei horvát nyelvváltozat nem önálló nyelv, hanem a horvát nyelv egy változata. A grádistyei horvátok identitása jellemzően hármas kötődést mutat. Mária Homišinova Grbić 1987-es kutatására hivatkozva megállapítja, hogy „a magyarországi horvátok etnikai tudatának többrétegűségét három szinten határozhatjuk meg: állampolgári magyar identitás, nemzetiségi horvát identitás a Horvátországgal mint hazával való azonosulás alapján, valamint lokális-regionális horvát identitás” (Homišinova 2003: 79). Meg kell azonban állapítanunk, hogy a grádistyei horvátok esetében a lokális-regionális identitáson nemcsak a magyarországi horvátokhoz való kötődést érthetjük, hanem a burgenlandi horvát régióval való azonosulást is. Fontos ugyanakkor megjegyeznünk, hogy a magyarországinál jelentősen nagyobb tömbben élő grádistyei horvátokat Ausztriában népcsoportnak (*nacionalna grupa*) tekintik, míg a Magyarországon élő grádistyei horvátokat nemzeti kisebbségként (*nacionalna manjina*) tartják számon.

## 2 Nyelvmegtartás - lehetőségek és kérdések

Joshua Fishman művei alapvetőek a nyelvi tervezéssel, revitalizációval és a kétnyelvű oktatással kapcsolatos kutatásokban. Számos más szerzőhöz hasonlóan (Gal 1979; Bartha 1999) úgy vélte, hogy a nyelv családban történő átadásának alapvetően fontos szerepe van egy nyelv további sorsának alakulásában. Kizárólag intézményes beavatkozással nem lehet hatékonyan befolyásolni a nyelvcseré folyamatát, ugyanakkor az iskolának is nagy szerepe van. Megállapításai szerint a folyamat visszafordításához a nyelvcseré folyamatát okozó hatalmi viszonyokat kell befolyásolni. Fishman (1991) kutatásai rávilágítottak arra, hogy a nyelv veszélyeztetett állapotát minden esetben megelőzi a kulturális asszimiláció folyamata, amely alapvetően megváltoztatja a nyelv státuszát, s ez meghatározza a nyelv jövőbeni sorsát. Fishman egy nyolclepcsős skálát – *GIDS*, vagyis *Graded Intergenerational Disruption Scale* – alakított ki, amelynek segítségével megállapítható egy adott nyelv veszélyeztetettségének a mértéke, valamint a nyelvcseré visszafordítására alkalmazandó stratégia is. A projekt során egy olyan eszköztárat fejlesztettek ki, amelynek segítségével megállapítható, hogy egy adott közösségben mennyire valósul meg a nyelvmegőrzés.

A grádistyei horvát nyelvváltozat a vizsgált nyelvi közösségben Fishman nyolclepcsős skáláján megfigyeléseim alapján a 7-es fokozattal írható le. A nyelvmegőrzés mértéke azonban a településen családonkénti eltérést mutat. Fishman szerint a 7-es fokozatra az jellemző, hogy „a veszélyeztetett nyelvet beszélő idősök (*Xish speakers*) a társadalomba integráltak, otthonaikban élnek [...] az új nyelvet (*Yish*) beszélő gyermekeik, unokáik és szomszédaiuk körében. Mivel azonban már nem reprodukcióképes korúak, nem tudnak demográfiailag hozzájárulni a veszélyeztetett nyelv beszélőinek növeléséhez, azonban szociológiai értelemben sokat tehetnek a nyelv fennmaradása érdekében” (1991: 90).<sup>i</sup> Fishman úgy vélte, hogy a veszélyeztetett nyelvek esetében, habár helyzetük sok hasonlóságot mutat, más és más eljárást kell alkalmazni az egyes nyelvekre jellemző tényektől függően” (2001: 1).<sup>ii</sup> A vizsgált nyelvi közösségben a képet a standard horvát nyelv jelenléte árnyalja, hiszen jelenléte esélyt ad arra, hogy fennmaradjon a horvát nyelv a településen. Ezt a feltételezést támasztja alá Paulston (1994) megállapítása, amely szerint a standard, valamint az írott változat ismerete lassíthatja a nyelvcseré folyamatát (idézi: Borbély 2001: 203). A standard horvát nyelv nagymértékben különbözik a grádistyei horvát nyelvtől, de a beszélők kölcsönösen megértik egymást. A település általános iskolájában a standard nyelvet tanítják, s számos felnőtt horvát nyelvhasználatára is hatással van a standard horvát nyelv egyebek mellett a kulturális cserekapcsolatok, az egyéni horvátországi utazások, tanulmányok, valamint munkavállalás következményeként. Radó Péter szerint „[...] a kisebbségek kulturális arculatát nem lehet csupán egy szociológiai és gazdasági alapját vesztett népi kultúra hagyományait ápolva őrizgetni, mert így »másságuk« – mely nélkülözhetetlen a csoport határainak kijelöléséhez – csupán a vasárnap délelőtti kóruspróbán lesz érzékelhető”. A szerző továbbá kifejti, hogy „a kulturális sajátosságok adaptív újratermelése közösségi tevékenységként képzelhető csak el” (Radó 1992: 144). Fishman szintén hangsúlyozza „az iskoláskort követő, valamint az iskolán kívüli programok fontosságát a nyelvmegőrzésre irányuló

<sup>i</sup> A szerző fordítása.

<sup>ii</sup> A szerző fordítása.

tevékenységek során [...]” (Fishman 2001: 15).<sup>iii</sup> A szerző továbbá fontosnak tartja a nyelvmegőrzés céljából, hogy „a veszélyeztetett nyelvnek osztoznia kell bizonyos funkciókon az erősebb versenytárs nyelvvel” (Fishman 2001: 21).<sup>iv</sup>

Mária Homișonova megállapítása, amely szerint a nyelvi kisebbségek asszimilálódásán belül kétféle trendet figyelhetünk meg, teljességgel jellemzi a Kópházán évtizedek óta megfigyelhető tendenciát is. „[...] a kulturális szocializáció magyar nyelvűvé válik, a másik oldalon viszont erőteljesebben jelentkezik a kisebbségi önkormányzatok, a civil szféra hatása, az etnikai tudat revitalizálódása, az anyanemzet kultúrájának hatása, az anyaországhoz fűződő kapcsolatok erősítése iránti igény mindkét részről” (Homișonova 2008: 67). A Kópházán élő horvát nyelvi közösség esetében nemcsak a horvátországi kapcsolatok ápolását figyelhetjük meg, hanem a burgenlandi (Ausztria) horvátokkal történő kapcsolatok fenntartását is, akikkel a trianoni békeszerződést megelőzően egy tömbben éltek, s nyelvváltozatuk nagyon kevésbé tér el.

### 3 Adatközlők és a kutatás módszertana

A kutatás számos éven keresztül tartó résztvevő megfigyeléssel kezdődött. Az adatgyűjtés előzményeként 2004 óta számos kulturális, iskolai és egyházi rendezvényen vettem részt a nyelvi közösségben. A résztvevő megfigyelés volt a módszerem az előzetes adatgyűjtés során is, s az így kapott adatokat, információkat egészítettem ki irányított interjúkkal. A kutatást rétegzett mintavétellel, tudatosan kiválasztott mintával végeztem. A kutatás adatközlői a vizsgált közösség egyetlen általános iskolájának, a Nakovich Mihály Általános Iskola és Óvodának 26 fő 13–14 éves osztályos tanulója, valamint a nyelvi közösség 24 fő 30–60 éves tagja voltak.

Az iskolában a 2016/2017. tanévben 145 diák tanul, a pedagógusok létszáma 15. Az iskola volt az első helyi nemzetiségi kisebbségi önkormányzat által működtetett iskola az országban. Az iskola a Nemzetiségi Oktatási Hivatal bázisintézménye, amely számos kapcsolatot tart fenn horvátországi, valamint boszniai intézményekkel. Az iskola 2017-ben a horvát köztársasági elnök elismerő oklevelében részesült a horvát anyanyelv és kultúra ápolásáért.

Az előzetes tájékozódás során szerzett tapasztalatok alapján, valamint a szakirodalom feltárását követően a gyermek adatközlők számára 10 nyitott kérdés tartalmazó kérdőívet állítottam össze. A felnőtt adatközlőkkel 16 kérdésből álló irányított interjút készítettem. Kérdéseim a horvát nyelvnek az adatközlők életében, illetve a településen elfoglalt helyére, valamint a nyelvmegtartás lehetőségére vonatkoztak. Azért vélem különösen fontosnak e kérdések vizsgálatát az általános iskolás korosztály esetében, mert a közösség nyelv- és identitásmegőrző törekvései eredményességét, valamint a horvát nyelvnek a településen betöltött jövőbeni szerepét a legfiatalabb korosztály attitűdjei mutatják meg a legnyilvánvalóbban.

A teljesség kedvéért dokumentumelemzést végeztem a község önkormányzata kiadványának 2016 decemberéig megjelent 108 számáról is, valamint figyelemmel kísértem a grádistryei horvátok eseményeivel foglalkozó közösségi portálokat is.

---

<sup>iii</sup> A szerző fordítása.

<sup>iv</sup> A szerző fordítása.



## 4 Eredmények

### 4.1 A gyermek adatközlők körében végzett kutatás eredményei

A 26 válaszadó diák közül 10 nem rendelkezik horvát gyökerekkel, 16 tanulónak pedig valamelyik, illetve mindkét szülője horvát identitású. Az iskolában minden diák tanul horvátul, s az iskola, illetve a falu befogadó légkörére utal, hogy a horvát gyökerekkel nem rendelkező 10 tanuló közül 8 tanuló szeretné érteni, illetve beszélni a faluban beszélt horvát nyelvet. Mindössze 2 tanuló válaszolta azt, hogy egyáltalán nem érti meg, ha az iskolán kívül horvátul beszélnek hozzá. Mivel dolgozatom egyik célja, hogy a falu nyelvmegtartó, identitásépítő törekvései összefüggésébe helyezze felmérjem, hogy a közösség 13–14 éves, nemzetiségi horvát származású általános iskolás tanulói milyen attitűdökkel rendelkeznek a horvát nyelvvel, valamint identitással kapcsolatban, a továbbiakban a horvát nemzetiségi gyökerekkel rendelkező válaszadók attitűdjeit vizsgálom, hiszen az identitásépítő törekvések őket érintik leginkább.

A horvát nyelvhasználattal kapcsolatos eredmények a vártak megfelelőek, s a nyelvcsere folyamatának lezárulására utalnak: a válaszadó gyermekek többsége „ritkán” és „soha” nem beszél horvátul az iskolán kívül, illetve családi körben. Mindössze 2 tanuló válaszolta azt, hogy nagyszüleivel mindig horvátul beszél, s 5 tanuló értékelt úgy, hogy szüleivel, illetve nagyszüleivel rendszeresen beszél horvátul. A horvát nyelv megértésével kapcsolatban viszont sokkal kedvezőbb képet adnak a válaszok: 13 tanuló véli úgy, hogy megérti, ha horvátul beszélnek hozzá. A válaszokból azonban az derül ki, hogy a tanulókhöz nem beszélnek gyakran horvátul az iskolán kívül, s nagyon megosztanak a vélemények arról, hogy milyen nyelven válaszolnak, ha horvátul szólítják meg őket. A tanulók túlnyomó többsége úgy ítéli meg, hogy a szülei és nagyszülei örülnek, hogy ők horvátul tanulnak. A nagyszülőket érintő kérdésre a gyermekek a következő válaszokat adták többek között: „mert ők horvátok, és szeretnék, ha én is olyan lennék, mint ők”, „mert szeretnék, ha tudnék horvátul”, „mert így tudnánk horvátul beszélni”, „mert így folytatom, amit ők elkezdtek”, „mert a nagymamám horvát, és közel áll hozzám”, „azért, mert akkor én is olyan leszek, mint ők”. A szülők véleményéről feltett kérdésre jellemzően más típusú válaszokat adtak a gyermekek. A különböző generációk attitűdjéről adott válaszokban is tükröződik a nyelvcsere folyamatának generációs rétegzettsége. A válaszadó tanulók szerint a szülők azért örülnek, hogy ők horvátul tanulnak, mert így több nyelvet tudnak majd, a többi szláv nyelvet is megértik, valamint hasznos lesz munkájukban, illetve külföldi utazásokkor a horvát nyelv ismerete. Tehát a tanulók úgy érzékelik, hogy míg a nagyszülők a nemzetiségi hovatartozásuk, valamint érzelmeik miatt tartják fontosnak azt, hogy ők horvátul tanuljanak, addig szüleiket gyakorlatiasabb szempontok vezérik. Arra a kérdésre adott válaszokból, hogy a német vagy a horvát nyelv tanulása mely-e jobban, arra lehet következtetni, hogy a tanulók nem támaszkodhatnak a családi környezetből hozott horvát nyelvi ismereteikre az iskolai horvátórákon.

Az adatközlő tanulók a horvát nyelv tanulásával kapcsolatos céljaik között a következőket említették: „horváttanár szeretnék lenni”, „az iskolánk *buševci* (Horvátország) testvériskolájában tanuló diákokkal szeretném tartani a kapcsolatot”, „szeretnék horvát emberekkel beszélni”, „később nyelvvizsgát szeretnék tenni”, „ha lesz szakmám, tudok dolgozni Burgenlandban vagy Horvátországban”, „meg fogom érteni, ha horvátul beszélnek hozzám”, „most még semmi célom nincs a horváttal, de később jól jöhet”, „könnyebb lesz majd munkát találnom”, „ha nyaralni megyek, tudjak

beszélni a horvát barátaimmal”, „fontos lehet, hogy nemcsak németül és angolul tudok majd, hanem horvátul is”. Azzal a kérdéssel kapcsolatban, hogy hol beszélnek még a horvát nyelvet Magyarországon és más országokban, a diákok megfelelően tájékozottak voltak: megemlítették a grádistyei horvát településeket (a grádistyei szót szinte kivétel nélkül használták), Pécs környékét, Baját, Horvátországot, Baranyát, Ausztriát. Az adatközlő tanulók tudatában vannak annak, hogy a Kópházán beszélt horvát nyelv eltér a „más területeken beszélt horvát nyelvtől”. Az eltérést a kiejtésben, valamint a szókincsben érzékelik. Használták a nyelvjárás szót is. Arra a kérdésre, hogy megértik-e a falubeli idősebb emberek horvát beszédét, az adatközlő gyermekek a következő válaszokat adták: „nem nagyon értem, mert mi más tanulunk”, „néha megérték pár szót, s néha még németet is hallok a beszédükben”, „megértem, mert a nagymamám is így beszél, és az ő beszédét már megszoktam”, „én az iskolában a horvátországi horvátot tanulom, de általában megértem”, „nem, mert én nem tudok jól horvátul”, „nem értem, mert gyorsan beszélnek”, „elég furcsán beszélnek, de megértem”, „pár szót megérték”, „igen, megértem”, „van, amikor megértem, van, amikor nem”, „érdekes, mert néha a magyarral és a némettel is keverik a beszédüket az idősebbek”, „nem, mert hadarnak”, „igen, megértem, de ha elmegyek mellettük, átváltanak magyarra, mert azt hiszik, nem értek horvátul”.

A gyermekek tudatában vannak annak, hogy Magyarországon más nemzetiségek is élnek. Válaszaikban a német nemzetiséget, valamint a szlovákokat és a románokat említik leggyakrabban. A tanulók mintegy fele említi a szlovén és a szerb nemzetiséget. Egy tanuló horvát nyelven írta le a horvát nemzetiségi csoportokat.

#### 4.2 A felnőtt adatközlők körében végzett kutatás eredményei

A 24 fő 30–60 év közötti tudatosan kiválasztott adatközlővel 16 kérdésből álló irányított interjú készíttettem. A 16 kérdés egyebek mellett az adatközlők nyelvhasználatára, a nyelvnek a következő generációra való átörökítésére, a standard horvát nyelv szerepére a beszélő közösség életében, a horvát nyelvvel kapcsolatos esetleges pozitív és negatív élettapasztalatokra, a kulturális rendezvényeknek a nyelvmegőrzésben elfoglalt szerepére, a horvát nyelv presztízsének változására, valamint az egyéni nyelvmegőrzés lehetőségeire vonatkoztak.

Az adatközlők a horvát nyelvet jellemzően az idősebb családtagokkal és a falubeliekkel történő kommunikáció során beszélnek. Utazással és munkavégzéssel kapcsolatban az adatközlők csak mintegy negyede használja rendszeresen a horvát nyelvet. A válaszadók több mint fele úgy nyilatkozott, hogy gyermeke valamennyire ért horvátul, míg 4-en úgy ítélték meg, hogy gyermekük nagyon jól beszél a nyelvet. Az utóbbi adatközlők gyermekei felsőfokú tanulmányokat folytattak horvát nyelven Magyarországon, illetve Horvátországban. Arra a kérdésre, hogy beszélnek-e a standard horvát nyelvet, 20 válaszadó azt nyilatkozta, hogy bizonyos fókig érti azt, míg 4 adatközlő úgy ítélte meg, hogy jól beszél a Horvátországban beszélt „irodalmi” horvát nyelvet. Az utóbbi adatközlők horvát tanulmányokat folytattak közép- és felsőfokon. A horvát nyelvvel kapcsolatos pozitív és negatív tapasztalatokkal kapcsolatban megoszlottak a vélemények. A negatív élmények jellemzően középiskolás korukban érték a válaszadókat diáktársaik részéről, a pozitív érzéseket pedig a horvát nyelvismeret miatt szerzett sikerélmények okozták. A válaszadók többsége volt már turistaként, illetve kulturális cserekapcsolatok során Horvátországban, ahol többnyire sikeresnek ítélték meg horvát nyelven történő kommunikációjukat. Négy válaszadó

nyilatkozta azt, hogy kommunikációjában frusztrálta a grádistyei horvát nyelv és a standard horvát nyelv között meglévő különbség. Húsz válaszadó találkozott már burgenlandi horvátokkal, s problémamentesnek ítélte meg a velük történt beszélgetést. A válaszadók kivétel nélkül találkoztak már Magyarországon élő horvátokkal, de jellemzően magyarul beszélgettek velük. Úgy ítélték meg, hogy a környező falvakban (Fertőhomok, Hidegség) élő magyarországi horvátok kevésbé jól beszélnek horvátul, mint a kópházaiak.

A válaszadók kivétel nélkül fontosnak tartották, hogy megmaradjon a horvát nyelv jelenléte a településen, de megoszlottak a vélemények az iskolai horvátnyelv-tanítás sikerét illetően. Az adatközlők 4 kivétellel úgy vélték, hogy emelkedett a horvát nyelv presztízse a településen. Az 50–60 éves válaszadók úgy nyilatkoztak, hogy gyermekkoruk óta jelentősen nőtt a horvát nyelv presztízse Kópházán. Arra a kérdésre, hogy aktivizálták, fejlesztették-e horvát nyelvtudásukat életük folyamán, 2 fő 60 év körüli válaszadó azt válaszolta, hogy mivel a nyelvet csak a településen használja, nem érezte szükségesnek, hogy fejlessze azt. A többi válaszadó szükségét érezte élete során, hogy jobban elsajátítsa a horvát nyelvet. Három válaszadó a munkájában használja a grádistyei horvát nyelvet, 19-en a horvátországi horvátokkal történő sikerebb kommunikáció miatt érezték szükségét annak, hogy fejlesszék horvát tudásukat. Az adatközlők fontosnak nevezték a horvát vonatkozású kulturális eseményeket, s szinte kivétel nélkül megemlézték a Kópházi lakodalmas (*Koljnofski pir*) című színdarab műkedvelő előadását.

#### 4.3 A nyelvi kontaktusok hatása a nyelvi közösség nyelvhasználatára

A kópházi horvát nyelvi közösség nyelvhasználatának feltehetően a standard horvát nyelv hatására bekövetkezett változásáról a grádistyei horvát nyelvnek egy, az USA-ban élő beszélője beszámolójából alkothatunk képet. Chicagóban egy kisebb Kópházáról kivándorolt kolónia él, amelynek még életben lévő tagjai rendszeresen tartják a kapcsolatot, s jellemzően horvátul kommunikálnak egymás között. A 60 éves női adatközlő szülei 1956-ban hagyták el Kópházát, a vizsgált települést. Szülei ahhoz a generációhoz tartoznak (jelenleg 80 évesnél idősebbek), melynek képviselői az egymás közötti kommunikációjukban egész életük során kizárólag a grádistyei horvát nyelv szülőfalujukban beszélt változatát használták, de kiválóan beszéltek magyarul is. Az adatközlő által küldött videofelvételek alapján elmondható, hogy édesanyja grádistyei horvát nyelvi kompetenciái semmiben sem különböznek a vele egykorú, mindvégig Kópházán élt személyek nyelvi kompetenciáitól. Az adatközlő angol domináns nyelve mellett megtanult szüleitől horvátul is, magyarul azonban nem. A standard horvát nyelvről nincsenek ismeretei. Négy alkalommal járt Kópházán. Először 1963-ban kisgyermekként, majd 1972-ben, 1977-ben és végül 2000-ben.

Az adatközlővel a burgenlandi, illetve nyugat-magyarországi horvátok eseményeit naprakészen követő *Gradišćanski portal* elnevezésű közösségi oldalon folytattam levelezést, majd ezt követően készítettem irányított interjút grádistyei horvát, illetve részben angol nyelven. A 60-as éveiben járó asszony spontán módon arról számolt be, hogy utolsó látogatása során azt tapasztalta, hogy nehezebben értette meg a grádistyei horvát közösség beszédét, miközben szüleivel változatlanul grádistyei horvát nyelven kommunikált az USA-ban.

Valószínűsíthető, hogy ennek az az oka, hogy a vizsgált településen a mára már kis létszámú legidősebb korosztállyal ellentétben – akiknek domináns nyelve a helyi horvát

dialektus – a fiatalabb korosztályok horvát nyelvhasználatára jelentősen hatott a standard horvát nyelv. Számos éven keresztül történt megfigyelés alapján az a következtetés vonható le, hogy minél intenzívebben vesz részt a horvát nyelvi közösség tagja a horvátországi kulturális cserekapcsolatokban, annál erőteljesebben érződik nyelvhasználatában a standard nyelv hatása. Az adatközlővel készített interjú, akinek rokonai ahhoz a körhöz tartoznak a településen, akik aktívan részt vesznek a közösség nyelvmegtartó törekvéseiben, s gyakran vannak kontaktusban a standard horvát nyelvvel, igazolta a feltevésemet.

A nyelvi közösségben megfigyelhető, hogy a közösség és az egyének nyelvmegtartó törekvéseinek jelentős alkotóeleme a standard horvát nyelv elsajátítása. A településre 1983-ban érkezett az első kulturális csoport a Zágráb környéki *Buševec* településről, melynek tagjait jellemzően horvát ajkú családok szállásolták el, akik a viszontlátogatás során élvezték a horvátországi horvátok vendégszeretetét. A kölcsönös látogatások rendszeressé váltak az évek során, s számos baráti kapcsolat is kialakult a magyarországi és horvátországi családok között. Ezek a kapcsolatok elősegítették a különböző nyelvváltozatok használatát a településen, s valószínűsíthető, hogy hatással voltak az egyének horvát nyelvhasználatára. A Horvátországból érkező vendégtanárok, valamint a diákok horvátországi és boszniai testvérfalvakba történő utaztatása is a standard horvát nyelv elsajátítását segítik elő.

#### 4.4 A dokumentumelemzés eredményei

A teljesség kedvéért dokumentumelemzést végeztem a *Kópházi Hírek* 1995 és 2016 között megjelent 108 számáról, amely egyebek mellett részben magyar, részben grádistyei és standard horvát nyelven ad hírt az identitásépítést és nyelvmegőrzést célzó kulturális eseményekről, a horvát kisebbségi önkormányzat rendezvényeiről, valamint a település iskolájának és óvodájának életéről, beleértve a horvát nyelv tanításának helyzetét is.

A nyelvmegtartást és identitásépítést segítő törekvések az intézményes támogatások mellett főként a helyi aktivisták tevékenységének köszönhetően valósulnak meg. A település minden háztartásába ingyenesen eljutó *Kópházi Hírek* tudósításaiból nyomon követhetjük a nyelvi közösség különböző szervezeteinek kapcsolatait, közös rendezvényeit a horvátországi és boszniai, valamint a burgenlandi (Ausztria) és magyarországi horvátokkal. A *Kópházi Hírek* az elmúlt több mint húsz év során számtalanszor számolt be Kópháza testvérfalvainak (a Zágráb melletti *Buševec*, a Zadar melletti *Bibinje*, valamint a boszniai *Kiseljak*) és a község iskolájának, valamint kisebbségi önkormányzatának és kulturális szervezeteinek közös rendezvényeiről. A kulturális csoportok kölcsönös látogatása 1983 óta rendszeres. A csoportok tagjai többnyire családtagjaikkal együtt vesznek részt a kölcsönös látogatásokon, ami lehetővé teszi a személyes kapcsolatok kiépítését, valamint a vizsgált nyelvi közösség kontaktusát a standard horvát nyelvvel.

A *Kópházi Hírek* rendszeresen hírt ad a helyi általános iskola tanulóinak utaztatásáról a testvérfalvak iskoláiba. A cserekapcsolatokról horvát nyelvű tanulói beszámolókat is olvashatunk a helyi kiadványban. A lap tudósít az iskola pedagógusainak horvátországi továbbképzéséről, valamint a horvátországi tanárok és pedagógusjelöltek tevékenységéről a település két tannyelvű általános iskolájában. A *Kópházi Hírek* beszámolt a nyelvi közösség kiemelkedő eseményéről, a Kópházi lakodalmas (*Koljnofski pir*) című műkedvelő színdarab előkészületeiről, valamint

előadásáról is. A falu népszokásairól szóló színdarabot mintegy 50 helyi aktivista részvételével, valamint horvátországi rendező irányításával adták elő nemcsak Kópházán, hanem a környező magyarországi és ausztriai grádistyei horvát településeken is.

A *Kópházi Hírek* beszámol a településnek a burgenlandi (Ausztria) grádistyei horvátokkal fenntartott kapcsolatairól is. Ezek közül kiemelkedik a közös zarándoklatok szervezése, a grádistyei horvát katolikus papok misézése egyházi ünnepek alkalmával, a színjátszó csoportok kölcsönös fellépése, a helyi óvoda pedagógusainak továbbképzése (amely 20 éves hagyományra tekint vissza), valamint az óvoda együttműködése a füles (németül: *Nikitsch*, horvátul: *Filež*) óvodájával.

Az aktivisták által szerkesztett lapból nyomon követhetjük a település szervezeteinek a magyarországi grádistyei horvátokkal fenntartott kapcsolatait is, melyek közül a legjelentősebbek az iskolai versenyek, valamint olvasótábor és tamburatór szervezése a közösség, valamint a grádistyei horvát falvak iskolásainak (Bezenye, Kimle, Und, Horvátzsidány, Szentpéterfa), a színjátszó csoportok kölcsönös fellépése, a sportversenyek szervezése, valamint a magyarországi grádistyei horvát papok misézése a település katolikus templomaiban.

A helyi lap rendszeresen jelentet meg horvát nyelvű cikkeket a település történetéről, hagyományairól, s tesz közzé horvát nyelvű verseket a falu szülöttének, a jelentős grádistyei horvát költőként számon tartott Sinkovits Mátyásnak (*Mate Šinković* 1927–1972) tollából.

## 5 Összegzés

A kisebbségi identitás megőrzésében alapvető szerepe van az adott nyelv ismeretének. Mivel a magyarországi horvát nemzetiség közösségeire – így az általam vizsgált közösségre is – a horvát nyelv ismeretét, illetve használatát illetően a generációs rétegzettség a jellemző, a nyelvmegtartás, illetve az identitásépítés egyetlen eszköze az iskolai oktatás, vagyis a horvát felmenők nyelvének újratanítása. A Nakovich Mihály Általános Iskola horvát nemzetiségi gyökerekkel rendelkező 13–14 éves korú tanulóinak válaszaiból arra lehet következtetni, hogy az iskola tanárai nagy erőfeszítéseket tesznek a horvát identitás és nyelv megőrzésnek céljából, hiszen a tanulók attitűdjei összességében pozitívak. A kapott adatok kvalitatív elemzése arra enged következtetni, hogy a tanulóknak sok élményt nyújt a horvát nyelv tanulása, különösen a tanórán kívüli programok: a táborok, a tamburaszakkör, a horvátországi vendégtanárokkal történő kommunikáció, a horvátországi látogatások, a horvát nyelven történő kommunikáció öröme, a horvátországi horvát kortársakkal történő találkozás. A diákok szinte kivétel nélkül úgy ítélik meg, hogy szüleik és nagyszüleik örülnek annak, hogy ők horvátul tanulnak. Néhány tanuló családi körben is használja esetenként a horvát nyelvet. Az iskolában végzős évfolyamonként 5–6 tanuló alapfokú nyelvvizsgát tesz a standard horvát nyelvből.

A grádistyei horvát nemzetiségű lakossággal rendelkező Kópháza esetében – ahol a generációs rétegzettséggel jellemzett nyelvváltás csaknem teljes mértékben lezajlott, s a nyelvmegtartást nehezíti az a tény, hogy a faluban beszélt horvát nyelvváltozat jelentős mértékben különbözik a horvát köznyelvtől – a leírtak alapján elmondható, hogy a helyi szakemberek és aktivisták jó eredményeket érnek el az identitás megőrzése, valamint a nyelvmegtartás terén. Az Országos Horvát Önkormányzat szerint jellemző az általános iskolai kisebbségi nevelésre, hogy „általában nagyon

pozitív szándékot és kis hatékonyságot tükröz” (Bindhoffer et al. 2011: 71). Ezzel a megállapítással nincs okunk vitába szállni, azonban a realitásokból kell kiindulnunk. Az adott helyzetben a cél a nyelvi sokszínűség fenntartása, valamint a nemzetiségi horvát identitás megőrzése. A leírtak alapján megállapíthatjuk, hogy a község különböző szervezeteinek, valamint az általános iskola pedagógusainak törekvései nem eredménytelenek.

Az eredmények továbbá azt igazolták, hogy a vizsgált településen a horvát származású, valamint a nyelvet valamilyen szinten beszélő egyének identitásának szerkezete, valamint nyelvtudása eltérő. Az eltérések több tényezőre vezethetők vissza. Függnek egyrészt a kutatásba bevont személy életkorától, valamint attól, hogy milyen mértékben beszéltek élete folyamán családjában a horvát nyelvet. A kutatás eredményei alapján elmondható, hogy a középkorú generáció azon tagjai esetében, akik tevékenyen részt vesznek a horvát kisebbség kulturális életében, valamint rendelkeznek horvátországi vagy burgenlandi horvát kapcsolatokkal, kimutatható mértékű nyelvi revitalizáció figyelhető meg. Az iskolás korosztály pedig a horvát nyelv standard változatának megtanulása útján kap lehetőséget a horvát identitás megőrzésére. Összességében tehát tetten érhető a községben a nyelvmegtartás és az identitásépítés folyamata.

## Irodalom

- Bartha, Cs. 1999. *A kényelvűség alapkérdései*. Budapest: Nemzeti Tankönyvkiadó.
- Bindhoffer, Gy., Főrika L., Kutassy-Nagy N., László B., Szabó O., Vass V. 2011. *Jelentés a nemzeti és etnikai kisebbség általános iskolai nevelés-oktatás helyzetéről*. Elérhető: [www.kisebbségiombudsman.hu/data/files217986220.pdf](http://www.kisebbségiombudsman.hu/data/files217986220.pdf). Letöltve: 2013. április 8.
- Borbély, A. 2001. A Nyelvcseré és folyamata kutatása. *Nyelvtudományi Közlemények*, 98: 193–215.
- Crystal, D. 1998. *The Cambridge Encyclopedia of Language*. Cambridge, New York: Cambridge University Press.
- Fishman, J. A. 1991. *Reversing Language Shift. Theoretical and Empirical Foundations of Assistance to Threatened Languages*. Clevedon, Philadelphia, Adelaide: Multilingual Matters LTD.
- Fishman, J. A. 2001. Why is it so Hard to Save a Threatened Language? In: Fishman, J. A. (szerk.) *Can Threatened Languages Be Saved?* Clevedon–Buffalo–Toronto–Sydney: Multilingual Matters LTD.
- Gal, S. 1979. *Language Shift. Social Determinants of Linguistic Change in Bilingual Austria*. New York – San Francisco – London: Academic Press.
- Grbić Jakopović, J. 1994. *Identitet, jezik, razvoj*. Zágráb: Institut za etnologiju i folkloristiku.
- Homišinoва, M. 2008. *Identitás, nyelvhasználat, asszimiláció. Etnikai folyamatok magyarországi kisebbségi családokban*. Budapest: MTA Kisebbségkutató Intézet – Gondolat.
- Horváth, S. 2005. A grádityei horvátok XVI–XX. századi asszimilációjának példái. *Kisebbségkutatás*, 14(2): 181–193.
- Paulston, C. B. 1994. *Linguistic Minorities in Multilingual Settings. Implications for Language Policies*. Amsterdam, Philadelphia: John Benjamins.
- Radó, P. 1992. A nemzeti kisebbségek nyilvános nyelvhasználata Magyarországon. *Regio*, 3(2): 135–146.
- Soksevcics, D. 2013. A magyarországi horvátok rövid története. Elérhető: <http://www.horvatok.hu/hu/magyarországi-horvatok/59-toertenelem>. Letöltve: 2015. május 2.
- Szarka, L. 2003. Etnikai változások a déli szláv kisebbségeknél. In: Kovács, N., Szarka, L. (szerk.) *Tér és térkép. Tanulmányok az etnicitás és az identitás kérdésköréből II*. Budapest: Akadémiai Kiadó. 285–312.
- Szoták, Sz. 2004. Veszélyeztetett nyelvek. *Kisebbségkutatás*, 13(1): 37–49.

Zoltán, A. 2001. A kisebbségi szláv nyelvek szociolingvisztikai tipológiája. *Kisebbségkutatás*, 10(1): 139–142.

### **Források**

*Kópházi Hírek. Kópháza Önkormányzatának tájékoztatója* (1995–2016). Elérhető: <http://www.kophaza.hu/magyar/kophazi-hirek-koljnfoski-visti>. Letöltve: 2016. november 1.

## Pokemon-go ポケモン語 – avagy hogyan kommunikálnak a Pokemonok?

Kaba Ariel

ELTE BTK Nyelvtudományi Doktori Iskola  
leira.abak@yahoo.com

**Kivonat:** A jelen írásban tárgyalt popkulturális produktumok, illetőleg médiumok megjelenése a digitális korszak paradigmaváltásának természetes hozadéka, amelyeket nemcsak érdekes vizsgálni, hanem egyenest a kutatási palettára szükséges venni, hiszen a populáció túlnyomó részének meghatározzák a mindennapjait – egyelőre reflektálatlanul, ugyanis sem a tudománynak, sem az oktatásnak nem fókusza a popkultúra, a virtuális kommunikáció kutatása, implementálása. Jelen kutatás célja, hogy kiderítse, hogyan kommunikál egy fiktív kommunikátorcsoport, esetünkben a Pokemonok. A feldolgozás módszertana a szerző kiterjesztettkommunikáció-elméletén és a karakterteóriáján alapul, annak okán, hogy a tudomány hagyományos módszerei nem adekvátak. A több mint húszéves Pokemon-franchise sikerét többnyire a karakterizáció jelenségében látjuk: valamennyi Pokemon karakteres a vizuális, az auditív, a nyelvi és a nem nyelvi reprezentációjában is. A vizsgálat elsősorban az anime hagyományos kommunikációjára fókuszál, lévén ez a leggazdagabb – szemben a videojátékokkal és a mangával. Az elemzésből kiderül, hogy a Pokemon és a Pokemon-nyelv, azaz a Pokemon-karakterlektus esszenciája a karakter, ugyanis prototipikus karakterjelölők útján kommunikálnak.

### 1 Bevezetés

2016 nyarán újra meghódította a világot a Pokemon-jelenség, ezúttal azonban szélesebb közönséget tudott megszólítani, mint korábban az animével, a játékfigurákkal és a konzoljátékokkal. Kutatásunk révén bebizonyosodik majd, hogy a tudományt is képes megszólítani.

Jelen tanulmány azonban nem az **elplatformosodott** Pokemon-GO elemzését tűzte ki célnak (l. Székely 2017), hanem a Pokemon-nyelv, azaz *pokemon-go* ポケモン語 vizsgálatát. A 'nyelv' jelentését tágan értelmezzük, így a Pokemon-világ nyelvhasználatát, illetőleg kommunikációját értjük alatta.

Az emberi és az állati kommunikációról is vannak kutatások, a reális és a digitális kommunikációról is, viszont a virtuális kommunikációról – amely elsődlegesen a popkultúra produktumaira jellemző, esetünkben egy elképzelt világ elképzelt lényeiének fiktív interakciójáról – nincsenek. A Pokemonokról is csak általános, felszíni leírásokat



lehet találni az angol lokalizációra fókuszálva; holisztikus, tudományos igényű munkát azonban nem.

Míthogy nem természetes nyelvről, de nem is mesterséges nyelvről és kommunikációról van szó, a felvázolt jelenség értelmezése kihívás mind a nyelvtudomány, mind a kommunikációtudomány számára. Annál is inkább, hiszen a tudomány hagyományos módszereivel nem lehetséges érdemben feltárni a Pokemon-jelenséget; még a (nemzetközileg) újszerű, s (Japánban) felkapott **szerepnyelvezet** (vö. Kaba 2016) sem képes a feladat megoldására.

Ennélfogva a kutatás felépítése, módszere, elmélete és eszköztára nem a konvencionális mintát követi, hanem jellegeből fakadóan innovatív és úttörő.

### 1.1 A kutatás célja

Jelen komplex karakterkutatás célja, hogy tudományos elméleti keretet és gyakorlati eszköztárat adjon a virtuális kommunikáció értelmezésére, jelen esetben a Pokemon-kommunikációra. A Pokemonok kommunikációját a kiterjesztett kommunikáció elméleti keretrendszerén belül értelmezzük a karakterteória eszközeit felhasználva, statikus és dinamikus médiumok alapján.

A kutatás fő kérdése, hogy hogyan kommunikálnak a Pokemonok, s hogy milyen módon értelmezhető ez a fiktív kommunikáció?

### 1.2 Módszer

Interkulturális nyelvészeti megközelítést alkalmazunk: Több nyelven és több médiumon analizáljuk a (nyelvi és nem nyelvi) Pokemon-kommunikációt, az I., a II. és a III. generációra nézve, így átfogó eredményeket közölhetünk majd.

A Pokemon-franchise valamennyi területét górcső alá vesszük eredeti japán és lokalizált koreai, angol és magyar változatban (l. 1. táblázat).

| Médium     | Nyelv        | Feldolgozott anyag    |
|------------|--------------|-----------------------|
| videojáték | jp/ko/eng    | >60 óra               |
| anime      | jp/ko/eng/hu | 147 epizód (×20 perc) |
| manga      | jp/ko/eng    | 22 kötet              |

1. táblázat. A felhasznált korpusz

A korpusz manuálisan került kiépítésre, annotálásra, és feldolgozásra is; a kutatás fókusza szolgált kritériumrendszerként: a fókusz perspektívájában releváns jelenségeket gyűjtöttük össze.

Az összegyűjtött anyagnak csak egy töredékét közöljük formai korlátozások miatt, a kutatás poszterén (Kaba 2017) azonban további szemléltető-példa látható. Az elemzett produktumok listája az irodalomjegyzékben található (l. Források).

## 2 A kiterjesztett kommunikáció

Még manapság is meghatározó Roman Jakobson kommunikációmodellje (1960); az oktatásban és a kutatásokban is alkalmazzák. Köztudott tehát, hogy a kommunikációban van feladó és vevő, köztük egy csatorna, amelyen kódolás történik, egymást követő sorrendben.

Meglehetősen régi ez a modell, a régiségnek valamennyi értelmében. Elavulttá teszi a kifejlesztése óta eltelt idő, a mai csúcstechnológia, a különféle médiumok és a mai ember kommunikációs szokásai. A naiv nyelvhasználó és a módszeres kutató egyaránt érzi, hogy valami változás történt az utóbbi néhány évtizedben, ezért is a szóbeliség–írásbeliség dilemmája (Ong 1982), és a terminológiai útvesztő (l. Veszelszki 2013: 435), mint például „másodlagos szóbeliség” vagy „új beszélnyelviség”, és így tovább.

A megnevezési problémákat félretéve elmondhatjuk, hogy megtörtént a digitális kommunikáció térhódítása. Nem elsődlegességről vagy másodlagosságról van szó, hanem a kommunikáció esszenciájában bekövetkezett változásról, azaz összemosódik a határ az (élő)beszéd és az írás közt, egyfajta „élőírás” vagy „beszédírás” figyelhető meg a technika vívmányainak révén.

A kép viszont még így sem teljes. A digitális kommunikációra (digilektus, l. Veszelszki 2010) még csak-csak értelmezhető Jakobson modellje, amit azonban sem Jakobson modellje, sem a digilektus elmélete nem magyaráz, az a virtuális kommunikáció.

A jakobsoni modell lefedi a hagyományos értelemben vett kommunikációt, a digilektus a digitális úton történőt, viszont mindkét teória kihagyja – a mára már – az egyik legmindennapibb és legelemibb kommunikációt, a virtuálisat. Egyáltalán nem új dologról van szó: minden kommunikáció virtuális, ami nem reális, tehát a videojátékoktól kezdve a televízió keresztül a szépirodalomig minden. A virtualitás abban nyilvánul meg, hogy egy közbelső médiumon keresztül, ugyanakkor ennek determinálva, kontrollált körülmények közt, nem egy térben, nem hagyományos módon történik a kommunikáció.

Azt a holisztikus kommunikáció-rendszert, amely képes a hagyományos, a digitális és a virtuális kommunikáció értelmezésére, a bennük fellelhető valamennyi kommunikatív tényezővel (audiovizualitás, meta-jelenségek<sup>1</sup>), **kiterjesztett kommunikációnak** nevezzük el. A kiterjesztett valóság mintájára a kiterjesztett kommunikáció olyan információkat aktualizál, amelyek bár jelen vannak, látszólag rejtettek.

A kiterjesztett kommunikáció a kommunikációs folyamatban detektálható tartalmak (kommunikációtartalom) és a benne lezajló események (hagyományos verbális/nem verbális/meta- és audiovizuális kommunikáció), s a kommunikáció létrejöttében esszenciális tényezők (médiummeghatározottság: virtuális, digitális és reális kommunikáció), valamint a kommunikátort (aktor/inaktor) meghatározó tényezők

---

<sup>1</sup> Lévéen integrálatlan, s nem eléggé *meta* egyik létező felfogás sem, a dolgozatban felvázolt paradigma a *meta*- előtagra nem a konvencionálisan vett 'metakommunikatív' jelentést (vö. [2/b.2]), illetve nem is a sokszor 'reflexió'-ként referált értelmezést alkalmazza (vö. metanyelv), hanem a *meta*- szó 'túl' jelentését: 'egy dolgon *jóval* túli, illetőleg kívül eső' nehezen megragadható jelenséget ért alatta a kiterjesztett kommunikáció rendszerén belül (vö. [2/meta-kommunikáció]). A terminus mögötti értelmezéskülönbséget kötőjellel tesszük explicitté: „meta” mint konvencionális vs. „meta-” mint kiterjesztett paradigma.

(szociokulturális identitás és karakter) strukturált, komplex, holisztikus kommunikációelméleti keretrendszere.

A kiterjesztett kommunikáció alapvető jellemzője, hogy megkülönbözteti az aktort és az inaktort, amelyek kommunikációtartalmakkal operálnak, kommunikációs vehikulumok révén. Tehát a kommunikációban nem feltétel az intencionáltság, sem az üzenet kódoltsága, sem a motivált kommunikatív szereplő (aktor), például a „beszélő” megléte sem, mert vannak inaktorok, amelyek pusztán a létükkel közölnek információt, azaz kommunikációtartalmat a kommunikációs vehikulum kommunikatív jellegének és információtartalmának függvényében.

A rendszer részei, egyben kutatható területei tehát a következők:

- I. konvencionális kommunikáció
  - a. nyelvi kommunikáció
  - b. 1. intencionált nem nyelvi: dinamikus és statikus nem nyelvi kommunikáció
  2. nem intencionált nem nyelvi: metakommunikáció
- II. auditív kommunikáció
  - a. hangok, zörejek
  - b. vokális kód
  - c. zene
- III. vizuális kommunikáció
  - a. statikus vizuális kommunikáció
  - b. dinamikus vizuális kommunikáció
- IV. szenzitív (tkp. érzékalapú) kommunikáció
  - a. szagok szaglás útján
  - b. anyagok és textúrák tapintás útján
  - c. atmoszféra érzés útján
- V. meta-kommunikáció
  - a. virtuális és reális kommunikáció
  - b. kultúrákőzi és médiumkőzi meta-kommunikáció
  - c. szociokulturális identitás

A kiterjesztett kommunikáció eddig ismert attribútumai közé tartozik az érzelem és a karakter. A kommunikáció attribútumai metajelenségek: valamennyi kommunikációs formában aktualizálódhatnak. Jelen kutatásban ez utóbbival foglalkozunk.

## 2.1 A karakter mint attribútum

A karakter jelentése triviális: filmek, mesék, könyvek szereplői a karakterek. Jelenthet írásjelet, írásjegyet is, vagy egyfajta jellegzetességet fejez ki egy italtra, ételre vagy akár emberre nézve; természettudományos munkákban rendszeres téma a karakterizálás.

A **karakter mint attribútum** azonban olyan jelenség, amely identifikálja, jellegzetessé, illetőleg egyedivé teszi a kommunikátort. Látható, hogy a karakterteória megalkotása során a 'karakter' magjelentését applikáltuk. A karakter lehet a nyelv és a kommunikáció attribútuma is, attól függően, hogy mi a vizsgálat fókusza.

Általában véve a karakter a kiterjesztett kommunikáció esszenciális és konkomitáns attribútuma, s mint olyan képes a kommunikáció velejáróinak értelmezésére (szituatív komponens), és képes generatív módon jelentést tulajdonítani a kommunikátorra nézve (emergens komponens).

A karakter attribútumát karakterjelölők realizálják. A jelölés történhet nyelvi úton, auditív vagy vizuális úton, ugyanakkor metaszinonim is. Ennek megfelelően beszélünk nyelvi karakterről, audiokarakterről, vizuális karakterről, illetve metakarakterről.

A karakterjelölők, s maga a karakter funkciójának meghatározása körülményes és illuzórikus a hagyományos módszertanok perspektívájában – akárcsak a diskurzusjelölők esetében a diskurzusszervezés mint funkció –, hiszen a karakterjelölők karakterépítési vagy karakterreprezentációs funkciót töltenek be, míg maga a karakter attribútumának funkciója az identitásalkotásban, a jellegzetes karakterminőségekben és az azonosítás jelenségében merül ki. Ezen kritériumok elkerülhetetlenül is tágak, homályosak az elmélet szintjén, viszont nem lehet konkrétan megfogalmazni őket, lévén a karakter emergens meta-nyelvi (vs. metanyelv), meta-kommunikációs (vs. metakommunikáció) kategória, így csak az adott szituációban és a szűkebb-tágabb kontextusban (*setting*) születnek meg a karakterminőségek és –relációk, azaz generálódik a karaktermodell.

Ennélfogva nem meglepő, hogy a karaktert kifejező tényezők tárházát a hagyományos nyelv- és kommunikációtudományi elemzőmódszerekkel nem lehet érdemben feltárni. A kommunikációban nem a résztvevők, vagy a nyelvi-, illetőleg nem nyelvi üzenet nyelvalapú szemantikai felszíni elemzése során kerül fény a karakterre, hanem a hagyományos kommunikáción túli kommunikáció, a meta-kommunikáció változói generálják a karakterjelentést.

A karakterjelentés lényegét tekintve emergens, azaz dinamikus jelentésképzés-folyamat eredményeképpen artikulálódik, szemben a konvencionális, szótári, nyelvstruktúrán alapuló szemantikai jelentésekkel. Ez azt jelenti, hogy tanulni kell, azaz fel kell fedezni ezeket a jelentéseket egyrészt a meglévő nyelvtudásunk, illetőleg kommunikatív kompetenciánk segítségével, másrészt a világról való ismereteinkre alapozva, mint a kultúra, értékek, popkulturális produktumok, érzések, és így tovább. Egyszóval karakterkompetenciára van szükségünk, amelynek révén felismerhetjük a karakterszenzitív elemeket.

Minden kommunikációtartalom potenciálisan karakteres, viszont nem arányosan karaktererősek. Minél több (karakter)referenciája van egy adott tartalomnak, annál kevésbé karakteres a karakterskálán; például a 'mindenki' vagy a 'hordó' szó kevésbé karakteres a „Civilizált/Kulturált jó estét kívánok!” (Puzsér 2016/2017: passim) frázishoz mérve. A karakter emergens jellegéből fakadóan azonban mindez nem jelenti azt, hogy az előbbi példák nem karakteresek (vö. sorjában: Deák 2016; PewDiePie 2012).

A karakter identifikáló és minősítő funkciója rávilágít a karakter hasznosíthatóságára is, ugyanis a legtöbb karakterjelölőt, karakteres jelenséget a reklámok, a különféle intézmények, társadalmi csoportok, és popkulturális produktumok termelik. A karakter tehát, a hatáskeltés és a befolyásolás révén, a legtágabb értelemben vett marketing alapvető stratégiai eszköze, s a marketolinguisztika (vö. Veszelszki 2014) egy bástyája is.

Nem szabad figyelmen kívül hagyni a manipulatív, akár destruktív felhasználási módját sem a karakternek. Az emberiség történetében az egyik legszörnyűbb karakterizáció a náci uralom alatt történt: két ősi vallási szimbólumot választottak a „jó” és a ’rossz’ (alternatív) karakterjelölőjének”, továbbá sok más karakterjelölő révén, mint például a ruházat, kézmozdulat, egyfajta náci karaktert építettek ki.

Annak ellenére, hogy eddig nem ismert teóriáról van szó, mindenkinek van tapasztalata arról, hogyan hat egy-egy márkanév és reklám, vagy a szimbólumok, mottók, szlogenek, de ugyanez a hatás érezhető egy-egy filmmel, könyvvel, tv-show-

nál: jellegzetes, egyedi sajátosságokkal vannak teli, amelyek a legtöbbször tudat alatt fejtik ki a hatásukat, még akkor is, ha nem implicit, hanem explicit tényezőkről van szó, mint az audiovizualitás.

Ilyen tényező, tehát karakterjelölő lehet egy mottó, egy szlogen, egy mém, egy melódia, egy dallam, egy szín, egy szó, egy frázis, egy szólás, egy cselekvés, egy ruha, egy stilisztikai eszköz, egy retorikai alakzat, és így tovább – a potenciális karakterjelölők vég nélkül sorolhatók, különösebb megkötések nélkül.

Ezen indirekt módon intencionált metaelemek akkor is hatnak ránk, ha nem vesszük észre, vagy ha nem akarjuk. A karakterteória tehát azért is hasznos, mert tudatosítja, detektálja és értelmezi a karaktert és a karakterizáció folyamatát egyaránt – természetesen a kiterjesztett kommunikáció keretében.

## 2.2 A karakterteória

A fentiekben megfogalmazott koncepció abban tér el az eddigi karakterképzetektől, hogy esetünkben nyelvi és kommunikációtudományi alapozású instrumentalizált elméleti keretrendszerrel van szó, amely emergens lényegéből fakadóan képes a dinamizmus megragadására. Így az idiolektustól kezdve a marketingstratégia kommunikációdimenzióját is képes leírni, illetve értelmezni.

Jelen teória többek közt azért képes erre, mert a stilisztikától, retorikától, szerepnyelvezettől és más nyelvváltozatoktól, valamint egyéb dependens és determinált módszerektől eltérően nem áll meg a makro- és mezoszinten, tehát felületi a priori tényeken, mint például a konvencionális jelentés elemzése, vagy a szöveg alaki és nyelv(tan)i jellemzőinek elemzése a hagyományos kommunikáció keretén belül, hanem egyaránt mélyebbre is (idiolektus), s följebb is (meta-kommunikáció) merészkedik, amely dimenziók lényegi megismerése csak a posteriori módon történhet.

A karakterteória a kiterjesztett kommunikáció keretén belül – így a hagyományos kommunikáció keretén belül is – operacionalizálható, akár metaszinten is, mivel a kommunikáció valamennyi területén fellelhető a karakter attribútuma.

A karakterteóriában megkülönböztetünk komplex és részleges karakterkutatást. Az utóbbi lehet (kiterjesztett vagy konvencionális) karakterkommunikáció-kutatás, amelynek célja a karakterépítés vizsgálata verbális és nem verbális úton; kiterjesztett karakterkutatás, amelynek célja a karakterépítés és -representáció kultúra- és médiumközi vonzatainak a megvilágítása; illetve metakarakter-kutatás, amelynek célja a metaelemek karakterre és kommunikációra tett hatásának vizsgálata, magyarán metakommunikáció-kutatás a karakter fókuszában. A komplex karakterkutatás pedig a részleges kutatások összesített implementációja.

A karakterkutatás során két fő irány és három stádium létezik: karakter- és/vagy karakterrepresentáció-elemzés, illetve az elsődleges, másodlagos, s a harmadlagos kutatás. Az elsődleges kutatás a karakterépítés módozataival, a másodlagos a karakterlektusok és a karaktertípusok/kommunikátorcsoportok megállapításával foglalkozik, míg a harmadlagos – a metakutatás lévén – a karakterrepresentáció módját vizsgálja.

Az elemzés történhet valós/virtuális dimenzióban, irányulhat karaktertípusokra, de konkrét karakterre is, akár csak aktorra és inaktorra is, s maga a kommunikátor karakterjelölőire vagy a kommunikátorcsoport jellegzetességeire is; továbbá értelmezhető hagyományos vagy kiterjesztett kommunikáció-keretek közt.

Jelen dolgozatban komplex karakterkutatót végzünk a Pokemon-karakterlektus fókuszában.

### 3 Analízis

A Pokemon-franchise eredeti formájában egy videójáték, amely valójában konzolra, GameBoy-ra készült 1996-ban. A rákövetkező évben manga és anime követte Japánban, majd 1998-ban nemzetközileg is teret hódított mind a videójáték, mind az anime (a manga csak 2000-től).

A Pokemonoknak van nevük, kiáltásuk, típusuk, természetük, támadásuk, élőhelyük, általában fejlődésük is, képességük, tojásuk, alternatív kinézetük, és a későbbi játékokban már genderük is.

Egy Pokemont, vagy mondhatni „a Pokemont” mindenki ismeri Pikacsu személyében. Egy másik Pokemont legyőzve Pikacsu lett a nem hivatalos kabalája a Pokemon-franchise-nak. Ez a Pokemon büszkélkedhet a legnagyobb hírnévvel, szinte a Pokemon-jelenségével egyszintűnek mondható, hiszen sokak Pikacsuról asszociálnak a Pokemonra és vice versa. Pikacsunak külön rövidfilmjei vannak azonkívül, hogy ő a főszereplő Pokemonja, s maga a főszereplő Pokemon a Pokemonok világában is. Érdekes, hogy a videójátékban csupán egy Pokemon a sok közül, még hozzá nem is a legerősebbek közül – az anime karakterizációja kreálta meg a Pikacsu-imázst.

A Pikacsu-hype-ban tapasztalható meg egyik legjellegzetesebb formájában a karakter. A karakter és a karakterizáció erejét demonstrálja a máig világhírű és (gazdasági tekintetben) sikeres franchise.

Mindent azonban nem vettek át a Pokemon-sorozat lokalizátorai, különösen az USA-ban és Koreában. A Pokemonok nevei különböznek egyes nyelveken, s rengeteg kulturális jelenség általánosításra vagy kihagyásra került, akárcsak az erőszakos vagy a szexuális témák.

A fentiek szellemében már megállapíthatjuk, hogy nem egy homogén nyelvvel fogunk találkozni, s hogy (karakterbeli) különbségek várhatóak a videójáték és az anime közt.

Az elemzés szemléletesebbé tétele érdekében érdemes az alábbi magyarázatot a kutatás poszterével (Kaba 2017) együttesen értelmezni.

#### 3.1 Állatnyelv

Természetesen az állatok valós és virtuális nyelvét mindenki ismeri, a tudomány is: tudjuk, hogyan beszélnek a lovak, kutyák, macskák, és így tovább. Tehát a nem nyelvi jeleket nyelviesítjük onomatopoeikus kifejezések formájában.

A Pokemonok dizájnja nagyrészt létező állatokon alapul (l. Kaba 2017: 1. Állatnyelv). A lovak lóhangot, a bikák bikahangot adnak ki, tehát minden állat-Pokemon állatként viselkedik. Ez azonban nem karakteres, hanem természetes. Egy bizonyos perspektívában mondhatjuk, hogy karakteres jelenség az, hogy a kutya ugat, hiszen csak a kutya ugat, viszont ez megszegné azt az irányelvet, miszerint minél több a referencia, annál gyengébb a karakter.

Tehát, ha egy kutya nyávogna, míg a többi kutya ugatna, akkor a nyávogó kutya karaktererős lenne a többihez képest.

### 3.2 Pokemon-nyelv (nem verbális)

Mi a helyzet azonban akkor, amikor egyáltalán nem létező egyedekről, teljesen fiktív lényekről van szó? Az eljárás végtére is ugyanaz: a durvábban kinéző egyedek nehezekebb, mélyebb hangot kapnak; a kicsinek és kedvesnek kinéző Pokemonok pedig cincogó, csipogó, magas frekvenciájú hangot.

Egyfajta pszeudo-állatvilágot teremt ezáltal a Pokemon. Ugyanazt az identifikációprocesszust láthatjuk, mint a valóságban az állatok esetében: először is hallunk egy hangot és látunk egy élőlényt, aztán felismerjük, értelmezzük az adatokat, majd megállapítjuk, melyik állatról van szó. A Pokemonok nem verbális kommunikációja során is ez történik, csupán egy közbülső implicit lépésben tér el a természetestől: a Pokemonok kommunikációjának megállapításakor felhasználjuk a tapasztalatunkat, amely a természeti jelenségeken, a valós állatokkal kapcsolatos – mindamelllett, hogy az audiovizuális reprezentáció hat az érzéseinkre, ítéletünkre.

A karaktermodell megállapítható a nem verbális kommunikációnál is, hiszen a szörnynek kinéző Pokemonok durva, nyers hangon, míg a kisebb, gyámoltalanabbnak kinéző Pokemonok nyöszörögve nyilatkoznak meg. Ez a fajta karaktermodell azonban jóval gyengébb, mint amelyenket a következőkben láthatunk; főleg a videojátékot és a mangát jellemzi, bár vannak kivételek.

### 3.3 Embernyelv

Néhány Pokemon, rendkívüli módon, képes beszélni az emberek nyelvét. A leghíresebb beszélő Pokemon Miau. A videojátékban és a mangában a Pokemonok nem tudnak beszélni még Pokemonul sem, nemhogy embernnyelven. Az anime szerint viszont Miau megtanult az emberek nyelvén azért, hogy meghódítsa a (viszonyatlan) szerelmét. Jellegzetes a nyelvi megnyilatkozása, hiszen valamennyi nyelven rendszeresen onomatopoeitikus kifejezést használ, magyarul ’nyávog’.

Az animében voltak olyan Pokemonok, amelyek Miau képességét kihasználva beszéltek az emberekhez, például Tentacruel. Az embernnyelv használatának egy másik különleges módja a telepátia. Lapras képes erre.

A Pokemon-társadalom megveti, mindenesetre fél a domesztikált Pokemonoktól – kiváltképp a beszélőektől.

### 3.4 Pokemon-nyelv (verbális)

A világon a Pokemonokat tekintve nem az állatnyelv, nem a MIDI-hangok, s nem az emberi nyelv jut eszünkbe, hanem az például, hogy „Pikacsu”, vagyis: a nevük. Már az anime első részében elhangzik a főszereplőtől Pikacsunak címezve, hogy „Csak a nevedet tudod ismételni?” – amire a válasz, hogy „*Pi-ka*”.

A tipikus Pokemon-beszéd a névmondás. A legtöbb Pokemon tehát a saját nevének az ismétlésével alakít ki kommunikatív nyelvi interakciót. Új értelmet nyer a beszélő név kifejezés, ugyanis minden egyes Pokemonnak egyedi neve van, amelyet többé-kevésbé realizálni is tud vokális úton.

Mivel a Pokemon-nyelvek a változói meglehetősen korlátozottak, tulajdonképpen a saját nevük, ezért főleg stilisztikai, illetőleg retorikai, illetve szupraszegmentális

módozatokkal tudnak elérni bizonyos funkciókat a kommunikáció során, például név- vagy névtöredék-ismétlés, nyomaték, intonáció és hangsúly.

Nem újdonság, hogy látszólag egy „egyszerű primitív nyelv” képes a kommunikáció funkcióinak kielégítésére, vegyük például a számítógépek bináris kódnyelvét vagy a programnyelveket, a különböző civilizálatlan törzsek nyelvét, s a fekete-fehér vizuális nyelvet, amely ugyancsak képes a kommunikáció szinte bármely funkcióját betölteni.

A Pokemonok tehát a nevük kimondásával ugyanabban a szituációban akár többféle kommunikációs funkciót is elérhetnek. A Pokemon-rajongók részéről vannak ugyan próbálkozások ennek megfejtésére, de annyira korlátozott a nyelvi egységek tárháza, hogy az alapvető érzelmeken kívül nem lehet elkülöníteni meghatározott konkrét vagy absztrakt értelemmel bíró egységeket – hacsak nem Pokemon az ember.

### 3.5 Kiterjesztett Pokemon-kommunikáció

Az anyag vizsgálata során a korábbi kategóriákon kívüli jelenségeket is felleltünk, amelyeket azonban nem lehet a felállított tipológia (vö. 4. Eredmények) keretén belül értelmezni. A kiterjesztett kommunikáció instrumentalizált implementációja viszont képes rendszerbe helyezni ezen eseteket is.

#### 3.5.1 Beszédzavar

Vannak olyan Pokemonok, amelyek nem rendelkeznek kifejlett kommunikatív kompetenciával, például Pysduck, Snorlax, Slowpoke, Slowbro, Ekans, Koffing – sem a verbális, illetőleg nem verbális kommunikációjuk, sem a beszédértésük nem fejlett. Ez kiderül a médiumfüggő sajátosságoknak köszönhetően, például a feliratból, ráadásul ezen Pokemonok audiovizuális reprezentációja is utal a karakterükre (l. Kaba 2017: Beszédzavar); mi több, nemcsak az emberek és a PokeDex, de még a Pokemon-társaik is butának tartják őket.

#### 3.5.2 Éneklés

Bár az éneklést nem tartják (nyelvi) kommunikációnak (Nemesi 2013: 179), amellet érvelünk, hogy az. Az éneklő Pokemonok általában altatókat énekelnek, például Kangashkan, Jigglypuff, vagy egyenest a Holdnak énekelnek, például Pikacsu.

Jigglypuff éneklése azonban eltér a többiekétől, ugyanis az ő célja nem az altatás, ez csak a kívánatlan velejárója, hanem a produkció, maga az éneklés és a hang szépségében való gyönyörködtetés.

Láthatjuk, hogy nem csupán nyelvi kommunikációról van szó a névvel történő éneklés jelenségében, hanem rendkívül erős kommunikatív funkciót is betölt: elaltatja az embereket és a Pokemonokat egyaránt.

#### 3.5.3 Az emberek Poke-nyelve

Nemcsak a Pokemonok képesek embernyelven beszélni, hanem az emberek is tudnak olykor Pokemonul. A híres Pokemonfilm-rendező például nagy rajongója Pikacsunak,



Pikacsuról készít filmeket, s a szavajárásává vált a *csú* ちゅう karakterjelölő, amelyet a mondatzáró kifejezések helyett, például a kopula helyett használ.

Az anime egyik részében Hypno hipnotizálta a gyerekeket, akik ezért azt hitték, hogy Pokemonok, így Pokemonként viselkedtek, úgy kommunikáltak. Ehhez az esethez hasonlít Tomi története, akit a dzsungelben felejtettek a szülei, így a Kangashkanok nevelték fel, így ő maga is Kangashkannak gondolta magát, úgy is kommunikált.

Mágika, az utánzómaster pedig Pokemonokat utánoz vokális és vizuális úton is. Ez a téma azonban egy következő ponthoz vezet (l. alább).

### 3.5.4 Karakterreprezentáció

Mágika utánzása nem szimplán utánzás, hanem karakterreprezentáció – ezen belül karakterimitáció, lévén egy másik karaktert játszik el ideiglenesen, csupán a felismerés kedvéért, identitásazonosulás nélkül.

Mágika Pokemonja, Ditto azonban pusztán imitálásnál többet tesz: karakterperformanciát végez, mivel egy másik Pokemon identitását, karakterét veszi föl, mindeközben a saját identitását impliciten megartva. Dittóhoz hasonlóan Gastly is képes a karakterperformanciára, de nemcsak Pokemonokra nézve, hanem emberre, állatra, szellemre és tárgyra is.

A karakterperformancia egy részleges változatát alakítja Squirtle, Tentacool, illetve Gastly, ugyanis eltulajdonítják a másik Pokemon tulajdonságát, képességét, karakterét, miközben a saját karakterük expliciten jelen van.

### 3.5.5 Médiumfüggő aspektusok

A legtöbb nyelvi példa az animéből származik, a videojátékokra és a mangára azonban teljesen más kommunikációstratégia jellemző. A videojátékban a Pokemonok kommunikációja emotikonok és függő beszéd útján, onomatopoeikus kifejezések révén kerül kifejezésre. A mangára szintén ezek a tényezők jellemzők, bár az animemangán érződik az anime hatása, így jelen van a nyelvi kommunikáció.

Az anime bizonyos részeiben feliratozzák a Pokemon-beszédet, így többet tudhatunk meg az adott Pokemon karakteréről.

### 3.5.6 Kultúrafüggő aspektusok

Abból kifolyólag, hogy a franchise lokalizációja során a legtöbb Pokemon-név átalakításra került, a Pokemonok kommunikációja is eltérő. Sok Pokemon az egyik nyelven kimondja a teljes nevét, a másikon pedig csak egy részletét; vagy az egyik nyelven verbális kommunikációt használ, a másik nyelven pedig nem verbálisat.

## 4 Eredmények

A korpusz elemzése során kiviláglott, hogy a Pokemonok nyelve karakterlektus, azaz csak a Pokemonokra jellemző virtuális nyelvváltozat. Ugyanakkor a névmondás mint

kommunikáció nem csak a Pokemonokra jellemző (vö. Game of Thrones: Hodor; egyes madárfajok: kakukk), ezért valamennyi névvel kommunikáló kommunikátor nyelvváltozata úgymond a névalapú karakternyelv szuperkategóriájába tartozik.

Az elemzett anyagok alapján a Pokemon-karakterlektus rendszerezésére a következő tipológiát állapíthatjuk meg:

- I. valós relációjú virtuális kommunikáció
  - a. állatnyelv
  - b. embernyelv
- II. fiktív relációjú virtuális kommunikáció
  - a. verbális Pokemon-nyelv
  - b. nem verbális Pokemon-nyelv

A Pokemonok tehát nyelvi és nem nyelvi elemeket, illetve természetes és mesterséges tényezőket egyaránt használnak. A karakterteória számára rendkívül hasznos anyaggal szolgált a Pokemon-kommunikáció vizsgálata, kiváltképp a nyelvi kommunikáció, hiszen bebizonyosodott, hogy prototipikus karakterjelölők, ugyanis valamennyi (verbális) Pokemon-nyelven történő megnyilatkozás tiszta karakterjelölőnek tekinthető, a szó legeredetibb értelmében.

A Pokemon-kommunikáció vizsgálata felfedte a karakterképzés néhány aspektusát is, így a karakterépítést és a karakterreprezentációt is. A karaktert a kiterjesztett kommunikáció elméleti keretrendszerében értelmeztük, ennek révén fény derült a Pokemonok meta-kommunikációjára, a kultúraközi és médiumközi aspektusok hatására is.

A megállapított tendencia szerint az animében az angol és így a magyar lokalizáció is a verbális Poke-nyelvet, esetleg az állatnyelvet implementálja, míg a koreai a nem verbálisat, az eredeti japán pedig arányosan ötvözi valamennyi kommunikációmódot. Nem mondhatjuk tehát, hogy a Pokemonok kizárólag így, vagy kizárólag úgy beszélnek, ugyanis kultúránként és médiumonként változik az adott lehetőségek, keretek függvényében – a kommunikáció esszenciája azonban változatlan: a legtöbb Pokemon a nevét használja kommunikációra. Arra is volt példa, hogy beceneve van a Pokemonnak, de ettől függetlenül a Pokemon továbbra is az anyanyelvén, Poke-nyelven beszél.

A Pokemonok kölcsönösen megértik egymást annak ellenére, hogy látszólag más nyelvet beszélnek, más fajhoz, mondhatni kultúrához tartoznak. Néhányuk képes egymás karakterének performálására, esetleg énekelni, viszont van olyan is, amelyik még csak alapvető kommunikációra sem képes.

A kiterjesztett kommunikációt alátámasztja, hogy a Pokemon-kommunikációnak nincs mindig iránya; nem létfontosságú az, hogy a feladótól címzetthez érkezzon az üzenet; csupán az adott cselekvés maga a célja a kommunikációs aktusnak, például előugrás a Poke-labdából. Ilyenkor a Pokemon csak magának beszél – a lényét fejezi ki a kommunikációja, tehát karakterjelölői funkciót tölt be.

Összegezve tehát: a Pokemonok kommunikációja rendkívül komplex rendszer, amelyet lényegében nem tárhatunk fel hagyományos módszerekkel, lévén a Pokemon-világ valamennyi aspektusa mikromenedzselve lett a készítő, lokalizátorok, játékfejlesztők részéről, aminek a vizsgálata a metaszintek elemzését kívánja meg – a hagyományos módszerek pedig csak a konvencionális kommunikációt képesek értelmezni.

## 5 Konklúzió

Jelen írás a virtuális kommunikációt vizsgálta, eddig példa nélküli paradigma szerint: egy komplex rendszeren alapuló instrumentalizált metakutatás keretében. A tanulmány megmutatta, hogy a populáris kultúra is lehet releváns a tudomány számára, s hogy a kommunikáció lehet bővebb, azaz metább, mint a hagyományos felfogás szerint, mi több: marketingcélokat is szolgálhat.

A kutatás jellegéből fakadóan, en bloc újszerűsége okán az új témák és az új jelenségek bemutatásához új módszertan és elméleti keretrendszer, illetőleg eszköztár volt szükséges.

Nem szorul bővebb magyarázatra az előbbi állítás, ha vetünk egy pillantást arra, hogy milyen eredményekre jutottunk volna a hagyományos módszerek által: kiderült (?) volna, hogy milyen szófajok, milyen beszédaktusok, milyen nem verbális és metakommunikáció-stratégiák jellemzők a Pokemonokra mint „homogén” (kommunikátor)csoportha. Jelen kutatás fényében azonban beláthatjuk, hogy ezen módszertani kereteken bőven kívül esik a Pokemon-kommunikáció esszenciája.

A dolgozatban használt paradigma tehát bebizonyította, hogy a „mai kommunikáció” értelmezéséhez a legalkalmasabb, a legadekvátabb, optimális megközelítésmód, legyen szó hagyományos, digitális vagy virtuális kommunikációról.

Nem túlzás, ha azt állítjuk, hogy paradigmaváltás van a küszöbön a nyelv- és kommunikációtudományok terén, akárcsak az oktatást illetően, hiszen a kiterjesztett kommunikáció elmélete és a karakterteória alapján formálja újra az eddigi axiómákat, miközben – az eddig megközelítésektől eltérő módon – reflektál a ma mindennemű kommunikációjára, és értelmezi is azt.

## Irodalom

- Jakobson, R. 1960. Closing Statement: Linguistics and Poetics. In: Sebeok, T. A. (ed.) *Style in Language*. Cambridge, Massachusetts: MIT Press. 351–377.
- Kaba, A. 2016. *A nemek pragmatikája: Nyelvi sztereotípiák a japán nyelvben*. MA-szakedolgozat. Budapest: Károli Gáspár Református Egyetem.

- Kaba, A. 2017. *Pokemon-go ポケモン語: avagy hogyan kommunikálnak a Pokemonok?* Poszter: XI. Alkalmazott Nyelvészeti Doktoranduszkonferencia, MTA Nyelvtudományi Intézet, Budapest, 2017. 02. 03.  
Elérhető: [http://www.nytud.hu/alknyelvdok17/prez/Kaba\\_poszter.pdf](http://www.nytud.hu/alknyelvdok17/prez/Kaba_poszter.pdf). Letöltve: 2017. március 5.
- Nemesi, A. L. 2013. A pragmatika mint kommunikációtudomány. *Magyar Nyelv*, 109(2): 177–184.
- Ong, W. J. 1982. *Orality and Literacy. The Technologizing of the Word*. Abingdon: Routledge.
- Székely, N. 2017. A Pokémon GO-jelenség és a kiterjesztett valóság. *Filológia*, 8(1–2): 66–73.
- Veszelszki, Á. 2010. Digilektus a lektusok rendszerében. In: Illés-Molnár, M., Kaló Zs., Klein L., Parapatics A. (szerk.) *Félúton 5. Az ELTE BTK Nyelvtudományi Doktori Iskolájának konferenciája*. Budapest: ELTE Nyelvtudományi Doktori Iskola. 199–215.
- Veszelszki, Á. 2013. A digilektus hatása a dialóguslevelekre. *Magyar Nyelv*, 109(4): 435–447.
- Veszelszki, Á. 2014. Marketolingvisztika és a marketing szaknyelve. In: Veszelszki, Á., Lengyel, K. (szerk.) *Tudomány, technolektus, terminológia. A tudományok, szakmák nyelve*. Budapest: Éghajlat Könyvkiadó. 321–328.

## Források

- Game Freak 2009/2010. Pokémon HeartGold and SoulSilver. The Pokémon Company. (jp/eng/ko)
- Game Freak 2004. Pokémon LeafGreen. The Pokémon Company. (jp/eng)
- Game Freak 1998/1999/2000. Pokémon Yellow. Nintendo. (jp/eng)
- Deák Kristóf 2016. Mindenki. Meteor-Film.
- PewDiePie 2012. BARRELS!. YouTube. Elérhető: [https://www.youtube.com/watch?v=PJ\\_FYyb0S94](https://www.youtube.com/watch?v=PJ_FYyb0S94). Letöltve: 2017. március 5.
- Puzsér R. 2016/2017. Sznobjektív. HírTV.
- 日下 秀憲 1997. Pokémon Adventures I. (1–6 epizód). 小学館/Viz Media/대원씨아이.
- 川崎 逸朗等 2013. ポケットモンスター THE ORIGIN I–IV. TV Tokyo.
- 田尻智 (原作) 石原恒和 (監修) 2001. 『ポケットモンスター』 「金・銀編」 20 巻. 小学館
- 湯山 邦彦 (総監督) 1997–2002. Pokémon (angol/magyar: 1. évad 82 rész, 3–5. évad 12 rész, 6. évad 2 rész, 18. évad 78. rész; ezek közül 30 részt japánul és 20 részt koreaiul is elemeztünk). TV Tokyo/4Kids Entertainment/SBS et. al./RTL-Klub.

## Az igekötők gépi annotálásának problémái

Kalivoda Ágnes

PPKE BTK Nyelvtudományi Doktori Iskola

k.agnes92@gmail.com

**Kivonat:** Az automatikus szófaji egyértelműsítésben nehézséget okoznak a homográf szavak, vagyis amelyek íráskéjükben egyeznek, de a jelentésük és gyakran a szófajuk is eltérő. Kiugróan magas a tévesztések aránya az igekötők esetében: a *meg* például kötőszóként és igekötőként is gyakori, és gyakran félrelemzett. Nemcsak más magyar szófajokkal lehetnek átfedésben az igekötők, hanem idegen nyelvi szóalakokkal (pl. *be* mint angol létige és mint magyar igekötő), valamint rövidítésekkel, mozaikszókkal is (pl. *LE* mint lóerő). Gyakori és nehezen kezelhető probléma a félregépelés és az ékezetek hiánya (pl. *fél* → *fel*). Az annotáció javítása fontos feladat, mivel a jobb annotáció bármilyen magyar nyelvű, korpuszt használó kutatáshoz jobb alapanyagot biztosít. A kutatásom olyan megoldást ad a homográfia problémájára, amely – a munka jelen szakaszában – a rossz annotációk több mint felét képes kiszűrni. A javítási módszer szabályalapú (reguláris kifejezéseket használ), és az igekötőként elemzett szavak kontextusára támaszkodik.

### 1 Bevezetés

Az automatikus szófaji egyértelműsítésben a homográf szavak jelentik az egyik legnagyobb problémát. Különösen gyakran látni hibás annotációt az igekötők esetében: például a *ki* használata kérdő vagy vonatkozó névmásként is gyakori, az *át* névutó is lehet. Jelen tanulmány egy szabályalapú javítási módszert mutat be, melynek során automatikusan, az igekötőként annotált szó kontextusa alapján dől el, hogy helyes-e a megadott szófaji címke.

Jelen tanulmány célja az, hogy választ adjon a következő kérdésekre:

- (1) Hogyan lehet automatikusan javítani az igekötők gépi annotációját?
- (2) Mennyire lehet hatékony a többértelműségek feloldása?

A tanulmány felépítése a következő: A 2. fejezet ismerteti a felhasznált korpuszokat és az egyes munkafázisokat. A 3. fejezet a homográfia problémakörét járja körül. Ezt követi a szabályalapú megoldás bemutatása (4. fejezet), és a javasolt megoldás kiértékelése a tesztkorpuszon a pontosság, fedés és f-mérték megállapításával (5. fejezet). Végül rövid összegzés után a további kérdések és teendők ismertetése következik.

## 2 Módszertan

### 2.1 Korpuszok

A kutatásban két korpuszt használtam fel. A 1,2 milliárd token nagyságú Pázmány-korpusz (Endrédy 2016) alapján térképeztem fel az igekötőket érintő annotációs hibákat, és állítottam össze a kontextusra illeszkedő reguláris kifejezéseket, amelyek segítségével elvégezhető a javítás. Tesztkorpuszként a Magyar nemzeti szövegtár (MNSz.) 2.0.4 verziója (Oravecz–Váradi–Sass 2014) szolgált, amely írásjelekkel együtt 1,348 milliárd tokent tartalmaz.

A két korpuszban nagyon hasonló elven működik a szófaji egyértelműsítés (az MNSz.-hez l. Oravecz–Dienes 2002, a Pázmány-korpuszhoz l. Orosz–Novák 2013), így a javító script mindkettőn alkalmazható, és közel azonos teljesítményt nyújt. Két nagyobb különbségre hívnám fel a figyelmet a szófaji címkék kapcsán:

(1) Az MNSz.-ben 79 egytagú igekötő szerepel, a Pázmány-korpuszban 76. Az *alább, benn, közé* szavak ez utóbbiban sosem jelennek meg igekötőként, ha elváltak az igétől.

(2) Az ismeretlen szavak kezelése az MNSz.-ben hatékonyabb. Itt SKIP címke járul az idegen nyelvi szavak jelentős részéhez, míg a Pázmány-korpuszban minden szó kap valamilyen magyar szófaji címkét. Például az angol *must be* 'kell hogy legyen' mind a 195 előfordulásában tárgyeesetű főnév + igekötő elemzést kap a Pázmány-korpuszban, az MNSz. 750 példájában 65 alkalommal alanyesetű főnév + igekötő, 695-ször SKIP + SKIP elemzés társul hozzá.

### 2.2 Munkafázisok

A javítás megkezdése előtt fel kellett mérni, mely igekötők milyen más szófaji címkét kaphatnak. A Pázmány-korpuszból lekértem minden olyan mondatot, amely igekötőként annotált szavakat tartalmaz, ezután olyanokat, amelyek tartalmazzák a vizsgált szóalakokat, viszont valamilyen más szófaji címkével. Szavanként átlagosan két különféle elemzéssel találkozunk, de néhány esetben ennél jóval hosszabb a lehetséges elemzések sora (példaként ld. 1. táblázat).

| Annotáció  | Annotáció jelentése              | Gyakoriság |
|------------|----------------------------------|------------|
| NU         | névtű                            | 89 212     |
| IK         | igekötő                          | 13 014     |
| HA         | határozószó                      | 128        |
| FN+POS+NOM | alanyesetű birtokos főnév (mié?) | 57         |
| FN+FAC     | factivus ragos főnév (mivé?)     | 1          |

1. táblázat. A *mellé* szó lehetséges elemzései és ezek gyakorisága a Pázmány-korpuszban

A kapott eredmények alapján megállapítottam a hibatípusokat és az ezek szűrésére alkalmas, kontextusra illeszkedő szabályokat, amelyeket reguláris kifejezésekkel adtam meg. Ezeknek a megalkotása kísérletező munka eredménye volt. A vizsgált mondatokat először csoportokba rendeztem aszerint, hogy mely igekötőket (pontosabban: igekötőnek annotált szavakat) tartalmazzák. Ezekben belül további

csoportokat hoztam létre az igekötőcímkés szó maximum 5-5 szavas környezetének a szófaji címkéiből. Így könnyebben láthatóvá váltak a hibás mintázatok.

Miután a szabályok elkészültek, az MNSz.-ből kigyűjtött 5000 mondatos korpuszon teszteltem a hatékonyságukat. A tesztkorpusz összeállításának szempont-jairól bővebben lesz szó az 5.1. alfejezetben.

### 3 A homográfia problémája

Az igekötők homográfiáját vizsgálva négy hibatípus különíthető el, ezek közül a legnagyobb gondot az elírásokból, ékezet nélkül gépelt szavakból adódó homográfia okozza. Az alábbi példában nem világos, hogy elírásról van-e szó (tehát a *meg + ad* igéről felesleges szóközzel) vagy a *meg* 'pedig' jelentéséről:

[1] Lemondani nem engedi mert amit *meg ad* számot azon nem lehet!!

A legtöbb (összesen 66) igekötőt érintő probléma az, hogy más szófajú magyar szóval lehet azonos az írásképe. Az alábbi példák esetén van egy preferált értelmezésünk, de automatikus módszerrel eldönthetetlen, hogy a vizsgált szónak melyik jelentése szerepel a mondatban:

[2] Ha a karbantartó sokáig vacakol, akkor az a konklúzió, hogy nem *ért hozzá*.

→ *Ért valamihez* vagy *hozzá + ért (valamihez)?*

[3] 4 telót rendeltem online tudom mit ír ki. → *Ki + ír* vagy *(vala)ki ír?*

Bár magyar nyelvű korpuszokról van szó, viszonylag sok idegen nyelvű mondatot is találunk a szövegekben. Ekkor nem ritka eset, hogy például a francia *le*, a spanyol és arab *el* névelő vagy a bolgár *tova* mutató névmás igekötői címkét kap. Néhány példa, amelyben az idegen szó igekötőként elemzett:

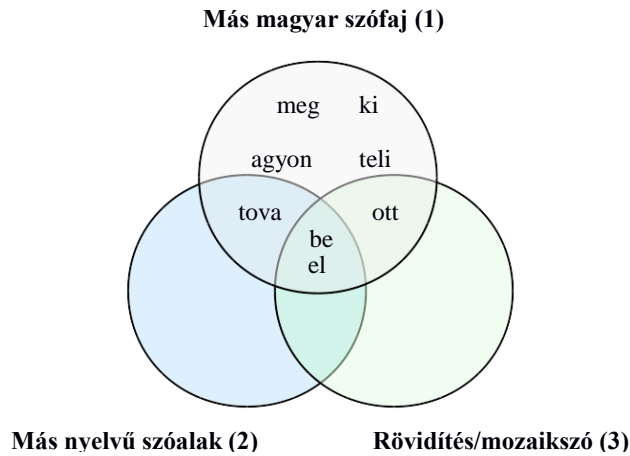
[4] [...] ugyanis a szám kerékpáros szakaszában összeütközött a kínai Ma Ming-hszüü, a dél-koreai Kim Hi Szun és a szingapúri Clara Vong *Van Ki*.

[5] [...] helikopter *vitte* a Sharm *el* Sheiki kórházból a per helyszínére, ahol kerekesszékesen tolták be a tárgyalóterembe.

Rövidítések, mozaikszók is kapnak tévesen igekötői címkét, például:

[6] *LE* (lóerő), *OTT* (oxytocin terheléses teszt)

A helyzetet tovább bonyolítja az, hogy egy igekötő akár több hibatípusban is érintett lehet, ilyenek láthatók az 1. ábra halmazainak metszetében:



1. ábra. Homográfiából adódó hibák az igeekötők esetében

Homográfia szempontjából a *be* bizonyult a leginkább problémásnak, mivel megfelel neki egy rövidítés (*BE = Bowling Egyesület*), idegen nyelvű szóalak (az angol *be* 'lenni'), a magyar nyelven belül pedig kétféle határozószó (pl. *Se ki, se be.; Be rít teremtés!*) Bár elvileg több homográf párja lehet, az annotációja elég pontos (7% körül hibaarányal) az MNSz.-ben és a Pázmány-korpuszban is.

#### 4 A javítás menete

A hibajavítás Python 3.4-ben írt scripttel történik. Ennek bemenete egy szövegfájl, amelyben minden sor egy mondat, a mondat szavainak a lemmatizált (szótővesített) alakjai és a morfológiai-szófaji elemzése is látható (az MNSz. esetében ez az *msd*, a Pázmány-korpuszban a *humor* mező), például:

```
tejet/tej/FN.ACC meg/meg/IK ilyeneket/ilyen/MN_NM.PL.ACC
```

A script megkeresi az *IK*, azaz igeekötő címkével ellátott szót, megállapítja a lehetséges hibatípusokat (a *meg* esetében a hibatípus 1, tehát csak más magyar nyelvi szófajjal keveredik, a *be* igeekötőnél 1, 2 és 3 is lehet, mert minden hibatípussal előfordulhat). Minden hibatípushoz tartozik egy függvény, a függvényen belül pedig számos további szabály (reguláris kifejezés), amelyeket egyesével ki kell próbálni az adott mondatra. Ha az egyik talált, a mondat feldolgozása véget ér.

A *meg*-hez Makrai (2007) már kidolgozott 7 szabályt, amelyeket kisebb módosításokkal felhasználtam. A többi igeekötőhöz írt reguláris kifejezéssel együtt a script összesen 98 szabályt használ. A 2. táblázat a *meg*-re vonatkozó egyik szabályt mutatja be részletesen.



| Pszudokóddal              | Reguláris kifejezéssel  |
|---------------------------|---|
| névszó + esetrag;         | $[\wedge ]+V(FN MN SZN MN\_NM NM\_MN)[+][\wedge ]*([\wedge ]\{3\})$ |
| 0–3 szó, amely nem .?!)   | $([\wedge ]\{0,3\})$  |
| <i>meg</i> (mint igekötő) | $meg\backslash meg\backslash IK$                                    |
| 0–3 szó, amely nem .?!)   | $([\wedge ]\{0,3\})$  |
| névszó + esetrag;         | $[\wedge ]+V(FN MN SZN MN\_NM NM\_MN)[+][\wedge ]*2$                |

**2. táblázat.** A *meg* egyik tipikus kötőszói használatára illeszkedő szabály: két azonos esetragot viselő névszó között a *meg* jellemzően kötőszó (a pszudokódban használt indexelés jelzi az esetragok egyezését)

A script kimenete egy olyan szövegfájl, amely az egyes mondatok előtt további információkat is tartalmaz, ezek a következők:

1. hibás volt-e az eredeti annotáció (Y/N)
2. melyik címkéről mire kell váltani  
(ha nem volt hibás, akkor marad IK/IK, ha hibás, akkor IK/HA, IK/NU stb.)
3. melyik hibatípus áll fel  
(ha nincs hiba: 0, ha van, akkor az 1. ábra halmazainak megfelelő számok)
4. melyik szabály illeszkedett a mondatra (ez a megfelelő szabály számával jelölt)

Ezekre azért van szükség, hogy fejleszteni lehessen a szabályrendszert, könnyen átlátható, melyik szabály milyen gyakran érvényesül és mennyire hatékony. Három főbb szabálytípust használók:

Leggyakrabban **általános, mintaalapú szabályokra** van szükség, ilyen a 2. táblázatban bemutatott reguláris kifejezés, vagy például az *át* javítására használt szabály, miszerint: ha superessivus ragos névszót követ az *át* (pl. *ezt hallgattuk órákon át*), akkor nem igekötő. Az általános szabályok hátránya az, hogy ritkán ugyan, de jól annotált szavakat is kiszűrnék. Például az előbbi *át*-tal kapcsolatos szabályra ez a mondatrész is illeszkedik: *esett a műtéten át*, ahol valóban arról van szó, hogy *át* + *esett a műtéten*, így nem kellene javítani.

Az idegen nyelvi szóalakkal való keveredés esetén a **speciális, lexikonalapú szabályok** bizonyultak a leghatékonyabbnak. Ez lényegében az idegen nyelvi kontextus, általában tulajdonnevek listázását jelenti, amelyet egy külön fájlból beolvas a script, és kivételként kezeli őket (pl. *El Greco, Sharm el Sheiki, Entre el amor y el odio*). Ennek hátránya az, hogy folyamatosan kerülhetnek be új, idegen nyelvi elemek a korpuszba, így a kivételek listája nem fedheti le az összes esetet.

A harmadik szabálytípus a **lehetséges ige-igekötő kombinációkat** veszi figyelembe (27 091 kombinációt ismer, a listát l. Kalivoda 2016). Ez olyan igekötők esetén hatékony, amelyek nem produktívak, és csak néhány igehez járulhatnak, pl. *észre, észbe, egyet*. A szabály lényege az, hogy ha az igekötőként annotált szó a tagmondat valamelyik igéjével létező kombinációt alkot, akkor marad az igekötői címkéje, ellenkező esetben új szófaji címkét kap. Például:

[7] szépség mellett *észre* is szükség van → *észre* + *van* nem létező kombináció

[8] ezt a hibát korábban *észre* sem *vettem* → *észre* + *vesz* létező kombináció

Ennek hátránya az, hogy ha félrelemzett az ige, akkor a script tévesen javítja a mondatot, például:

[9] *észre* is *vétette* → *vétet* helyett *vét* a lemma, *észre* + *vét* nem létezik

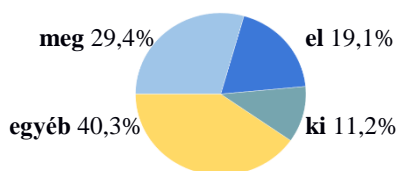
Mint látjuk, mindegyik szabálytípusnak megvannak a korlátai, és maga az igekötő, valamint annak hibatípusa határozza meg, mikor melyiket célszerű alkalmazni.

A további technikai részletek és a forráskód elérhető az alábbi linken: [http://github.com/kagnes/prt\\_rules](http://github.com/kagnes/prt_rules) (letöltve: 2017. október 16.).

## 5 Kiértékelés

### 5.1 A tesztkorpusz

A legfontosabb szempont a tesztkorpusz összeállításánál az volt, hogy az igekötők mennyiségét tekintve az MNSz. kicsinyített mása legyen. Ehhez először megmértem, hogy a teljes MNSz. igekötő-állományának hány százalékát alkotják az egyes igekötők. Az eredményt a 2. ábra szemlélteti.



**2. ábra.** Az MNSz. igekötő-állománya: leggyakoribb igekötők a *meg*, *ki*, *el*, a további 76 igekötő összesen 40,3%-ot tesz ki

Az MNSz.-ből ezután lekértem véletlenszerűen 5000 mondatot, amely igekötőnek annotált szót tartalmaz – nem az ige vagy igenév közvetlen környezetében, ezzel növelve a hibás annotáció esélyét. Ügyeltem arra, hogy a teljes MNSz.-ben megfigyelt arányokat megőrizzem (így pl. a mondatok 29,4%-a, azaz 1470 mondat tartalmazta a *meg* igekötőt). A kapott adatokat átnéztem, és az 5000-ből 1033, tehát a tesztkorpusz egyötöde bizonyult hibásnak. Mivel a hibás mondatok esetén azt is meg kellett határozni, hogy mi lenne a helyes annotáció, a feladat 8-10 órát vett igénybe. A hibák arányát a 3. táblázat mutatja be részletesen.

| Igekötő | Darab | Helyes | Hibás | Hibásak %-a |
|---------|-------|--------|-------|-------------|
| meg     | 1470  | 960    | 510   | 34,7        |
| el      | 956   | 901    | 55    | 5,8         |
| ki      | 563   | 463    | 100   | 17,8        |
| be      | 368   | 341    | 27    | 7,3         |
| fel     | 355   | 332    | 23    | 6,5         |
| vissza  | 123   | 96     | 27    | 21,9        |
| át      | 118   | 21     | 97    | 82,2        |
| hozzá   | 77    | 37     | 40    | 51,9        |
| rá      | 70    | 22     | 48    | 68,6        |

**3. táblázat.** Néhány gyakori igekötő eloszlása a tesztkorpuszban (az igekötő mellett látható a tartalmazó mondatok mennyisége, a helyes és hibás annotáció mennyisége, végül a hibás annotáció aránya a helyeshez képest)

Három olyan igekötő van, amelyik viszonylag gyakori, és az esetek több mint felében hibás szófaji címkét kap, ezek az *át*, *rá* és *hózzá*. Az *át* 82,2%-ban igekötőként annotált akkor is, ha valójában névutó, így kimagasló hibaarányal bír. A *vissza* esete azért problémás, mert egyik korpuszban sincs soha határozószóként annotálva, pedig 21,9%-ban ez volna a megfelelő címke.

## 5.2 Eredmények

A szabályok hatékonyságát a számítógépes nyelvészetben hagyományosnak számító módszerrel, a pontosság, fedés és f-mérték segítségével értékeltem ki. A pontosság (1) azt fejezi ki, hogy a script által hibásnak jelölt annotációk közül mennyi valóban hibás. A fedésre (2) kapott százalékos érték azt jelzi, hogy az összes hibás annotáció közül hányat talált meg a script. Az f-mérték (3) a pontosság és fedés harmonikus közepe.

(1) pontosság = (valóban hibásak × 100) / hibásnak jelöltek

(2) fedés = (megtalált hibásak × 100) / összes hibás

(3) f-mérték =  $2 \times ((\text{pontosság} \times \text{fedés}) / (\text{pontosság} + \text{fedés}))$

A 4. táblázat tartalmazza a script eredményeit az 5000 mondatos tesztkorpuszon.

| Pontosság | Fedés  | F-mérték |
|-----------|--------|----------|
| 88,2 %    | 57,5 % | 69,6 %   |

4. táblázat. A javító script teljesítménye a tesztkorpuszon

A pontossága viszonylag jó, 88,2%-ban igaz, hogy hibás annotációt ismert fel, nem egy eredetileg jót rontott el. A fedése viszont gyenge, az összes hibás szófaji címkének valamivel több mint a felét sikerült felismernie. Ugyanakkor fontos, hogy a felismerésen túl 98%-ban helyesen is javította az annotációt (a tökéletlenség néhány olyan esetenél mutatkozik meg, ahol rosszul választ a névutó és a határozószó címke között).

A fedés két okból alacsony. Egyrészt, ha maga a kontextus is hibásan elemzett (pl. az *ír* ige melléknévi annotációt kap), akkor a megfelelő szabály nem illeszkedik rá. Másrészt a fel nem ismert mondatok átnézése során kiderült, hogy rengeteg köztük az elírás (pl. *el* helyett *le*), hibás szóközt tartalmazó, vagy egyszerűen ékezetek nélküli mondat (pl. *fél* → *fel*, *még* → *meg*). A script az elírásokat jelenleg semmilyen formában nem kezeli.

## 6 Összegzés, további kérdések

A kutatás eredménye egy olyan script, amellyel az igekötők annotációját utólag lehet javítani. A kutatási kérdésekre adott válaszaim a következők: (1) A kontextus mintázatai alapján hatékonyan javítható az igekötők annotációja, de a hibásan elemzett kontextus és elírások felismerése, javítása – egyelőre – nem megoldott. Tökéletes eredményt nem is várhatunk. (2) A hibás annotációt az esetek több mint felében sikerült egyértelműen kiszűrni. Legnehezebbnek azok a szerkezetek bizonyultak – az elírásokat

leszámítva –, ahol a kérdéses szó (főként a *meg* és *ki*) olyan pozícióban áll, ahol igekötő is várható, és a tagmondat igéjével létező kombinációt alkothat.

További teendők a fedés arányának javítása, az összetett igekötők (pl. *bele-bele*, *ki-be*) annotációjának ellenőrzése, valamint a javítási irány megfordítása: olyan szabályokat is létre kell hozni, amelyek nem igekötőként annotált szavakról döntenek el, hogy igekötők-e.

Jelen kutatásban olyan mondatokkal foglalkoztam, ahol az igekötőnek elemzett szó közelében ige is szerepelt. Viszont az ige jelenléte nem kötelező egy elliptikus szerkezetben, az eldöntendő kérdésre adott rövid válaszban. A problémát az alábbi két példa szemlélteti, az elsőben a *meg* igekötő, a másodikban diskurzuspartikula (jobb híján KOT, azaz kötőszó címkét lehetne hozzárendelni a korpuszokban):

[10] Megírtad a cikket? Én *meg*.

[11] Te megint ülsz le játszani... Én *meg*?

Végül nehézséget jelent az is, hogy az igekötő-állomány az MNSz.-ben és a Pázmány-korpuszban sem konzisztens. Például igekötőként elemzett a *zokon*, de a *cserben* inessivus ragos főnévként szerepel, pedig a mondatbeli viselkedése és a produktivitás szempontjából ez a két szó nem különbözik egymástól. Sok esetben nem dönthető el, hogy az adott szó határozószó vagy igekötő-e, pl. *oda*, *haza*; illetve névmás vagy igekötő, pl. *rá*. Mivel ezeknek nyelvészeti szempontból sem egyértelmű a besorolása, érdemes lehet egy vagylagos szófaji címke használata, az *oda* esetén például [IK|HA]. Ez könnyen kivitelezhető módosítás – és az MNSz.-ben már most is találunk hasonlót.

## Források

MNSz. = *Magyar Nemzeti Szövegtár* v2.0.4. Elérhető: <http://corpus.nytud.hu/mnsz>. Letöltve: 2017. szeptember 26.  
Pázmány-korpusz. Nyilvánosan nem elérhető.

## Irodalom

- Endrédi, I. 2016. *Nyelvtechnológiai algoritmusok korpuszok automatikus építéséhez és pontosabb feldolgozásukhoz*. PhD-értekezés. Elérhető: [https://itk.ppke.hu/uploads/articles/163/file/EI\\_PhD\\_Dissert%C3%A1ci%C3%B3.pdf](https://itk.ppke.hu/uploads/articles/163/file/EI_PhD_Dissert%C3%A1ci%C3%B3.pdf). Letöltve: 2017. szeptember 26.
- Kalivoda, Á. 2016. *A magyar igei komplexumok vizsgálata*. Mesterszakos szakdolgozat. Elérhető: [https://github.com/kagnes/hungarian\\_verbal\\_complex](https://github.com/kagnes/hungarian_verbal_complex). Letöltve: 2017. szeptember 26.
- Makrai, M. 2007. *Többértelműségek magyar mondatok számítógépes elemzésében – a „meg” szó szófajának vizsgálata gyakoriságokkal*. Témalabor-dolgozat. Elérhető: [http://hlt.bme.hu/media/pdf/makrai07\\_temalabor\\_meg.pdf](http://hlt.bme.hu/media/pdf/makrai07_temalabor_meg.pdf). Letöltve: 2017. szeptember 26.
- Oravecz, Cs., Dienes, P. 2002. *Efficient Stochastic Part-of-Speech Tagging for Hungarian*. In: Calzolari, N., Choukri K., Maegaard B., Mariani J., Municio A. M. Tapias, D. Zampolli, A. (szerk.) *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*. Párizs: ELRA. 710–717. Elérhető: [https://www.researchgate.net/profile/Csaba\\_Oravecz/publication/2481141\\_Efficient\\_Stochastic\\_Part-of-Speech\\_Tagging\\_for\\_Hungarian/links/0912f5139e302a44c0000000/Efficient-Stochastic-Part-of-Speech-Tagging-for-Hungarian.pdf](https://www.researchgate.net/profile/Csaba_Oravecz/publication/2481141_Efficient_Stochastic_Part-of-Speech_Tagging_for_Hungarian/links/0912f5139e302a44c0000000/Efficient-Stochastic-Part-of-Speech-Tagging-for-Hungarian.pdf). Letöltve: 2017. szeptember 26.
- Oravecz, Cs., Váradi T., Sass B. 2014. The Hungarian Gigaword Corpus. In: Calzolari, N., Choukri K., Declerck T., Loftsson H., Maegaard B., Mariani J., Moreno A., Odijk J.,

- Piperidis S. (szerk.) *Proceedings of Ninth International Conference on Language Resources and Evaluation (LREC 2014)*. Elérhető: [http://www.lrec-conf.org/proceedings/lrec2014/pdf/681\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/681_Paper.pdf). Letöltve: 2017. szeptember 26.
- Orosz, Gy., Novák, A. 2013. *PurePos 2.0: a hybrid tool for morphological disambiguation*. In: Angelova, G., Bontcheva K., Mitkov R. (szerk.) *Proceedings of Recent Advances in Natural Language Processing (RANLP 2013)*. Sumen: INCOMA Ltd. 539–545. Elérhető: <http://aclweb.org/anthology/R13-1071>. Letöltve: 2017. szeptember 26.

## Az óvodások megnyilatkozásairól az *izé* kapcsán<sup>1</sup>

Kondacs Flóra

SZTE BTK Nyelvtudományi Doktori Iskola

kondacs.flora@gmail.com

**Kivonat:** Kutatásomban az *izé* nyelvi elemmel foglalkozom az óvodások megnyilatkozásában. Megállapítottam, hogy a 3–7 éves korosztály spontán beszédében a szóhelyettesítő *izé* fordul elő nagyobb arányban, akárcsak a felnőtt és a 6 évesnél idősebb korosztálynál (Gyarmathy–Neuberger 2013). A kutatás során megfigyeltem továbbá azt is, hogy már a 3 éves óvodások is javítják az *izét*. A gyermekek jellemzően főnevet, igét, igenevet korrigálnak, akárcsak a felnőttek (Gyarmathy 2012), azonban megállapítható még, hogy számnevet, névmást és teljes frázisokat is javítanak a vizsgált elemmel.

### 1 Bevezetés

Tanulmányom első fejezetében az *izé* elméleti keretét mutatom be részletesebben. Megismertetem a társalgáselemzők, a beszédkutatók álláspontját, illetve vázolom az *izé* helyzetét a diskurzusjelölői szakirodalomban, melyet kiegészítek saját kutatási eredményeimmel. Emellett bemutatom az *izé* vizsgálata során eddig elért eredményeket. A második fejezetben a kutatásom módszertanát ismertetem. A harmadik részben az elért eredményeket prezentálom, míg a negyedik fejezet az eredményekből levont következtetéseket tartalmazza.

A társalgáselemző megközelítés szerint az *izé* olyan lexikális kitöltő elem, mely „a beszédprodukciónak mechanizmus különféle hibáinak javítására használatos, az adott tervezési hiba függvényében” (Fabulya 2007: 324). A lexikális kitöltő elemek segítségével a beszélgető partnerek irányíthatják a társalgást, illetve koordinálhatják a beszélőváltásokat (Clark 1994).

Fabulya Márta (2007) a beszédprodukciónak nehézsége alapján megkülönbözteti a szóhelyettesítő *izét* és a töltelékshói *izét*. A szóhelyettesítő *izé* a lexikális hozzáférés hibái alapján jellemezhető, ekkor ugyanis a mentális lexikonban megtalálható elemhez vezető út során valamilyen probléma lép fel. Két típusát különíti el: a szókereső és a szópótló *izét*. A szókereső *izé* esetén a probléma az, hogy nem sikerül a célszó elérése az adott pillanatban, azonban a beszélő a célszót később elő tudja hívni a mentális lexikonjából pl.

---

<sup>1</sup> A tanulmány az Emberi Erőforrások Minisztériuma ÚNKP-ÚNKP-16-3 kódszámú Új Nemzeti Kiválóság Programjának támogatásával készült. Köszönettel tartozom a Szegedi Tudományegyetem Általános Nyelvészeti Tanszéke Fejlődéses és Neuropragmatikai Kutatócsoportjának a szakmai segítségért.

[1]

F: És mit csinált a pipi?

Gy: a meglátta a **izét** (.) a bácsit.<sup>ii</sup>

A szópótló *izénél* ezzel szemben teljesen hiányzik a keresett szó a beszélő mentális lexikonjából, ezért ezt a beszélő nem is tudja pótolni. A szóhelyettesítő *izé* szintaktikai szerkezetbe illeszkedve látható. Például:

[2]

\*FW1: és mit csinálnak akkor?

\*CH1: békát keresnek .

\*FW1: mhm.

\*CH1: és a kutya pedig a méhek [//] mézet [//] és odament a méhek **izéjéhez** vagy mi ez .

\*FW1: mhm<sup>iii</sup>

A töltelékszói *izé* jellemzője pedig, hogy a beszélő még a mondanivalója megformálásánál tart, így a beszédhez való jogát biztosítja az *izé* által. Gyakran toldalékolt alakban jelenik meg, ami miatt számos szófaj helyett előfordulhat (Fabulya 2007). Fabulya Márta (2007: 339) szerint az *izé* „töredezettségmentesítő” funkcióban használatos a beszédben, amely segíti a hallgatót a beszélő megnyilatkozásának megfelelő interpretálásában.

A beszédkutatói hozzáállás (Horváth 2004: 187) szerint a beszédprodukciónak egy összetett folyamat, hiszen a beszédtervezés és a beszédprodukciónak szinte egyszerre zajlik, ezáltal a két folyamat működése zavarhatja egymást, és így a spontán beszéd folytonosságában megjelennek a megszakítók, a megakadásjelenségek. A megakadásjelenség (Gósy 2003: 257) a spontán beszédben fellépő különféle bizonytalanságok és hibák gyűjtőfogalma. Két nagyobb csoportra bontható (Gósy 2002): a bizonytalanságból adódó megakadásokra és a téves kivitelezésre. Az utóbbi jellemzője, hogy a beszédprodukción minden szintjén jelen van, míg az előbbi a nyelvi tervezés és a fogalmi szint problémáit mutatja meg. A tervezés bizonytalanságának következménye többek között az *izé* is. Az *izé* töltelékszói megítélése kétféle képet mutat. Bár hagyományosan töltelékszóként nevezik, mégsem tipikus töltelékszó. A töltelékszók ugyanis önálló jelentéssel rendelkeznek, és csak a spontán beszédben jelennek meg töltelékszói használatban, vagyis amikor az adott lexémát a beszélők beszédprodukciónak mankóként használják (Pierrehumbert 2001; Bybee 2001). Ekkor elveszítve „konvencionális tartalmukat eredeti funkciójuk megváltozik a spontán közlésekben” (Gósy–Horváth 2009: 26). Az új, töltelékszói funkció megléte mellett azonban „a régi tartalom is használatban marad” (Gósy–Horváth 2009: 26). Az *izé* ezzel szemben nem rendelkezik önálló jelentéssel. Egyes nyelvtörténészek nézőpontja szerint is bizonyos töltelékszavak gyakori használatuk miatt szemantikailag kiüresednek (Hadrovics 1992: 395), azaz önálló jelentéssel nem rendelkeznek a spontán beszédben. A nyelvművelői munkák (Grétsy–Kemény 2005: 334) azonban töltelékszóként hivatkoznak a vizsgált elemre, illetve a Magyar értelmező kéziszótár (2003: 586) a szótári jelentését is megadja az *izének* „olyan pótló szócseka, amely nem jut eszünkbe, ill. amelyet nem akarunk kimondani”, illetve „emlékezet- v. fogalmazásbeli zavar áthidalására használatos henye szó”.

<sup>ii</sup> A példaanyag forrása: saját korpusz (Kondacs 2017). Az átírásban a konverzációelemzési átírási konvenciókat használtam fel (Jefferson 1984; Németh 2014).

<sup>iii</sup> A példaanyag forrása a MONYEEK adatbázis (Mátyus–Orosz 2014). A CHILDES átírási rendszert használták fel.

A megakadások egyfajta időnyerő lehetőségek (Gósy 2003: 275), melyek mind a beszélőnek, mind a hallgatónak hasznosak, hiszen hozzájárulnak a megnyilatkozás előállítási és feldolgozási folyamatának sikeréhez. A beszélők öntudatlanul létrehozhatnak egy olyan eszközrendszert, mellyel a problémás esetek feloldhatókká válnak. A bizonytalanságból adódó megakadások tehát nem a beszélés problémái, hanem a beszélők problémamegoldó stratégiájának működési eredményei (Gósy 2003: 275). A beszélő egy-egy megakadás elkövetése utána korrekciós műveleteket hajthat végre. A hibák felismerése és javítása tehát az általános folyamatműködés közben zajlik (Gósy 2008). A felnőtt nyelvhasználókban működik egy bizonyos „önmonitorozó mechanizmus”, amely felfedezi a hibát, majd megszervezi és elvégzi a javítást (Gósy 2009: 141). Beszédprodukciónk ellenőrzésére kétféle eszközt használhatunk fel: egy rejtett (belső) monitorozást végezhetünk, amely a belső tervezéskor megy végbe, illetve egy felszíni (külső) monitorozást is alkalmazhatunk, amikor a beszédünk megértése által ellenőrizzük már elhangzott közléseinket. A szóhelyettesítő *izére* jellemző a javíthatóság, ahogy az a következő példában is látszik:

[3]

F: oké (.) mit csináltak?

Gy: megették a *izét*? (.) a kajájukat?

Ezzel szemben a töltelékszó *izé* nem javítható, mivel ezt nem egy szó vagy egy kifejezés helyettesítőjeként használjuk, hanem a beszédjog fenntartására.

[4]

\*FW1: mit szeretsz még kint játszani , XY ?

\*CH1: *öö@fp* focit ész [: és] *mmm@fp izé* mm futni .

\*FW1: mhm .

Schirm (2013: 326) felsorolja a leggyakoribb diskurzusjelölőket, melyek között az *izé* is helyet kapott, azonban a vizsgált elem nem tipikus diskurzusjelölő. Dolgozatomban az *izé* beszédtervezői, hezitáló funkcióját vizsgálom meg részletesebben, amely a beszélő bizonytalanságából ered.

A diskurzusjelölőkre jellemző a terminus sokfélesége (Dér 2009), azonban ezek az elnevezések (pl. modális partikula [Kiefer 1988], pragmatikai kötőszó [Németh T. 1998], diskurzusjelölő [Dér 2005]) különböző formai vagy tartalmi jellemzőt emelnek ki. A diskurzusjelölők szófaji besorolása (kötőszó, ige, módosítószó, interakciós mondatszó) sokféle, emiatt az elnevezésben a funkcionális, pragmatikai kategóriát hangsúlyozó, „semleges” (Schirm 2011: 10) megnevezést választottam (Kondacs 2016: 45).

Bruce Fraser (1999: 938) egy kéthelyiértékű viszony (S1 és S2) relációjában ábrázolja a diskurzusjelölőket (DJ), amelyben a funkciójuk az általuk bevezetett és a korábbi diskurzusrész kapcsolatának explicitté tétele (Kondacs 2016: 45). A diskurzusjelölők tehát arról adnak információt, hogy hogyan értelmezzük az S2 által hordozott információt az S1 interpretációjához képest (Fraser 1998: 302). Pl. Megjöttek a szüleim (=S1), pontosabban (=DJ) az édesanyám (=S2) (Schirm 2011: 13). A pontosabban diskurzusjelölővel a beszélő önjavítást végez, tehát pontosítja az elhangzott diskurzusrész és az új kapcsolatát. Az *izé* esetén azt láthatjuk, hogy nem az a funkciója, hogy egy kéthelyiértékű viszony kapcsolatát egyértelművé tegye, hanem pótolhat, pl. Gy: *vagy melyik izé harapta le?(0.1) vagy Pumukli (.) vagy Csöpi.* ,vagy időt biztosíthat a beszélőnek megnyilatkozása teljessé tételére. Például:

[5]



\*CH1: és akkor ész [: és] akkor ettek meg **izé** öö@fp hát .

\*FW1: igen ?

Fraser a diskurzusjelölők textuális szerepét hangsúlyozza, azonban ha hezitáló, beszédtervezői funkcióban szerepel az *izé*, akkor használatuk oka a pótlás vagy az időbiztosítás.

Furkó (2011: 53) összegzi a relevanciaelméleti felfogást a diskurzusjelölők szerepéről, amelynek lényege, hogy optimalizálják a megnyilatkozások feldolgozását, tehát felgyorsítják az interpretációs folyamatot azáltal, hogy „az előtérbe kerülő megnyilatkozások, illetve a közös kognitív környezet (szövegelőzmény, szöveggörnyezet) egyes elemei közötti kapcsolatokra nézve instrukciókat adnak”. Ezek az elemek tehát a diskurzus ragasztóanyagaiként biztosítják a koherenciát (Zhao 2014). A diskurzusjelölők tehát megkönnyítik a hallgatónak, hogy elérje az optimális relevanciát azáltal, hogy korlátozza a hallgató megnyilatkozásainak interpretációját az optimális relevancia felé. Az *izé*-féle beszédtervezői, hezitáció funkciójú elemek használatának fő funkciója azonban nem az, hogy korlátozzák a megnyilatkozás interpretációját, hanem hogy a beszélő saját kognitív zavarait leplezze használatuk által (Németh T. 2016). Ezek az elemek tehát nem az interpretációs folyamat megkönnyítésére használatosak. Az *izé* azonban diskurzusjelölőként is megjelenhet, ugyanis tabukerülésre vagy beszélői attitűd kifejezésére is használható, viszont ezzel tanulmányomban nem foglalkozom részletesebben.

A beszédkutatók, társalgáselemzők megállapításai szerint ezek az elemek hatással vannak a beszélőre és a hallgatóra egyaránt. A beszélőnek gondolkodási időt adnak, amely által az információdúsabb megnyilatkozásait elő tudja készíteni (Huszár 1985), illetve segítséget nyújtanak a beszélő saját közlésének megtartására, avagy átadására irányuló törekvésehez is (Schachter et al. 1991), s ezek tipikus diskurzusjelölői tulajdonságok.

A hallgatónak pedig szintén, mint „időnyerési lehetőségek” (Gósy 2003: 275) segíthetnek a feldolgozási műveletek elvégzésében, azaz az elhangzottak pontos megértésében van jelentőségük. A relevanciaelméletben a hezitáló, beszédtervezői *izé* használata a beszélőnek a kognitív zavarai leplezésében segít, míg használata a hallgató számára az optimális relevancia eléréséhez szükséges műveleti erőfeszítést könnyíti meg, nem az optimális kontextuális hatás elérését (Németh T. 2016).

Az *izével* kapcsolatban mind a felnőtt, mind a hatévesnél idősebb korosztály nyelvhasználatában végeztek már kutatásokat. Gyarmathy–Neuberger (2013) megállapította, hogy a beszédprodukciónak a nehézség alapján felírható kétféle *izé* közül a szóhelyettesítő *izé* a gyakoribb, kb. 62% mindkét korcsoportban. Az általuk vizsgált 94 felnőtt beszélőről elmondható, hogy 20–76 éves kor közöttiek voltak. A vizsgált 70 gyermek pedig 6–14 éves korú volt. A töltelékszó *izé* kb. 38%-ban volt jellemző a két korcsoportban. Az *izé* javítása esetén a következő eredmények születtek (Gyarmathy–Neuberger 2013): a felnőttek 51%-ban javították az *izét*, míg a gyerekek 60%-ban tudták korrigálni a vizsgált elemet, tehát 6 éves korban már működik az önmonitorozó mechanizmus (Gósy 2009: 141).

Gyarmathy (2012: 187) kutatásában a szóhelyettesítő *izé* szófaji megoszlását mutatja be 94 felnőtt beszélő nyelvhasználatában. A 86 db *izé* szófaji eloszlása nála a következő volt: 89,77%-ban főnevet, 5,68%-ban melléknevet, 3,41%-ban igét, míg 1,14%-ban igenevet helyettesítettek a vizsgált elemmel.

Gyarmathy (2012: 189) a szóhelyettesítő *izé* szófaji megoszlásának javítását is megvizsgálta. Megállapította, hogy a felnőtt beszélők leggyakrabban főnevet javítanak (86,36%), majd melléknevet (9,09%-ban). A nem javított *izék* esetében szintén a főnév

volt a leggyakoribb szófaj (93,18%), de előfordultak még igék (4,54%), melléknevek (2,27%).

A beszédkutatók számos vizsgálatot végeztek a gyermeknyelv megakadásainak sajátosságairól. Megállapították (Gósy 2009, Horváth 2006), hogy a 6 éveseknél az 5. leggyakoribb megakadásjelenség a töltelékszó, míg a felnőtt nyelvhasználóknál a 2. helyen áll, s az is kiderült, hogy a felnőttek dupla annyi töltelékszót használnak, mint a 6 évesek. Ezzel szemben a 3 éveseknél jellemző diszharmóniás jelenségek az ismétlések, illetve a néma és a kitöltött szünetek.

Dolgozatomban arra keresem a választ, hogy megjelenik-e az *izé* mindhárom óvodás korosztály, azaz a kis-, a középső- és a nagycsoportosok nyelvhasználatában. Kiinduló hipotézisem, hogy a vizsgált korosztály spontán beszédében a szóhelyettesítői *izé* található meg nagyobb arányban, nem a töltelékszó. Feltételeztem, hogy a szóhelyettesítő *izén* belül a szókereső *izék* aránya magasabb lesz, mint a szópótlóké. Emellett feltételeztem még, hogy a szóhelyettesítő *izé* szófaji vizsgálata során megállapítható, hogy a főnevet helyettesítik leggyakrabban a gyermekek a vizsgált elemmel. Úgy vélem, hogy az óvodásokra is jellemző lehet az *izé* korrekciója, amely során leggyakrabban a főnevet javítják.

## 2 Metodológia: anyag, módszer, kísérleti személyek

A kutatáshoz a saját korpuszomat (Kondacs 2017) és a Magyar Óvodai Nyelvi Korpusz (MONYEEK)<sup>iv</sup> anyagát használtam fel. A saját korpuszomban összesen 148 felvétel készítem, 61 különböző beszélővel. A saját korpuszom 70 310 szóból épül fel, a felvételek teljes időtartama 14 óra 48 perc. Az adatközlők egy csongrád megyei óvoda egyetlen kiscsoportjából, két középső és két nagycsoportjából valók. Életkoruk: 2,9–6,9 év. Az adatrögzítés 2016 őszén történt, csoportszobán kívül, egyesével. Az alkalmazott módszerem történetmesélés volt képek és rövidebb-hosszabb bábmesék segítségével. A MONYEEK-ben pedig 62 óvodással készítették interjút az adatrögzítők. Életkoruk: 4,5–5,5 év. A felvételek 2012 tavaszán készültek, több budapesti óvodában. A korpusz 37.714 megnyilatkozása összesen 178.377 szóból áll (Mátyus 2015: 48). A fiúkkal és lányokkal készített felvételek aránya közel azonos. A felvételeket a csoporton kívül készítették el. A gyermekeknek többféle feladatot kellett megoldaniuk: történetmesélés hallás után képek segítségével, egy képtörténet elmesélése önállóan, egy 4–4 képből álló történet elmesélése, egy játékszabály ismertetése, végül a „szép” beszédről való elmélkedés (Mátyus–Orosz 2014). Mindkét korpuszban előkerültek a gyermekeket érdeklő spontán beszédtemák.

A két korpuszban 120 óvodás közül 15 beszélőnél 62 *izé* fordult elő. A MONYEEK-ben 62 beszélőből 9-nél jelent meg (5 fiú, 4 lány), míg a saját korpuszomban 60 beszélőből 6 beszélő nyelvhasználatában (3 fiú, 3 lány) jelent meg az *izé*: három kiscsoportosnál, egy középsősnél és két nagycsoportosnál. Jellemző volt még, hogy egy-egy óvodásnál több *izé* is előfordult. A MONYEEK-ben egy adatközlő 35 *izét* használt spontán beszédében, tehát nála ténylegesen egy bizonyos megakadási stratégia volt megfigyelhető (Gósy 2003: 271).

<sup>iv</sup> Köszönettel tartozom Mátyus Kingának a MONYEEK-adatbázis elérésében nyújtott segítségéért, illetve a szegvári Kurca-parti Óvodának, hogy lehetőséget biztosított a gyermeknyelvi anyagok rögzítésére.

### 3 Eredmények

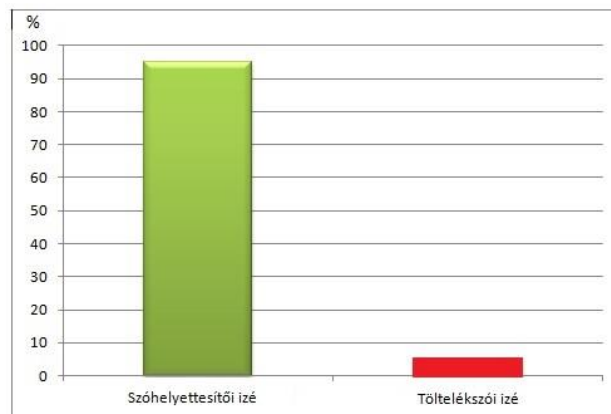
Kutatásom kiinduló hipotézise az volt, hogy a szóhelyettesítő *izé* fordul elő nagyobb arányban az általam vizsgált korosztályban. Az 1. ábrán látható, hogy a szóhelyettesítő *izé* aránya 95% volt, tehát 59 darab szóhelyettesítő *izé* volt az anyagban, míg a töltelékszó *izé* mindössze 5%-ban fordult elő, azaz 3 darab jelent meg belőle a korpuszban. Az *izét* használó 15 beszélő átlagosan 3,933 szóhelyettesítői, míg 0,2 töltelékszó *izét* használt. Azonban mivel az átlagérték nem teljesen reprezentálja a mintát, ezért a szórást, az átlagtól való átlagos eltérést is kiszámoltam. Ennek mértéke a szóhelyettesítő *izé* esetén 8,36, a töltelékszóinál 0,54. Mivel a szórás számomra önmagában nem volt elég kifejező, ezért a relatív szórást is megállapítottam, amely által még inkább érzékelhetővé vált a mintában lévő kiugró érték, amely nem reprezentálja az átlagot. A szóhelyettesítő *izé* esetén 220%, míg a töltelékszóinál 271% volt a relatív szórás. A kétféle *izé* elkülönítésénél segítséget nyújtott Fabulya (2007) jellemzése. Azonban a hanganyagok megléte ellenére is akadtak olyan problémás esetek, ahol nem teljesen egyértelmű, hogy melyik típusú *izéről* van szó. A következő példa szemlélteti ezt:

[6]

F: NA mit láttál ?

Gy: öö (.) egy **izét**, (0.3) **izét**, **izét** öö (0.2) várjál már, (.) várjál, várjál, várjál, várjál öhm, (0.2) öö nem tudom. (0.1) máma voltuk a kastályba?

Ebben a példában az első *izét* szókeresőnek vettem, mivel utána egy hosszabb néma szünetet tart a kisgyerek, míg a 2. és 3.-at töltelékszóinak írtam fel, mivel a kisgyermek célja ekkor már nem a szó megtalálása, hanem a beszéd jogának fenntartása volt, melyet a hezitáció (*öö*) és az ige halmozása is mutat (*várjál, várjál*). Habár mindhárom *izé* ragozott alakú (*izét*), de ez a ragozás az adatrögzítő által feltett kérdés (NA mit láttál?) szintaktikai kényszere miatt jelenik meg a töltelékszó *izé* esetében is.



1. ábra. szóhelyettesítő és a töltelékszó *izék* megoszlása 15 óvodás beszélő alapján

Kutatásomban feltételeztem azt is, hogy a szóhelyettesítő *izén* belül a szókereső *izé* aránya lesz a magasabb, mivel a gyermekek mindkét adatforrásként szolgáló korpuszban életkoruknak megfelelő képeket, meséket, témákat dolgoztak fel. Valóban a szókereső *izék* (77%) aránya volt a nagyobb, míg a szópótlóé csak 33% volt. A 15 *izét* használó beszélő alapján elmondható, hogy átlagosan 3,066 szókeresői, míg 0,866 töltelékszoói *izét* használtak. Szórás tekintetében a szókeresői *izé* 7,76, a hozzátartozó relatív szórás mértéke 253% volt, a szópótló *izé* átlagtól való átlagos eltérése pedig 0,8, amelyhez 93%-os relatív szórás kapcsolódik. A szókereső *izé* a következő példában látható: „Gy: a mesében is láttam. (.) Livi meg azt hitte, hogy az egy *izé*,(.) madár.”, míg a szópótló *izére* a következő példa mutatható be:

[7]

F: NA most mi lehet a mesében? (0.2) mi történt?

Gy: az a zöld *izé* MI volt?

F: hát káposzta mondjuk, (.) vagy saláta? .

Feltételeztem továbbá azt is, hogy a gyermekek leggyakrabban főnevet helyettesítenek az *izé* elemmel. A szókereső *izé* szófaji megoszlását a 2. ábra mutatja. Ahogyan látható a felnőtt nyelvhasználókhöz hasonlóan (Gyarmathy 2012: 187), leggyakrabban főnév (54,3%) helyett áll, ahogy azt az alábbi példa is mutatja.

[8]

\*FW1: minél részletesebben .

\*FW1: itt kezdődik .

\*CH1: őö@fp ott van a *izé* .

\*CH1: ott van a *izé* a kiszfiú [: kiszfiú] .

Ebben a példában a kisgyermek a kétszeri *izét* a *kiszfiú* főnév helyett használja. De előfordul igenév (4,4%) és ige (13%) helyett is, ez utóbbira láthatjuk a következő példát:

[9]

\*CH1: meghúzta a farkát .

\*CH1: és akkor *ráizézett*, ment .

\*CH1: rá (..) ment [:ráment] a macska a lányra .

A 9. példában a *ráizézett* alak a *ráment* ige helyettesítőjeként jelent meg.

Azonban a 3–7 év közötti gyermekeknél elég magas arányban (23,9%) jellemző volt még a kifejezések helyett álló *izé*. Pl.

[10]

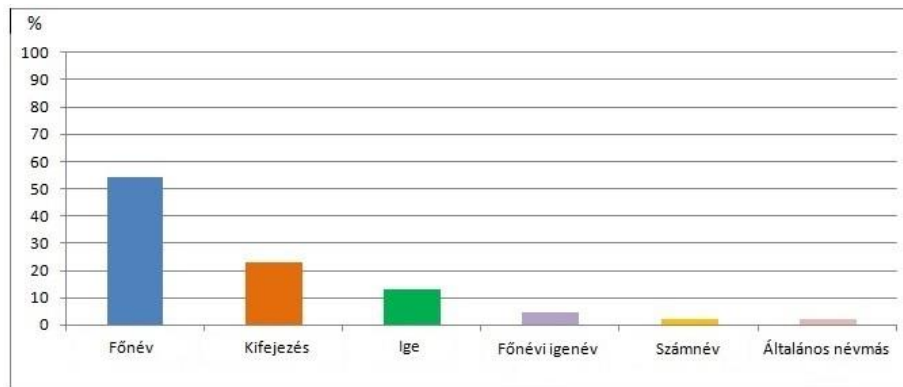
\*FW1: mi történt velük, XY ?

\*CH1: megkarmolta .

\*CH1: és akkor *izé* kiabált a kislány .

Ebben a példában a szókereső *izét* a kisgyermek egy egész kifejezéssel, a *kiabált a kislánnyal* javítja.

Emellett megjelent még egy számnév (2,2%) és egy általános névmás (2,2%) helyett is. A 15 *izét* használó óvodás átlagosan 1,666 főnevet, 0,733 kifejezést, 0,4 igét, 0,133 főnévi igenévet, 0,066 számnevet és általános névmást használt. Az átlagtól való átlagos eltérés értékei a következők: a főnév esetén 3,92 (relatív szórásérték 235%), az igenél 1,25 (rel.  $\sigma$ : 313%), a kifejezés esetén 2,49 (rel.  $\sigma$ : 339%), a főnévi igenév értéke 0,34 (rel.  $\sigma$ : 254%), a számnév és általános névmásnál 0,25 (rel.  $\sigma$ : 374%).



2. ábra. A szókereső *izé* szófaji megoszlása a gyermekek nyelvhasználatában.

A szópótló *izé* szófaji megoszlásáról elmondható, hogy itt is a főnév volt a leggyakoribb megjelenésű 92,3%-kal, míg igei 7,7%-kal egy volt. A 15 beszélő, akinek nyelvhasználatában megjelent az *izé* a szópótló *izén* belül, átlagosan 0,8 főnevet és 0,066 igét használt. A főnév átlagtól való átlagos eltérése 0,75, relatív szórásértéke 94% volt, míg az igének 0,25-ös szórása állapítható meg 374%-os relatív szórásértékkel. Pl.

[11]

F: Na mit láttál ?

Gy: a (.) a róka megette a **izét**? (.) a bácsi **arrébb izélte** a pókot .

A szópótló *izé* esetén a kisgyermek mentális lexikonjából hiányzik a keresett szó, ezért nem is tudja javítani az *izé* elemet. A példában megjelenő első *izé* egy főnevet pótol, melyet alátámaszt a *-t* tárgyrag is, míg a második, az *arrébb izélte* a pótlandó ige helyett áll.

A szóhelyettesítő *izé* jellemzője még a javíthatóság:

[12]

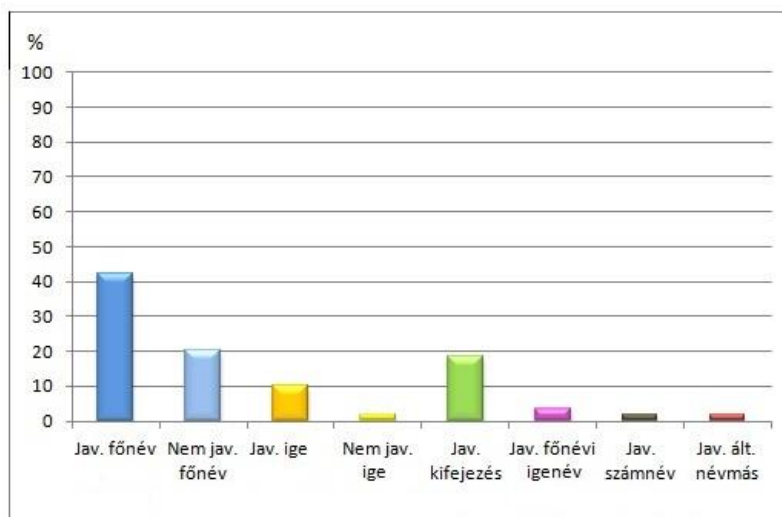
\*CH1: a **izét** nem szeretem, (.) a sóskát, meg a spenótot .

\*FW2: há:t azt nem sokan szeretik .

\*CH1: azt én se szeretem (.)

\*CH1: meg a, (.) meg a **izét** sem szeretem (.) a halas kenyért .

A javíthatóság már a 3 éveseknél is megfigyelhető, tehát működik az önmonitorozó mechanizmusuk. A 12. példában az óvodás az első *izé* alakot sóskára, míg a második *izét* halas kenyérré korigálja. Az 3. ábrán látható, hogy legnagyobb arányban a főnevet javítják (42,3%) és nem javítják (20,3%) az óvodások, de megjelent az ige is javított (10,2%) és nem javított (1,7%) formában. Javították még a kifejezéseket (18,7%), a főnévi igeneveket (3,4%), és egy-egy számnevet (1,7%) és általános névmást (1,7%). A gyermekek a szókereső *izét* minden alkalommal, míg a szópótló *izét* egyszer sem javították.



3. ábra. A szóhelyettesítő *izé* szófaji megoszlása a javítás alapján

#### 4 Következtetések

Dolgozatomban az egyik bizonytalanságból adódó megakadásnak, az *izé*nek jellegzetességeit vizsgáltam meg az óvodások diskurzusaiban. A kutatás során bebizonyosodott, hogy a vizsgált elemet már a nyelvelsajátítás korai szakaszában használják a gyermekek, és megfigyelhető volt még az is, hogy önmonitorozó mechanizmusunk (Gósy 2009: 141) már igen fiatalon is működik. A kapott adatok alátámasztják, hogy a fiatalabb és az idősebb korosztály *izé*használatára között nincsen számottevő különbség, csupán a százalékok terén tapasztalhatunk eltérést: (1) 95%-ban a szóhelyettesítő *izé* jelent meg, töltelékszó *izé* pedig csupán 5%-ban volt megfigyelhető. (2) Főként főnevet helyettesít (62,6%) a szóhelyettesítő *izé*, de megjelent ige (11,9%) és igenév (3,4%) helyett is. (3) Ha az óvodások javítják a szóhelyettesítő *izét*, akkor főként főnevet javítanak (42,3%), illetve nem javítanak (20,3%).

A mennyiségi mutatók alapján továbbá elmondható, hogy ha az egész korpuszt vesszük alapul, akkor 248 684 szóból mindössze 62-szer fordult elő a vizsgált elem, tehát átlagosan 4011 szavanként jelenik meg egy *izé* a korpuszban. A kapott eredmények azt mutatják, hogy az *izé* elég ritkán használt megakadásjelenség mind a gyermekek, mind a felnőttek nyelvhasználatában. Ennek okát Gyarmathy (2012: 181) fogalmazza meg, aki szerint a nyelvi elem stigmatizáltsága miatt a felnőtt beszélők kevésbé használják ezt az elemet, emiatt a gyermekek nem is hallhatják, ezáltal nem tudják utánozni, tehát nem tud megjelenni a spontán beszédükben. A felnőtt beszélőkre jellemző, hogy gyakran kicserélik az *izé* elemet. Helyette használják pl. a *cucc*, a *szar*, az *a valami*, a *tudodmi*, a *bizbasz* alakokat.

A kutatásom rávilágított arra, hogy a gyermekek nem egyforma mértékben és mennyiségben használják a különböző funkciójú *izé*ket, ami bizonyítja, hogy az óvodások egy bizonyos megakadási stratégiát (Gósy 2003: 271) alkalmaznak a spontán

beszédükben létrejött problémák áthidalására. Ezen mennyiségi változók megfigyelhetők az általam megvizsgált szórás és relatív szórás értékeiben is, hiszen az átlagtól eltérő kiugró értékek miatt minden megvizsgált esetben magas volt a két mérték.

## Irodalom

- Bybee, J. L. 2001. *Phonology and language use*. New York: Cambridge University Press.
- Clark, H. M. 1994. Managing problems in speaking. *Speech Communication*, 15: 243–250.
- Dér, Cs. I. 2005. A diskurzusjelölők kialakulásáról a grammatikalizációs paraméterek tükrében. In: Mártonfi, A., Papp K., Sliz M. (szerk.) 2006. *101 írás Pusztai Ferenc tiszteletére*. Budapest: Argumentum. 83–88.
- Dér, Cs. I. 2009. Mík is a diskurzusjelölők? In: Keszler, B., Tátrai, Sz. (szerk.) 2009. *Diskurzus a grammatikában – grammatika a diskurzusban*. Budapest: Tinta. 293–303.
- Fabulya, M. 2007. *Izé, hogyhívják, hogymondjam*. Javítást kezdeményező lexikális kitöltőelemek. *Magyar Nyelvőr*, 131(3): 324–342.
- Fraser, B. 1999. What are discourse markers? *Journal of Pragmatics*, 31: 931–952.
- Furkó, B. P. 2011. Diskurzusjelölők és szövegösszefüggés: a kohézió vagy a koherencia eszközei? Az implikáció mint a szövegösszefüggés eszköze. In: Dobi, E. (szerk.) *A szövegösszefüggés elméleti és gyakorlati megközelítési módjai: poliglott terminológiai és fogalmi áttekintés: Petőfi S. János 80. születésnapjára*. Debrecen: Debreceni Egyetemi Kiadó. 37–55.
- Gósy, M. 2002. A megakadásjelenségek eredete a beszédprodukción tervezési folyamatában. *Magyar Nyelvőr*, 126(2): 192–204.
- Gósy, M. 2003. A spontán beszédben előforduló megakadásjelenségek gyakorisága és összefüggései. *Magyar Nyelvőr*, 127(3): 257–277.
- Gósy, M. 2008. Önellenőrzési folyamatok a beszédben. *Magyar Nyelvőr*, 54(4): 402–426.
- Gósy, M. 2009. Önjavítási stratégiák a beszédben gyermekeknél és felnőtteknél. In: Bárdosi, V. (szerk.) 2009. *Quo vadis philologia temporum nostrorum? Korunk civilizációjának nyelvi képe*. Budapest: Tinta. 141–150.
- Gósy, M., Horváth, V. 2009. Hogyan tükrözi a kiejtés a nyelvi funkció változását? In: Keszler, B., Tátrai, Sz. (szerk.) 2009. *Diskurzus a grammatikában – grammatika a diskurzusban*. Budapest: Tinta. 26–33.
- Grétsy, L., Kemény, G. 2005. *Nyelvművelő kézikönyvtár*. Budapest: Tinta
- Gyarmathy, D. 2012. Az *izé* funkciófüggő realizációi. In: Gósy, M. (szerk.) 2012. *Beszéd, adatbázis, kutatások*. Budapest: Akadémiai. 178–194.
- Gyarmathy, D., Neuberger, T. 2013. *The Hungarian filler word izé in children's and adults' spontaneous speech*. Elhangzott: XV. Summer School of Psycholinguistics, Balatonalmádi, 2013. május 26–31.
- Hadrovics, L. 1992. *Magyar történelmi jelentéstan*. Budapest: Akadémiai.
- Horváth, V. 2004. Megakadásjelenségek a párbeszédokban. *Beszédkutatás 2004*. 187–199.
- Horváth, V. 2006. A spontán beszéd és a beszédfeldolgozás összefüggései gyerekeknél. *Beszédkutatás 2006*. 134–146.
- Huszár, Á. 1985. A rádió és a televízió beszélt nyelvének mondattana. In: Grétsy, L. (szerk.) 1985. *Nyelvészet és tömegkommunikáció*. Budapest: Tömegkommunikációs Központ. 73–117.
- Jefferson, G. 1984. On the organization of laughter in talk about troubles. In: Atkinson, M. J., Heritage, J. (szerk.) *Structures of Social Action. Studies in Conversation Analysis*. Cambridge: Cambridge University Press. 346–369
- Kiefer, F. 1988. Modal particles as discourse markers in questions. *Acta Linguistica Hungarica*, 38: 107–125.
- Kondacs, F. 2016. A *hát* diskurzusjelölőről az óvodai diskurzusokban, In: Várad, T. (szerk.) *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből*. 45–58.
- Mátyus, K. 2015. *A szociolingvisztikai stílus: Stratégiák a gyermek-felnőtt diskurzusban*. PhD-értekezés. Szeged: Szegedi Tudományegyetem. Elérhető: [http://doktori.bibl.u-szeged.hu/2593/1/Jelencsik\\_Matyus\\_Kinga\\_disszertacio.pdf](http://doktori.bibl.u-szeged.hu/2593/1/Jelencsik_Matyus_Kinga_disszertacio.pdf). Letöltve: 2017. október 6.

- Mátyus, K., Orosz, Gy. 2014. MONYEK – Morfológiailag egyértelműsített óvodai nyelvi korpusz. *Beszédkutatás* 2014: 237–245.
- Németh, T. E. 1998. *A hát, így, tehát, mert* kötőszók pragmatikai funkciójának vizsgálata. *Magyar Nyelv*, 94(3): 324–31.
- Németh T., E. 2016. *A diskurzusjelölők általános jellemzői*. Elhangzott: Az SZTE Pragmatikaelméletek c. PhD-kurszán, Szegedi Tudományegyetem, 2016. november 22.
- Németh, Zs. 2014. Javítási műveletek jelöltségi hipotézise. *Jelentés és Nyelvhasználat*, 1: 29–54. Elérhető: <http://www.jeny.szte.hu/jeny/images/issues/2014/JENY-2014-NemethZs.pdf>. Letöltve: 2017. október 6.
- Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J. L., Hopper, P. J. (szerk.) 2001. *Frequency effects and the emergence of lexical structure*. Amsterdam: John Benjamins. 137–157.
- Pusztai, F. 2003. *Magyar Értelmező Kéziszótár*. Budapest: Akadémiai.
- Schachter, S., Christenfeld N., Ravina B., Bilous F. 1991. Speech disfluency and structure of knowledge. *Journal of Personality and Social Psychology*, 60: 362–367.
- Schirm, A. 2011. *A diskurzusjelölők funkciói: a hát, az -e és a vajon elemek története és szinkrón státusa alapján*. PhD-értekezés. Szeged: Szegedi Tudományegyetem. Elérhető: [http://doktori.bibl.u-szeged.hu/759/1/schirm\\_anita\\_doktori\\_disszertacio.pdf](http://doktori.bibl.u-szeged.hu/759/1/schirm_anita_doktori_disszertacio.pdf). Letöltve: 2017. október 6.
- Schirm, A. 2013. Fókusz(csoport)ban a diskurzusjelölők. In: Kontra, M., Németh M., Sinkovics B. (szerk.) 2013. *Elmélet és empiria a szociolingvisztikában*. Budapest: Gondolat. 324–336.
- Zhao, H. 2014. The textual function of discourse markers under the framework of relevance theory. *Theory and Practice in Language Studies*, 4(10): 2105–2113.



## Koreferenciaviszonyok vizsgálata enyhe kognitív zavarban szenvedők beszédátirataiban<sup>1</sup>

Kovács Viktória

SZTE BTK Nyelvtudományi Doktori Iskola  
viktoria.kovacs12@gmail.com

**Kivonat:** A jelen tanulmányban az enyhe kognitív zavarban szenvedők beszédátiratain végzett koreferenciaannotációt mutatom be, valamint az annotálás eredményeül kapott adatokon végzett statisztikai számításokat és egy gépi tanulási kísérletet. Az enyhe kognitív zavar lefolyása során megfigyelhető a memória működésének romlása, ez tetten érhető a koreferenciaviszonyok használatán. A láncok mennyisége és hossza, valamint a különböző szintaktikai és szemantikai szabályszerűségeken alapuló visszautalási stratégiák vizsgálata segíthet az enyhe kognitív zavar korai diagnosztizálásában, valamint a kapott adatokon végzett gépi tanulási kísérletek alapján az automatikus azonosításban is.

### 1 Bevezetés

A kutatás során arra a kérdésre kerestem a választ, hogy az enyhe kognitív zavar (a továbbiakban: EKZ) lefolyása során egyre hanyatló emlékezeti rendszerek hatására sérül-e a koreferencialáncok használatának képessége. Ehhez egy, a Szegei Tudományegyetemen korábban zajló kutatáshoz készített korpuszt használtam fel (Vincze et al. 2015). A korpusz EKZ-ban szenvedő személyek és egy egészséges kontrollcsoport tagjai által elmesélt történetek átíratait tartalmazza, ezáltal alkalmassá válik arra, hogy a két csoport számítógépes nyelvészeti módszerekkel összehasonlítható legyen a koreferencialáncok használata alapján. Mivel az EKZ és a belőle kialakuló komolyabb mentális deficitiek gyógyíthatatlanok, kiemelt fontosságú a betegség korai diagnosztizálása. A különböző nyelvészeti megközelítésekben történő vizsgálatok eredményei segíthetnek a szakembereknek a diagnózis felállításában és ezzel a betegség lefolyásának lassításában.

A továbbiakban ismertetem azokat a beszédben fellelhető és nyelvtudományi szempontból releváns tüneteket, amelyek segítenek az EKZ diagnosztizálásában. A tünetek bemutatása után a koreferencia definícióját fogom megadni és ezzel együtt a kutatásom motivációját, valamint az előzetes hipotéziseimet. Ezután bemutatom magát a kutatást, ezen belül is a kutatás alapjául szolgáló korpuszt, az annotálási módszereket, valamint a kapott adatokat és az azokon végzett statisztikai számításokat. Végezetül pedig bemutatom egy gépi tanulási kísérletet, és levonom a kutatás konklúzióit.

---

<sup>i</sup> A kutatást az EFOP-3.6.1-16-2016-00008 azonosítójú, EU társfinanszírozású projekt támogatta.

## 1.1 Kognitív háttér

A kutatás alapját képező korpuszt EKZ-ban szenvedők beszédátiratai alkotják. A következőkben a deficit tüneteit és a memóriával kapcsolatos fogalmakat szükséges tisztázni. Az EKZ-nak nem pusztán a beszédben megjelenő tünetei vannak, viszont mivel nyelvészeti szempontból vizsgáltam a szövegeket, csak azokat és a memóriával kapcsolatos tüneteket veszem sorra.

### 1.1.1 Az enyhe kognitív zavar tünetei

Az EKZ az egészséges és a demens állapot közötti folyamatban jelenik meg köztes állomásként, tehát a diagnosztizálás során a demencia tüneteire kell odafigyelni (Negash et al. 2007). A demenciák közül az Alzheimer-kór az, amelynek a tünetei megjelenhetnek az EKZ-ban, egy sokkal enyhébb formában. Az Alzheimer-kór és az EKZ közötti erős összefüggést alátámasztja az a tény, hogy az idős, EKZ-ban szenvedők 50%-ánál három éven belül az Alzheimer-kórt is diagnosztizálják (Kempler 2005).

Az EKZ során jelentkező deficit az emlékezeti rendszerek tekintetében az epizodikus memóriát, a prospektív memóriát és a végrehajtó működéseket érinti (Tariska 2002; Turi–Németh–Hoffmann 2014). Az epizodikus memória az explicit (deklaratív) emlékezeti rendszer része, amely a konkrét eseményekre való emlékezést biztosítja (Baddeley–Eysenck–Anderson 2010), és ezekhez az eseményekhez kapcsolódó érzéseink tárolására alkalmas (Gósy 2005). A prospektív memória a terveinkre való emlékezést biztosítja (Baddeley–Eysenck–Anderson 2010), ezeknek az emlékeknek alacsonyabb az információtartalma, spontán beszédben nem is igazán fellelhető a romlása. A végrehajtó működések a terveink végrehajtásáért felelnek, legyen az egy megnyilatkozás kimondása vagy egy fizikai cselekvés végrehajtása.

A deficittel kapcsolatban meg kell említeni a munkamemóriát, amely ugyan az EKZ-ban még nem érintett, de az enyhe Alzheimer-kórban már igen. A munkamemória a rövid távú memória része, az új ismeretek feldolgozásáért, az adatrögzítésért és a szabályalkotásért felel, kapacitása egyénenként eltérő. A munkamemória-modell egyik fő komponense a központi végrehajtó, amely a feladatmegoldásért is felel, pl. a következtetésért, a nyelvi megértésért, a problémamegoldásért, valamint azért, hogy kapcsolatba tudjuk hozni az egymástól független információkat. A munkamemóriának az egyik perifériás alrendszere a fonológiai hurok, amely a beszédszerű információk feldolgozásáért felelős (Gósy 2005).

### 1.1.2 A beszédben megjelenő tünetek

Az EKZ diagnosztizálási nehézségei abból fakadnak, hogy az egészséges idős egyének neuropszichológiai vizsgálatokon elért eredményei nem sokban térnek el az enyhe kognitív zavarban szenvedők teljesítményétől. Ezzel párhuzamosan a spontán beszédben is csak nehezen figyelhetők meg a deficitre utaló jellegzetességek, hiszen az egészséges adatközlők is alkalmaznak különböző stratégiákat a szókeresésre.

Az EKZ eredményesebb szűréséhez hozzájárul egy, a Szegedi Tudományegyetemen zajló projekt (Vincze et al. 2015; Vincze et al. 2016), amely a deficit automatikus azonosítására törekszik beszédben fellelhető jellemzők statisztikai vizsgálata alapján.

Kutatásuk során az alábbi jellemzők statisztikailag szignifikánsnak bizonyultak az EKZ szűrésében: a szöveg szavainak száma, a hezitációs jelenségek száma és aránya, a bizonytalan (körülró, megnevezési nehézségre utaló) szavak száma és aránya, az ismeretlen szavak (az adatközlők által alkotott szavak) száma, mondatbeli szavak száma, a néma szünetek száma, a nyújtások száma, a főnevek száma, az írásjelek száma (Vincze et al. 2015).

## 1.2 Nyelvtudományi háttér

Az enyhe kognitív zavar tehát egy olyan kognitív deficit, amelynek az egyik legjellemzőbb tünete a rövidtávú emlékezet zavara. Olyan beszédből kinyerhető további információkat szükséges jelölni az annotáció során, amelyek az emlékezet zavarára utalhatnak, és nyelvészeti szempontból vizsgálhatók.

### 1.2.1 A koreferencia

A memória romlását jól mutathatja az is, hogy hogyan utalunk vissza egy korábban megemlített objektumra. A koreferencia két szövegbeli szó vagy szóelem, szókapcsolat között jön létre a szövegvilágon belüli ugyanazon dologra utalással, melynek alapja a referencia. A második szóelem az elsőre utal vissza, és rajta keresztül referál a már említett dologra (Tolcsvai 2003). Tehát a koreferencia objektumok és események említése és az azokra való folyamatos visszautalás. Ezek a visszautalási formák különböző szintaktikai és szemantikai szabályszerűségeken keresztül valósulnak meg a kifejezések között. Az azonos objektumokra és eseményekre vonatkozó kifejezések ezáltal láncokat alkotnak a szöveg mentén. A koreferencia bonyolult pszicholingvisztikai, valamint logikai jelenség, hiszen a mondandónk részben ez által lesz logikusan felépített, tartalmas egység.

### 1.2.2 Visszautalási kategóriák

A visszautalási kategóriák a koreferencialánc egymást követő tagjai közötti szintaktikai, illetve szemantikai kapcsolatokon és a visszautaló szó szófaján alapulnak. Számos jellemző felsorolható, amelynek a segítségével antecedens választható a visszautaló szóhoz, pl. a két kifejezés közötti szemantikai kapcsolat, a kötési elvek vagy a megnyilatkozás kontextusából származó információk. Az annotálás szempontjából olyan kategorizálást kell alkalmazni, amely a lehető legáttekinthetőbb és leggazdaságosabb. A kategóriák megválasztása során az egyik szempont az volt, hogy elkülönítsem egymástól a deiktikus kifejezéseket, ezen belül is a térre és időre, valamint a személyekre és objektumokra vonatkozó kifejezéseket. Egy másik szempont az események és objektumok elkülönítése. A harmadik pedig az objektumok közötti szemantikai kapcsolatok figyelembevétele.

Az annotálás során ezek alapján a következő kategóriákat különböztettem meg: Adverbial (határozó), Apposition (értelmező), Epithet (jelző), Meronym (meronímia), Hypernym (hiperonímia), None (egyik sem), Other (egyéb), Pronominal (névmás), Repetition (ismétlés), Synonym (szinonímia), Verbal (igei).

A következő példákban a visszautaló szót félkövérrel emeltem ki, a (...) jelölés a példa megszakítottságát jelöli.<sup>ii</sup> Mivel a kutatás középpontjában a memória működése és az abban bekövetkező változások felismerése áll, ezért a visszautalási kategóriát mindig a visszautaló szó és a láncban őt megelőző elem közötti kapcsolat határozta meg.

Az Adverbial (határozó) kategóriába tartozott a határozói visszautalásnak az összes típusa. Ezek többnyire helyhatározók voltak, de előfordult a szövegekben időhatározó is, mivel a térre és időre vonatkozó információk az elmesélt történetek kontextusát határozták meg, egyaránt ebbe a kategóriába kerültek.

[1] Hát egy bizonyája egy vendéglő volt, **ahol** leült ugye a vendég.

Az Apposition (értelmező) kategória az értelmezést és a kiegészítő magyarázat típusú elemeket tartalmazza. Azok a kifejezések kerültek ebbe a kategóriába, amelyek valamilyen információ hozzáadásával azonosították az objektumot vagy az eseményt, amire visszautaltak, tehát többlettartalmat adtak hozzá, értelmezték azt. Ezek a visszautalások azért nem egyszerűen az ismétlés kategóriába kerültek, mert a többlettartalom hozzáadása fontos volt az objektum beazonosításának szempontjából. A többlettartalom hiányában nem lett volna egyértelmű, hogy melyik láncba tartozik az elem.

[2] Hát valahol ... egy asztal két szék ... egy férfi már ül ... a másik meg jön befelé. (...) **Az aki ült férfi** elővette a botját.

Epithet (jelző) kategóriába kerültek azok a visszautalások, ahol egy objektumot egy tulajdonsága alapján azonosított be az adatközlő, és ennek segítségével utalt vissza rá. Ez a kategória is az objektumok elkülönítésében játszott szerepet, viszont ebben a kategóriában nincs ismétlés.

[3] Férfi ilyen nyeles fával volt (...). De **a hálónál** van egy virág...

A Hypernym (hiperonímia) kategóriába a két kifejezés közötti fogalmi alá-fölé rendeltség alapján kerültek a visszautalások.

[4] ... de valami madzagon volt ez a bagoly (...) hogy talán el akarta fogna **a madarat**.

Azokban az esetekben, amelyekben az egyik szó által jelölt dolog része volt a másik szó által jelölt dolognak, a visszautaló szó a Meronym (meronímia) kategóriába került. Ezekben a szövegekben ez leggyakrabban testrészeket és bűtordarabokat jelentett. Ezek a kifejezések nem alkottak külön koreferencialáncot, mivel csak egy-egy alkalommal említették őket az adatközlők.

[5] Akkor .. utána ugye jött egy .. hölgy , gondolom én , hogy .. helyet foglalt és akkó kínálta volna , **fejére** öntötte az italt ...

A None kategóriába minden esetben a láncok első elemei kerültek, mivel ezek még semmire sem utaltak vissza.

<sup>ii</sup> A hanganyagok átiratai számítógépes feldolgozásra készültek, ezért a további központosítási jelek a számítógépes feldolgozás szempontjából jelentősek, mivel a koreferenciaannotációt nem befolyásolják, változtatás nélkül közlöm a példákat. Bővebben az átiratok elkészítésénél figyelembevett módszertani elvekről: (Vincze et al. 2015, Vincze et al. 2016).

A Pronominal (névmás) kategóriába azok a visszautalások kerültek, amelyekben a visszautaló szó személyes, mutató vagy vonatkozó névmás volt. Ezek a visszautalások a jellemzőkészlet gazdaságossága érdekében kerültek egy kategóriába, de elválasztva a határozói visszautalásokhoz tartozó névmásoktól és a határozatlan névmásoktól.

[6] Hát szerintem... az illető az süketnéma volt ... mer.. azé mutogatott **neki** a másik

Az egyszerű ismétléssel történő visszautalások a Repetition (ismétlés) kategóriába kerültek. Ebbe beletartoznak azok a visszautalások is, amelyek egyébként más kategóriába tartoztak volna.

[7] Behívja az inast aki (...) jön utána a ... az **az inas** ...

[8] És akkor ő odavitte neki. Nem, először vitt **neki** virágot.

A Synonym (szinonímia) kategóriába a két kifejezés közötti szinonímián alapuló visszautalások kerültek.

[9] Aztán a hölgy le akar ülni (...) Éső lekuporodik **a nő** és fölötte vívna.

Az eseményekre való visszautalás a Verbal (igei) kategóriába tartozik, mivel ilyen esetekben a lánc feje legtöbbször ige. Elhanyagolható mennyiségű példa volt arra, hogy pontosan ugyanarra az eseményre utaljon vissza az adatközlő, viszont az áttekinthetőség érdekében érdemes elkülöníteni az objektumokra történő visszautalásoktól.

[10]Hát az , hogy kiment ez a részeg ember , ki **ki** . **osont** vagy **ki** . **suttyogott** sunnyogott.

Other (egyéb) kategóriába azok a visszautalások kerültek, ahol valamely kifejezésnél megnevezési nehézségei akadtak az adatközlőnek, ezt legtöbbször határozatlan névmással fejezte ki, de ettől függetlenül egyértelmű volt az, hogy mire utal vissza.

[11]...valószínű hogy gyűrűt vett a férfi elő . Kis dobozba ... volt **valami**.

### 1.3 Hipotézisek

Mivel a koreferenciaviszonyok használatához szükséges a memória épsége, ezért feltételeztem, hogy az enyhe kognitív zavarban szenvedők az egészségesektől eltérő stratégiát alkalmaznak a visszautalásra ugyanazon történet elmesélése során. Elsődleges hipotézisem az volt, hogy az enyhe kognitív zavarban szenvedők kevesebb és rövidebb koreferencialáncot fognak használni ugyanazon történet elmeséléséhez. Másodsorban pedig, hogy az enyhe kognitív zavarban szenvedők csoportjában az ismétlés a legtöbb, a kontrollcsoportban pedig a szinonímia és egyéb, a két kifejezés közötti bonyolultabb kapcsolatok felismerését feltételező visszautalási formák használata a leggyakoribb.

## 2 Kísérleti személyek, anyag, módszer

A kutatás a már említett Szegedi Tudományegyetemen zajló projektet veszi alapul, amely az enyhe kognitív zavar automatikus azonosítására törekszik (Vincze et al. 2015; Vincze et al. 2016).

### 2.1 Korpusz

A kutatáshoz készített korpuszt a kísérleti személyek által elmesélt történetek beszédátiratai alkotják. A hanganyagok felvétele minden esetben a Szegedi Memóriaambulancián történt, majd ezeket később jegyezték le nyelvészek. A hanganyagok lejegyzése során figyelembe vették a spontán beszéd sajátosságait, mint a néma szünetek: *előveszi a ... .. görbebotját*, vagy kitöltött szünetek: *Őő és akkor*, a kétszer indítások: *ért ... . értettem belőle*, fonológiai nyújtások: *képrű*. Az átiratok nem tartalmaznak azonban az intonációs jelenségekre vonatkozó információkat.

### 2.2 Kísérleti személyek

Összesen 86 adatközlő vett részt a felmérésben, ez 172 átíratot jelent, amelyből 35 adatközlő, azaz 70 átírat alkotta a kontrollcsoportot. A 35 kontrollszemély közül 23 volt nő és 12 férfi, az 51 enyhe kognitív zavarban szenvedő beteg közül 33 volt nő és 18 férfi. Az adatközlők mind magyar anyanyelvűek voltak, és rendelkezésemre állt a klinikai diagnózisuk. A feladatuk egységesen az volt, hogy két fekete-fehér rövidfilm megnézése után elmeséeljék az első filmet, majd beszéljenek a legutóbbi napjukról, ezután meséeljék el a második filmet.

### 2.3 Módszertan

Az annotálás az MMAX2 annotációs eszközzel és kizárólag manuálisan történt a hibázási lehetőségek kiküszöbölése miatt. A kutatáshoz a hanganyagok semmilyen formában sem lettek felhasználva. Az elliptikus szerkezetek és a morfológia szintjén jelölt kapcsolatok – mivel az annotációs eszköz nem volt rá alkalmas – nem lettek jelölve. Azokat az eseteket, amelyek nem voltak egyértelműek, vagy hezitációnak minősíthetőek voltak, nem kerültek bele a koreferencialáncokba az áttekinthetőség és egyértelműség végett. A két csoport átíratait véletlenszerű sorrendben, anonimizálva annotáltam, hogy a klinikai diagnózis ismerete ne befolyásolja az annotációt.

Az annotálás után az átíratokat két csoportba soroltam a klinikai diagnózis alapján, majd először az átíratoknak olyan általános tulajdonságait gyűjtöttem ki, mint a szövegek hossza, a bennük található láncmennyiség és lánchossz. Ezek után gyűjtöttem ki a két csoportban külön a használt visszautalási kategóriák számát. A kapott adatokon ezután statisztikai számításokat végeztem, hogy megvizsgáljam, melyek azok az átíratokból kinyerhető jellemzők, amelyek segíthetnek az enyhe kognitív zavar azonosításában és az egészséges állapottól való elkülönítésében.

A kategóriákra vonatkozó adatok alapján ezután gépi tanulási kísérleteket végeztem a Weka szoftver (Hall et al. 2009) segítségével. Ennek alapja a már említett Szegedi

Tudományegyetemen zajló projekt volt (Vincze et al. 2016). A gépi tanítás során ők a következő jellemzőkészletet használták fel: A spontán beszéden alapuló jellemzőik voltak a kitöltött és kitöltetlen szünetek száma, a hezitációk száma és aránya a tokenek számához képest, a névelő és a tartalmas szó közötti szünetek száma és a nyújtások száma. A morfológiai jellemzőik a tokenek és szavak száma, a lemmák száma és aránya, a középpontozási jelek száma, a főnév, ige, melléknév, névmás, kötőszó száma és aránya, az egyes szám első személyű igék száma, az ismeretlen szavak száma. Szemantikai tulajdonságaik a bizonytalan szavak száma és aránya az összes tokenhez képest, az emlékezési tevékenységgel kapcsolatos szavak száma és aránya, a tagadó szavak száma, a tartalmas és funkciósavak száma és aránya, a film tartalma szempontjából kulcsszavak száma. A demográfiai jellemzőik a nem, életkor és képzettség (Vincze et al. 2016). Az általuk használt jellemzőkészletet egészítettem ki az annotált visszautalási kategóriákkal. Minden beszélő esetében hozzáadtam jellemzőként az általa használt egyes visszautalási típusok számát az egyes szövegekre lebontva.

### 3 Eredmények

A következőkben a szövegekkel és a kategóriákkal kapcsolatos adatokat, valamint az azokon végzett statisztikai számításokat ismertetem, majd a gépi tanulási kísérletek közül is bemutatom azt, amelyik a legeredményesebb lett.

#### 3.1 Az átiratokkal kapcsolatos adatok

Először az átiratok adatait vizsgáltam meg az egyes adatközlők alapján. Mivel a két történet elmesélése között eltelt idő is befolyásolhatja az eredményeket, így a két történet átiratait is elkülönítettem.

|                                 | EKZ         |             | Kontroll    |             |
|---------------------------------|-------------|-------------|-------------|-------------|
|                                 | 1-es szöveg | 2-es szöveg | 1-es szöveg | 2-es szöveg |
| <b>Láncok átlagos száma</b>     | 4,53        | 5,02        | 5,66        | 5,40        |
| <b>Láncok átlagos hossza</b>    | 2,31        | 2,32        | 2,32        | 2,42        |
| <b>Leghosszabb szöveg (szó)</b> | 369         | 222         | 206         | 202         |
| <b>Legrövidebb szöveg (szó)</b> | 24          | 29          | 37          | 52          |
| <b>Átlagos szöveg (szó)</b>     | 109,14      | 95,71       | 104,37      | 97,86       |

1. táblázat. A szövegek adatai. EKZ: enyhe kognitív zavar.

A szöveggel kapcsolatos adatok alapján elmondható, hogy a kontrollcsoport valóban több koreferencialáncot használ átlagosan az egyes szövegekben, a láncok hossza között azonban nem mutatkozik különbség. Az enyhe kognitív zavarban szenvedők csoportjának átiratai – a deficitből adódóan – valószínűleg nagyobb mértékben tartalmaznak hezitációs jelenségeket, mint a kétszer indítás és a gyakori szóismétlés, ez pedig befolyásolja a leghosszabb és legrövidebb elmesélt történet jellemző eredményeit

az egyes csoportokban. A leghosszabb szöveg jellemző nagy eltérést mutat a két csoport között, az EKZ-csoport kiugró értékét a hezitációs jelenségek okozhatják. Tehát az, hogy az adatközlő sokkal több szót használ, nem jelenti azt, hogy tartalmasabban fejezte ki magát, ezért ez a jellemző nem alkalmas a két csoport összehasonlítására. A legrövidebb szöveg jellemző alapján a kontrollcsoport tartalmasabban fejezte ki magát. Az átlagosan használt szavak száma szövegenként a leghosszabb szöveg jellemző miatt szintén nem alkalmas az összehasonlításra.

### 3.2 A kategóriákkal kapcsolatos adatok

Az átiratokon végzett annotáció után a jelölt kategóriákat kinyertem az egyes átiratokból és csoportonként összesítettem.

|                   | EKZ  |          | Kontroll |          |
|-------------------|------|----------|----------|----------|
|                   | szám | százalék | szám     | százalék |
| <b>Adverbial</b>  | 13   | 1,15%    | 15       | 1,64%    |
| <b>Apposition</b> | 48   | 4,26%    | 33       | 3,61%    |
| <b>Epithet</b>    | 6    | 0,53%    | 5        | 0,55%    |
| <b>Hypernym</b>   | 16   | 1,42%    | 10       | 1,09%    |
| <b>Meronym</b>    | 43   | 3,64%    | 40       | 4,36%    |
| <b>Other</b>      | 76   | 6,75%    | 42       | 4,60%    |
| <b>Pronominal</b> | 404  | 35,88%   | 284      | 31,04%   |
| <b>Repetition</b> | 407  | 36,15%   | 410      | 44,81%   |
| <b>Synonym</b>    | 111  | 9,86%    | 73       | 7,98%    |
| <b>Verbal</b>     | 4    | 0,36%    | 5        | 0,55%    |

2. táblázat. A kategóriák adatai. EKZ: enyhe kognitív zavar

A kategóriákkal kapcsolatos adatok alapján kiemelkedő mind a két csoportban az ismétlések (repetition) száma. Megfigyelhető, hogy amíg az EKZ-csoport ismétlései az összes visszaulásnak a 36,15%-át teszik ki, addig a kontrollcsoport ismétlései a 44,81%-át. Ennek fényében az a hipotézisem, hogy az enyhe kognitív zavarban szenvedő adatközlők több ismétlést fognak használni a kontrollcsoportnál, nem helytálló.

Érdeemes azonban a többi kategóriát is megvizsgálni. A második leggyakrabban használt kategória szintén egyöntetűen mind a két csoportnál a névmási (pronominal) kategória volt. Ez az EKZ-csoportban 35,88%-ot jelentett, tehát majdnem megegyezett az ismétlések számával. A kontrollcsoportban azonban csak 31,04%-ot ért el, ami elég nagy eltérés az ismétlések arányától. A harmadik leggyakoribb kategória a szinonímia volt, ez az EKZ-csoportban 9,86%, a kontrollcsoportban pedig 7,98%.

### 3.3 Statisztikai vizsgálatok

A statisztikai adatok számítása során a University of Washington által kiadott elveket követtem (Writing an APA Empirical (lab) Report, 2010). Mivel a kontrollcsoporton belül a kategóriák használati arányai között ekkora különbségek voltak, de ez az enyhe kognitív zavarban szenvedők csoportjára nem volt jellemző, a kategóriákon



alkalmaztam a  $\chi^2$  statisztikai eljárást. Az eredmény azt mutatja, hogy szignifikáns az eltérés a kategóriák eloszlásának tekintetében a két csoport között ( $p = 0,010$ ).

A kapott adatokon egyéb statisztikai vizsgálatokat is végeztem, hogy meg tudjam vizsgálni az egyes kategóriák szerepét a két csoport tagjainak elkülönítésében. Ezért minden egyes kategóriára, szövegenként kétmintás  $t$ -próbát végeztem. A kutatás során a két minta szabadságfoka kontrollcsoport: 34, EKZ-csoport 50. A kategóriákhoz tartozó szignifikanciaszinteket ( $p$  értékeket) a következő táblázat mutatja. Ennek alapján elmondható, hogy egyedül a repetitio, azaz az ismétlés mutat statisztikailag szignifikáns különbséget a két csoport között.

|                   | <i>t</i> -próba az 1-es szövegen | <i>t</i> -próba a 2-es szövegen |
|-------------------|----------------------------------|---------------------------------|
| <b>Adverbial</b>  | 0,4837                           | 0,0688                          |
| <b>Apposition</b> | 0,2985                           | 0,2102                          |
| <b>Epithet</b>    | 0,1565                           | 0,2116                          |
| <b>Hypernym</b>   | 0,3560                           | 0,3407                          |
| <b>Meronym</b>    | 0,2714                           | 0,0925                          |
| <b>Other</b>      | 0,3432                           | 0,1771                          |
| <b>Pronominal</b> | 0,3847                           | 0,4949                          |
| <b>Repetition</b> | 0,0135                           | 0,0234                          |
| <b>Synonym</b>    | 0,4549                           | 0,3581                          |
| <b>Verbal</b>     | 0,1565                           | 0,3945                          |

3. táblázat. A kategóriák szignifikanciaszintjei

### 3.4 Egy gépi tanulási kísérlet

A már említett, a Szegedi Tudományegyetemen zajló projektet kiegészítve gépi tanulási kísérleteket is végeztem. Az általuk alkotott teljes jellemzőkészletet felhasználva a legeredményesebb módszer számukra az SVM algoritmus (Cortes–Vapnik 1995) volt, amellyel 69,1%-os pontosságot értek el. Ez azt jelenti, hogy 84 esetből 58-szor állapított meg a rendszer pontos diagnózist (Vincze et al. 2016).

Az általuk használt teljes jellemzőkészletet kiegészítettem az annotált visszautalási kategóriákkal. Minden beszélő esetében hozzáadtam jellemzőként az általa használt egyes visszautalási típusok számát, az egyes szövegekre lebontva. Így egy nagyobb mintát felhasználva az SVM algoritmus 75,58%-os pontosságot ért el. A mérésben a 86 adatközlő eredményei szerepeltek. Az összes esetből 65-ször állapított meg helyes diagnózist a rendszer. A 21 hibás diagnózisból 11-szer fals pozitívat, 10-szer fals negatívát állapított meg, ami azt jelenti, hogy a rendszer 11 személyt tévesen enyhe kognitív zavarban szenvedőnek, 10 személyt pedig tévesen egészségesnek minősített.

## 4 Következtetések

Az általam végzett vizsgálat alapján elmondható, hogy az enyhe kognitív zavarban szenvedők az egészségesektől részben eltérő stratégiát alkalmaznak a visszautalásra ugyanazon történet elmesélése során. Az elsődleges hipotézisem az volt, hogy az EKZ-sok kevesebb és rövidebb koreferencialáncot fognak használni: ez csak részben

igazolódott. Az EKZ-s adatközlők valóban kevesebb koreferencialáncot használtak átlagosan, azonban a hosszaik között nem mutatkozott különbség.

A második hipotézisem az volt, hogy az EKZ-s csoportban az ismétlés használata a legjellemzőbb, a kontrollcsoportban pedig a szinonímia és egyéb visszautalási formák. A vizsgálat alapján ez a hipotézis sem helytálló, mind a két csoportban az ismétlés volt a leggyakoribb visszautalási forma. Ennek egyszerű magyarázata lehet, hogy az elménkben emléknymokként tárolt információ könnyebben hívható elő, minél többször ismétlődik egy szó, annál erősebb és frissebb az emléknym. Ezért lehetséges az, hogy a leggyakoribb kategória a kontroll- és az EKZ-csoportban is az ismétlés volt. Emellett a kontrollcsoportban közel 8%-kal több lett az ismétlés használata, mint az EKZ-s csoportban. Ennek oka lehet az, hogy mivel a kimondott szó a fonológiai hurokban tárolódik (Gósy 2005), így egy ép emlékezeti rendszerrel rendelkező egyénnek könnyebb onnan előhívni a már egyszer kimondott szavakat, míg egy sérült emlékezeti rendszerrel rendelkező egyénnek ugyanakkora erőfeszítés az ismétlés használata, mint más visszautalási formáé.

A hipotézisek ugyan nem igazolódtak, de a gépi tanítás alapján érdemes a koreferenciára vonatkozóan további kutatásokat végezni a mentális betegségek kapcsán. Mivel az annotált kategóriák közül nem mind mutatott szignifikáns különbséget a két csoport között, ezért gazdaságosabbá lehetne tenni a jellemzőkészletet egyes kategóriák összevonásával. További kategóriák felvételével szintén árnyaltabbá lehetne tenni a különbséget. Mivel a névmási kategóriát elég nagy arányban használta mind a két csoport, érdemes lenne megvizsgálni azt, hogy ezen a kategórián belül mekkora az előfordulása az egyes névmási visszautalásoknak.

Mivel a gépi tanítás során az olyan kategóriák javítottak jelentősen az elért eredményen, amelyek egyébként nem mutattak szignifikáns különbséget használati arányaik szempontjából, a minta növelése is indokolt lenne.

## 5 Összegzés

Kutatásom során az enyhe kognitív zavarban szenvedők és egészséges adatközlők által elmesélt rövid történetekben fellelhető koreferencialáncokat hasonlítottam össze. A két csoport között eltérést mutatott a koreferencialáncok száma, az ismétlés mint visszautalás használatának aránya és az egyes visszautalási kategóriák csoporton belüli eloszlása. Nem mutat különbséget a használt koreferencialáncok hossza és a többi visszautalási stratégia használati aránya a két csoport között. A kapott eredményeim javítják a különböző automatikus felismerésre törekvő eljárásokat.

## Irodalom

- Baddeley, A., Eysenck M. W., Anderson M. C. 2010. *Emlékezet*. Budapest: Akadémiai Kiadó.
- Cortes, C., Vapnik, V. 1995. Supportvector networks. *Machine Learning*, 20(3): 273–297.
- Gósy, M. 2005. *Pszicholingvisztika*. Budapest: Osiris Kiadó.
- Hall, M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H. 2009. The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1): 10–18.
- Kempler, D. 2005. *Neurocognitive disorders in aging*. Thousand Oaks: Sage Publications.
- Tolcsvai Nagy G. 2003. A szöveg. In: Kiefer, F. (szerk.) *A magyar nyelv kézikönyve*. Budapest: Akadémiai Kiadó. 245–256.
- Negash, S., Petersen L. E., Geda Y. E., Knopman D. S., Boeve B. F., Smith G. E., Ivnik R. J., Howard D. V., Howard Jr J. H., Petersen R. C. 2007. Effects of ApoE genotype and Mild Cognitive Impairment on implicit learning. *Neurobiology of Aging*, 28(6): 885–893.
- Tariska, P. (szerk.) 2002. *Kortünet vagy kórtünet? Mentális zavarok az időskorban*. Budapest: Medicina Könyvkiadó.
- Turi, Zs., Németh D., Hoffmann I. 2014. Nyelv és emlékezet. In: Pléh, Cs., Lukács, Á. (szerk.) *Pszicholingvisztika: Magyar pszicholingvisztikai kézikönyv*. 2. kötet. Budapest: Akadémiai Kiadó. 743–776.
- Vincze, V., Hoffmann I., Sztalóczki G., Biró E., Gosztolya G., Tóth L., Pákáski M., Kálmán J. 2015. Az enyhe kognitív zavar automatikus azonosítása beszédátiratok alapján. In: Tanács, A., Varga V., Vincze V. (szerk.) *XI. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY2015*. Szeged: JATEPress. 249–256.
- Vincze, V., Gosztolya G., Tóth L., Hoffmann I., Sztalóczki G., Bánréti Z., Pákáski M., Kálmán J. 2016. Detecting mild cognitive impairment by exploiting linguistic information from transcripts. In: Katrin Erk, Noah A. Smith (szerk.) *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Berlin, 2016. augusztus 7–12.)*. Berlin: Association for Computational Linguistics. 181–187.
- University of Washington, Psychology Writing Center. 2010. *Reporting Results of Common Statistical Tests in APA Format*. Elérhető: [https://depts.washington.edu/psych/files/writing\\_center/stats.pdf](https://depts.washington.edu/psych/files/writing_center/stats.pdf). Letöltve: 2017. október 6.

## Cigány identitáskonstrukciók heterogenitása egyetemisták nyelvhasználatában

Kresztyankó Annamária

ELTE BTK Nyelvtudományi Doktori Iskola  
annamari915@gmail.com

**Kivonat:** A Budapesti Egyetemi Kollégiumi Korpusz (BEKK) létrehozásával és folyamatos bővítésével számos empirikus vizsgálat válik lehetővé, mivel a korpusz nemcsak a Budapesten élő fiatalok egymás között folytatott spontán nyelvi interakcióiból, hanem a felvett spontán társalgásokat követően rögzített szociolingvisztikai interjúkból épül fel, így sokféle – pl. társalgáselemzési és korpusznyelvészeti – elemzési szempont érvényesítését teszi lehetővé. A társalgáselemzést metodológiai kiindulópontnak tekintve a kutatás során több típusú módszert alkalmazva vizsgálom a különböző társadalmi kategóriák, jelen esetben a *cigány* mint társadalmi kategória konstrukcióit.

### 1 Bevezetés

A rendszerváltást követően a magyarországi cigányságot érintően számos sztereotípiakutatás és attitűdvizsgálat született, melyek alapvetően szociológiai vagy antropológiai szemléletűek (vö. Csepeli 2014; Kende 2013; Neményi 1998; Postma 1996). Írásom célja annak bemutatása, hogy az egyetemisták körében készített interjúkban artikulálódó etnicitáskonstrukciók mennyiben hasonlítanak a cigányokkal kapcsolatban a rendszerváltás idején kimutatott sztereotípiákra és attitűdökre (vö. Postma 1996), valamint hogy a BEKK interjúi milyen módon kapcsolódnak a sztereotípiakutatások eredményeihez. Feltételezésem szerint a negatív sztereotípiák kevésbé artikulálódnak explicit módon az egyetemisták körében, mint az 1990-es évek elején.

### 2 A Budapesti Egyetemi Kollégiumi Korpusz rövid bemutatása

A 2015-ben megalakult Budapesti Egyetemi Kollégiumi Korpusz kutatócsoportja kettős célt tűzött ki maga elé. Egyrészt elsőként létrehozni a magyar nyelvű hétköznapi társalgások adatbázisát, amely a Budapesten élő fiatal egyetemisták egymás között folytatott nyelvi interakcióiból épül fel, és sokféle korpusznyelvészeti elemzési szempont érvényesítését teszi lehetővé. Másrészt a korpuszt a társadalmi nem, a szexualitás és az etnicitás konstrukciói mentén vizsgáljuk a társalgáselemzés keretében.

A kutatás célja annak feltérképezése, hogy miként jönne létre, és miként változnak társadalmi kategóriák a mindennapi társalgások során. A kollégiumi szobákban rögzített társalgásokat követően a résztvevőkkel irányított szociolingvisztikai interjúkat

veszünk fel az etnicitást, a nemzet és a szexualitást középpontba állító modulok feldolgozásával, s e társadalmi kategóriák nyelvi konstrukcióit kvalitatív tartalomelemzéssel vizsgáljuk a Bihalbocs szoftver (Vékás 2007) felhasználásával. Társas-konstruktivista szemléletű kutatásunk során emellett etnográfiai megfigyeléseket is végzünk, hiszen a beszéd etnográfiája is kutatási szempont, ám annak ellenére, hogy a BEKK nem sztereotípiakutatás, a felvételek elemzését követően az etnicitás, a sztereotípiák és az imitáció kapcsolatának vizsgálata is megfigyelési szemponttá vált.

### **3 A kutatás céljai**

Tanulmányomban a kutatásban vizsgált kategóriák közül a cigány identitáskonstrukciókkal foglalkozom. Kiinduló kérdéseim a következők:

- 1) Milyen stratégiák mentén hozzák létre a cigányságot mint társadalmi kategóriát a vizsgálatban résztvevő egyetemisták?
- 2) Milyen negatív és/vagy pozitív sztereotípiatartalmakkal, belső és/vagy külső tulajdonságokkal és konnotációkkal hozzák összefüggésbe a cigányság kategóriáját a fiatalok?

E kérdések megválaszolásával képet kapunk a bevezetőben leírt célkitűzésekről is, miszerint analógiás kapcsolatot mutathatunk ki a rendszerváltás- és a jelenkori etnicitáskonstrukciók között.

### **4 Módszertani kérdések**

Írásomban a BEKK kiscsoportokban (3–5 fővel) végzett interjúanyagát dolgozom fel, melyben ún. modulokat alakítottunk ki, amelyek az egyes interjúk összehasonlíthatóságát biztosítják, valamint elősegítik eltérő beszédstílusok vizsgálatát. E modulokból az etnicitásra/cigányságra vonatkozó részt vizsgálom meg. Ezek ugyan irányított beszédhelyzetek, de az itt létrejövő diskurzusok vélhetően olyan társadalmi gyakorlatra utalnak, amelyek meghatározzák az interjúalanyok nyelvi konstrukcióit – és gondolkodásmódját. Elemzésem során 10 interjút, közel 50 kollégista etnicitáskonstrukcióit vizsgáltam meg.

Az irányított interjú első lépéseként a résztvevők egy cikkrészletet olvashattak el, amely a magyarországi cigánygyilkosságok tárgyalásáról tudósított. Ezt követően a cigányság társadalmi megítéléséről, valamint a saját cigánydefiníciójuk kifejtése után annak jellemzőiről kellett véleményt formálniuk a résztvevő egyetemistáknak.

Az interjúk feldolgozása két szakaszban zajlott: elsőként kvalitatív tartalomelemzést végeztem, majd a fellelhető jelenségek összefüggéseire fókuszáltam.

### **5 Etnicitásdefiníció(k), etnikai és kisebbségi sztereotípiák**

A kutatás során az etnicitásra – az antropológiai és a szociológiai diskurzusok „többszólamú” etnicitás- és identitásmeghatározásait alapul véve – mint előre nem adott tényezőre tekintek. Andorka Rudolf definíciója mentén „etnikai csoportnak

nevezük a társadalmon belül azoknak a csoportját, akik olyan közös kulturális identitás tudatával rendelkeznek, amely elkülöníti őket a többségtől vagy a többi etnikai csoporttól” (Andorka 2003: 336). Emellett Andorkához hasonlóan mindvégig hangsúlyozom azt is, hogy az egyén identitásválasztása a legfontosabb tényező, és kiemelem annak a lehetőségét is, hogy ugyanaz az egyén több identitást is választhat önmaga számára. Továbbá az etnocentrikus ideológiai tradíción (vö. Andorka 2003; Csepeli 2014) továbblépve úgy vélem, hogy a kulturális különbségek társadalmilag és alkalomhoz kötötten szerveződnek, tehát kiindulópontom szerint a társalgások résztvevői az etnicitás kategóriáját és jellemzőit közösen hozzák létre, így az etnicitás egy állandóan változó, dinamikus jelenség (vö. Barth 1996; Brubaker 2004; Eriksen 1993, 2000). E gondolat a nemzetközi szociolingvisztikai kutatásokban már az 1980-as évektől kezdve jelen van, hiszen a társadalmi struktúrák, gyakorlatok és a társas jelentés közös konstruálásának kutatása során már nem az eleve adottnak vélt társadalmi tényezők, hanem a nyelvhasználók által létrehozott és az általuk láttatott identitások és társadalmi struktúrák nyelvhasználati vizsgálatai domináltak a társas-konstruktivista megközelítésmód szellemében (vö. Bartha–Hámori 2010: 302).

Az etnikai és kisebbségi sztereotípiákat a különböző csoportok tagjainak pozitív és negatív tulajdonságaira vonatkozó nézetek alapján vizsgálják a szociológusok. Magyarországon az ilyen típusba tartozó attitűdvizsgálatok a rendszerváltást követően kezdődtek meg, ám e kutatások nyelvi kérdésekkel csupán érintőlegesen foglalkoztak, míg a BEKK épp az attitűdök nyelvi és diskurzív vonatkozásaira fókuszál.

A vizsgálat során attitűdnek tekintek minden tapasztalat révén szerveződött mentális és idegi készenléti állapotot, amely irányító vagy dinamikus hatást gyakorol az egyén reagálására mindazon helyzetekben, melyekre az attitűd vonatkozik (vö. Allport 1999). Kutatásunkban a személyes előítélet a nyilvános attitűddel együtt válik vizsgálhatóvá.

A szociológiai szakirodalom szerint a sztereotípiák általános érvényűnek tartott, túlzáson alapuló, az ember sajátos észlelési képessége nyomán többféleképpen létrejövő leegyszerűsített képek, Allport szerint: a sztereotípiák kategóriával társult túlzáson alapuló nézet. Feladata a kategóriával kapcsolatos viselkedés racionalizálása (Allport 1999), ám a társas interakciókra is nagy hatással van, hiszen a sztereotíp sémák befolyásolják az észlelést, a gondolkodást és a viselkedést is (Binderffer 2006: 14). A sztereotípiakutatás funkcionális megközelítése szerint a sztereotípiák prototípusa egy kategória – jelen esetben etnikum – felruházása egy emberi tulajdonsággal (Hunyady–Nguyen 2001: 484). A továbbiakban e funkcionális irány mentén értelmezem a sztereotípiák fogalmát a csoportközi kapcsolatokban. A nyelvet, nyelvhasználatot érintő előítéletek a kutatásokban látens módon szerepelnek, ami funkcionalista társadalomelméleti megközelítésből értelmezve a manifeszt funkcióval egyenértékű elem a társadalmi előítéletek vizsgálata és értelmezése során (Merton 2000).

### 5.1 Sztereotípiák és nyelvi előítéletek a '90-es években és a 21. században

Az 1980-as évek végén és az 1990-es évek elején Koos Postma gyűjtött cigányokkal kapcsolatos vélekedéseket Magyarországon. 1987-ben, 1992-ben és 1993-ban elvégzett kvantitatív kutatása eldöntendő kérdésekre adott válaszai alapján arra a következtetésre jutott, hogy negatív előítéletesség jellemezte a rendszerváltás idején a magyar társadalom a cigányokkal szemben. A kutatásban részt vevőknek 16 pozitív, valamint 16 negatív tulajdonságról kellett véleményt formálniuk aszerint, azok vajon jellemzőek-e a cigányokra. A negatívumok között a megkérdezettek körében a

legnagyobb gyakorisággal az agresszivitás (70%), a lustaság (68%), a műveletlenség (61%), a megbízhatatlanság (60%), az élősködés (59%) és a piszkosság (56%) szerepelt (vö. Huszár–Géczi–Sramó 2003; Postma 1996; Prónai 2004). Egy nyelvhasználatot vizsgáló kutatás szempontjából is értékes, felhasználható adatok ezek, hiszen a felsorolt tulajdonságok közül az agresszivitás, valamint a műveletlenség a nyelvhasználatra is kiterjedő fogalmak (vö. az ÉKsz. agresszivitás szócikkével: 'erőszakos, támadó, sértő hang', ami Michael Stewart 1987-es észak-magyarországi kutatásával is összekapcsolható, melyben a hangos és céltalan zajongó jelzőkkel illetve a cigány közösség tagjainak nyelvhasználatát, lásd Stewart 1987).

Egy évtizeddel ezen eredmények publikálását követően észak-magyarországi 10–18 évesek körében végeztek hasonló kutatást, itt Murányi István a cigányok és a magyarok attribúciós megítélését vizsgálta: a csoporton belüli összetartás jelent meg mint leggyakoribb tulajdonság, ezt az erőszakosság, egészségtelenség, megbízhatatlanság, valamint a műveletlenség és a piszkosság követte a cigányokra vonatkozó jellemzők sorában (Murányi 2006).

Bár a nyelvet, nyelvhasználatot érintő sztereotípiák e kutatásokban csak látens módon szerepelnek, az előítéletek a nyelvre is vonatkoznak, és az imitációk egy regiszterképző-folyamatba épülnek be. Susan Gal szerint pedig a regiszterképzés (*enregistment*) mint metafogalom háromféle elképzelést köt össze: elképzelést a beszélőkről, az értékekről és a nyelvi variánsokról (Gal 2016: 14). Érdemes lehet e megállapítást akár a médiában megjelenő (pl. talk-show, valóságshow) magyarul beszélő cigányok beszédmódjával is összefüggésbe hozni (vö. Császi 2011, Hegedűs 2009).

## 6 Eredmények

Tanulmányom a cigány identitáskonstrukciók heterogenitását helyezi előtérbe. Heterogenitás azonban az interjúelemzést elvégezve nem mutatható ki, hiszen ezen identitáskonstrukciók megegyeznek a rendszerváltás idején kimutatott sztereotípiákkal, az egyetemisták a cigányságot mint társadalmi kategóriát a Polyanna-elv maximájára (Leech 1983) látenssen hivatkozva hozzák létre. Ezen udvariassági – pragmatikai kompetenciára épülő – elv szerint a társalgások során el kell kerülni a kellemetlen beszédtemákat, így a kommunikáció nemcsak téma, hanem kontextus és kulturális szokások alapján is vizsgálhatóvá válik (vö. Szili 2007). Ám az interjúk egy pontján mégis negatív sztereotípiatartalmakkal és belső és/vagy külső tulajdonságokkal, valamint a nyelvhasználat szupraszegmentális „mátságával” hozzák összefüggésbe a cigány kategóriát a megkérdezett egyetemisták, tehát Leech elve az interjúk egy pontján már nem érvényesül.

### 6.1 Mi vezérelte az elkövetőket? – avagy etnicitáskonstrukciók és kisebbségi sztereotípiák egy jogi kérdésen keresztül

A 2008–2009-ben történt cigányok ellen elkövetett támadások tárgyalóperéről tudósító szöveg elolvasása után az elkövetők lehetséges indítékairól kérdeztük az egyetemistákat. A válaszok három kategóriába sorolhatók, ezek gyakoriságuk szerint a következők voltak: (1) random, véletlenszerű támadások történtek; (2) a cigány

emberek életvitele volt az indíték; s végül legnagyobb arányban előforduló válasz (3) a romabűnözés visszaszorítása volt az elkövetők célja.

A következőkben olyan interjúrészleteket mutatok be, amelyek a feltett jogi kérdésen túl már saját etnicitáskonstrukciókkal is kiegészülnek.

[1] A: Eléggé [...] érdekes életmódot folytatnak a romák, nem az összes, de a nagy részük. Meglincselgetnek autókat, akik bemennek a faluba.

Az [1] példán keresztül jól látható, hogy az „érdekes életmód” megjelölése az etnicitáskonstrukció alapja, melynek negatív mivolta a lincselés említése után válik egyértelművé. A példa az áldozathibáztatáson túl azt is megmutathatja, hogy miként működik a funkcionális sematizálás gyakorlata, hiszen egyetlen kategóriát érvényesít abban az esetben is, amikor a bűncselekmény áldozatai cigány emberek.

A [2] példa ugyancsak az életmódon keresztül hozza létre a cigányság kategóriáját úgy, hogy azt jól elválaszthatóként, mintegy evidenciaként értelmezi más csoportokhoz – jelen esetben a magyar csoporthoz – képest:

[2] Az életvitelük eltér a magyarok szokásaitól.

A cigányok ellen elkövetett gyilkosságok leggyakrabban említett indítéka azonban nem az életmód, hanem a bűnözés, pontosabban a cigánybűnözés jelensége volt. Az etnicitáskonstrukciók létrehozása során ez így jelent meg:

[3] ... válasz Olaszliszkára. Magyarországon tíz bűncselekményből nyolcnál cigány származású az elkövető. Vagy legyen mondjuk hét.

[4] Azt mindannyian tudjuk, hogy a roma származásúakból gyakrabban kerülnek ki bűnözők. [...] De ez embertelen, általános közvélekedés.

[5] „Megunták, hogy a cigány etnikum egy része hogyan viselkedik.

[6] Az tény, hogy kriminális állapotok vannak ott, azok a gyerekek nem gyerekként viselkednek. A helyzet folyamatosan romlik, egyre többen vannak, nem csodálkozom azon, hogy vannak rasszista hajlamú emberek.

A példákon keresztül látható, hogy a megkérdezettek akár számadatokkal, meggyőzésre alkalmas formulákkal („azt mindannyian tudjuk”, „az tény, hogy”) is élnek, ha a kisebbségi sztereotípiák igazolására kerül sor. A [4]-es szövegrészlet is így épül fel, ám az interjú egy későbbi pontján visszatér a megkérdezett e kérdésre, és önmagára reflektál, miszerint vélekedése „embertelen, általános” volt, de mégsem vonja azt kétségbe.

## 6.2 Szerinted ki számít cigánynak? – avagy az etnonimák megítélése

Az interjú második része a „*Ki a cigány?*” (Ladányi–Szelényi 2000) kérdésre kereste a választ, majd az etnonimák használatának tudatosságát mérte fel.

A megkérdezett egyetemisták élesen elválasztható kategóriák mentén értelmezték a „Ki a cigány?” kérdést, e kategóriák interjúrészletekkel illusztrálva a következők voltak:

[7] G: Az, aki a cigánysoron születik.



Ez a válaszcsoport a szociális helyzetet helyezte középpontba, amikor a cigány csoport(ok) egészét kellett meghatározniuk a szegénység, a putriban lakás és az iskolázatlanság tényezőinek diskurzusba emelésén keresztül.

A második típusba azok a válaszadók tartoznak, akik a szociológiai felmérések szerinti cigány-meghatározást követték, tehát akik szerint a többségi társadalom véleménye és magatartása számítanak ebben a kérdésben (vö. Havas 1999:23):

[8] H: Cigány az, akit a környezet annak tart.

Végül arra is található példa, akik az önmeghatározást tartják mérvadónak ebben a kérdésben, ami a (kulturális) antropológiai vizsgálatok módszertanát követi:

[9] I: Aki magát annak tartja.

Nyelvet vagy nyelvhasználatot érintő válasz – előzetes elvárásommal ellentétben – egyetlen esetben sem született. A *cigány* és a *roma* meghatározások használata között azonban vannak különbségek a megkérdezett egyetemisták szerint, ám e vélekedések esetenként akár ellent is mondhatnak egymásnak. A *cigány* és *roma* megnevezés a következő esetekben legitim a válaszok szerint:

[10] J: A cigány egy pejoratív, negatív jelző.

[11] ... viselkedésük alapján. A cigány negatív jelző: hát szerintem, nagyjából ha megnézed a híradóban meg itt, ott, amott a különböző oldalakon így hozzák fel, hogy cigány, aki egy negatívabb személyiség és a roma, aki [...] aki meg azér' azt mondom, hogy kevésbé vagy úgymond próbál kitörni ebből a cigányos dologból. Meg hogy ha ilyen hivatalosabb dolgokban halljuk, akkor roma kisebbségről beszélnek, nem mondják cigánynak.

[12] ... vannak a tisztességesebb roma származású családok. És 200 kilósak és –20 fokban strandpapucsban ülnek kinn a főtéren, és kiabálnak egymásnak, hogy amúgy így dikk, hozd már ide a nemtom micsodát: nálunk ez a két kategória van. Azokat hívjuk cigányoknak, akik ilyen alpáribb módon viselkednek, azokat meg akik így integrálódnak a társadalomba, azokat romának. De van, aki sértésnek veszi (mert hogy a roma eredetileg egy csoport a cigányságon belül).

[13] Ezt használom (cigány), nem tartom sértőnek, mert nem annak szánom.

[14] A roma kicsit kedvesebb megnevezése a cigánynak.

[15] A cigányt arra is mondom, aki nem cigány.

Az [15]-ös interjúrészlet túlmutat az etnicitáson és ennek nyelvi leképeződésén, hiszen egy jellemmel, viselkedéssel, tehát más attribútumokkal kapcsolja össze a *cigány* megnevezés használatát, így ez már ikonizációként (Irvine–Gal 2000) is jelen lehet a fiatalok nyelvhasználatában.

### 6.3 A nyelvhasználat kapcsolata az etnicitással

Az interjú befejezésekként külső és belső tulajdonságok mentén kellett a cigányságot meghatározni. Ennél a résznél volt a leginkább megfigyelhető, hogy kezdetben a „Nem tudom” válasz jelent meg (vö. az udvariasság társalgási maximájával, Leech 1983; Szili

2007), ám ezt követően e jellemzők fordultak elő, már nyelvhasználati kérdéseket is megemlítve:

[16]P: Külsőre meg igenis látszik rajta, merthát az igenis látszik.

Külső tulajdonságok közt említették a bőrszín, a szemszín, az arcszerkezet és az öltözködés „*jellemző más milyenségét*”. Ezt követően a már bemutatott sztereotípiavizsgálatok eredményeivel nagy hasonlóságot mutató belső tulajdonságok következtek. Az erkölcstelenség (lásd: [17]-es párbeszéd), az erőszakosság, a más életmód, a nemtörődömség, a más járás, a vagánykodás, a forróvérűség, a hirtelenség, az érzékenység, valamint az összetartás jellemzői mentén alakították ki közösen a cigányság csoportját úgy, hogy az elemzett interjúk során nem található példa az etnikai kategória mentén létrehozott különbségtétel megtagadására.

[17]

- ... teljes erkölcstelenség.
- Én inkább úgy mondanám, hogy nagyon más munkamorál. <nevet>
- Nem létező munkamorál.
- Igen, tehát ilyen szabadabb felfogás mindenféle dologról.
- Igen, az egy dolog, de amikor már zavarják a környezetüket is, na.”

A nyelvhasználati különbségek ennél a kérdésnél artikulálódtak elsőként az interjú során. Nem véletlenül használom a különbségek kifejezést, ugyanis magyar anyanyelvüként jelennek meg a cigányok az egyetemisták konstrukcióiban, azonban akcentussal, másként beszélő magyar anyanyelvűekként:

[18]R: ... nyelvezet is egyébként más. Egyfajta akcentus is hallatszik a beszédén.

[19]S: ... de inkább a beszédük, ami elárulja őket. Hát az akcentus.

Az interjúk során cigány nyelvekről, más anyanyelvről vagy kétnyelvűségről egy alkalommal sem esett szó.

## 7 Következtetések

Tanulmányom cigány identitáskonstrukciók heterogenitásának bemutatását helyezte előtérbe egyetemi gyakorlati közösségek (Wenger 1998; Bucholtz 2001) nyelvhasználatát megvizsgálva. Heterogenitás azonban az interjúelemzést elvégezve nem mutatható ki oly módon, amilyenre számítottam a kutatás kezdetén. Eredményeim azt mutatják, hogy az egyetemisták a cigányságot mint társadalmi kategóriát homogenizált csoportként konstruálják, mégpedig olyan – akár személyes tapasztalatok nélkül is kialakított – sztereotípiák mentén, amelyek a rendszerváltás idején is jelen voltak a társadalomban. A bemutatott szövegrészek és példák alapján elmondható, hogy negatív sztereotípiatartalmak és belső és/vagy külső tulajdonságok a jellemzők, valamint a nyelvhasználat szupraszegmentális „mássága” az etnicitáskonstrukció alapja, s ezek explicit módon vannak jelen e közösségekben.

A BEKK szociolingvisztikai interjúi a bemutatott etnicitáskonstrukciók mentén rávilágítanak arra, hogy a budapesti egyetemisták társalgásaikban hogyan konstruálják meg a *cigány* kategóriáját, valamint arra is, hogy e kategória hogyan válik a társadalmi különbségtétel tárgyává egyetemi közösségekben. A kutatás eredményei számtalan

területen alkalmazhatók a gyakorlatban, hiszen az előítéletes magatartásra vonatkozó érzékenyítés az oktatástól a munkaerőpiacon át az élet minden területén meghatározó. Az eredmények pedig azt mutatják, hogy ezen közösségek esetében is szükség lenne a társadalmi különbségek kialakulását és újratermelését értelmező beszélgetésekre.

## Irodalom

- Allport, G. 1999. *Az előítélet*. Budapest: Osiris Kiadó.
- Andorka, R. 2003. *Bevezetés a szociológiába*. Budapest: Osiris Kiadó.
- Barth, F. 1996. Régi és új problémák az etnicitás elemzésében. *Regio*, 7(1): 3–25.
- Bartha, Cs., Hámori, Á. 2010. Stílus a szociolingvisztikában, stílus a diskurzusbán. Nyelvi variabilitás és társas jelentések konstruálása a szociolingvisztika „harmadik hullámában”. *Magyar Nyelvőr*, 134: 298–321.
- Binderffer, Gy. 2006. Stereotipizáció az interetnikus kapcsolatokban. In: Bakó, B., Papp R., Szarka L. (szerk.) *Mindennapi előítéletek*. Budapest: Balassi Kiadó. 9–36.
- Brubaker, R. 2004. *Ethnicity without Groups*. Harvard University Press.
- Bucholtz, M. 2001. Minek legyek normális? Nyelvi és identitásgyakorlatok külön lányok közösségében. *Replika*, 45–46: 191–211.
- Császi, L. 2011. *A Mónika-jelenség kulturális szociológiája*. Budapest–Pécs: Gondolat Kiadó.
- Csepeli, Gy. 2014. *Szociálpszichológia mindenkiben*. Budapest: Kossuth Kiadó.
- Eriksen, Th. H. 1993. Ethnicity and Nationalism: Definitions and Critical Reflections. In: Lindholm, H. (szerk.) *Ethnicity and Nationalism*. Göteborg: Nordens & Padrigu. 40–50.
- Eriksen, Th. H. 2000. Ethnicity and Culture: A Second Look. In: Bendix, R., Roodenburg, H. (szerk.) *Managing Ethnicity. Perspectives from folklore studies, history and anthropology*. Amsterdam: Het Spinhuis. 185–205.
- Gal, S. 2016. Nyelvi sztenderdizáció: modellek és ideológiák. In: Kozmács, I., Vančo, I. (szerk.) *2016. Sztenderd – nem sztenderd. Variációk egy nyelv változataira*. Lakitelek: Antológia Kiadó.
- Havas, G. 1999. Cigányok a szociológiai kutatások tükrében. In: Glatz, K. (szerk.) *A cigányok Magyarországon*. Budapest: MTA.
- Hegedűs, A. 2008. „Nem hazudol, az őszintét mondom”. *Magyar Nyelv*, 105(1): 73–77.
- Hunyady, Gy., Nguyen, L. L. A. 2001. *Stereotípiakutatás. Hagyományok és irányok*. Budapest: ELTE Eötvös Kiadó.
- Huszár, Zs., Gécsi J. Sramó A. 2003. Kisebbségi attitűdök szocio-demográfiai háttere. *Iskolakultúra*, 4: 40–50.
- Irvine, J., Gal, S. 2000. Language ideology and linguistic differentiation. In: Kroskrity, P. (ed.) *Regimes of language: Ideologies, politics and identities*. Santa Fe, New Mexico: School of American Research Press. 35–83.
- Kende, Á. 2013. Normál gyerek, cigány gyerek. *Esély*, 2. Elérhető: [http://www.esely.org/kiadvanyok/2013\\_2/kende.pdf](http://www.esely.org/kiadvanyok/2013_2/kende.pdf). Letöltve: 2016. június 1.
- Ladányi, J. Szelényi, I. 2000. Még egyszer az etnikai besorolás „objektivitásáról”. In: Horváth, Á., Landau E. Szalai J. (szerk.) *Cigánynak születni*. Budapest: Aktív Társadalmi Alapítvány – Új Mandátum Könyvkiadó. 239–241.
- Leech, G. N. 1983. *Principes of pragmatics*. London – New York: Longman.
- Neményi, M. 1998. Cigány anyák az egészségügyben. Budapest: Nemzeti és Etnikai Kisebbségi Hivatal.
- Merton, R. K. 2000. Társadalmi struktúra és anomia. In: *Szociológiai irányzatok a XX. században*. Budapest: Új Mandátum Kiadó. 62–90.
- Murányi, I. 2006. *Identitás és előítélet*. Budapest: Új Mandátum Kiadó.
- Postma, K. 1996. *Changing Prejudice in Hungary. A Study of the Collapse of Socialism and its Impact on Prejudice Against Gypsies and Jews*. Groningen: Rijksuniversiteit ICS Dissertation Series.
- Prónai, Cs. 2004. A kulturális antropológia jelentősége a cigánykutatásban. In: Kovács, N., Osvát A., Szarka L. (szerk.) *Tér és terep. Tanulmányok az etnicitás és az identitás kérdésköréből*. Budapest: Akadémiai Kiadó. 252–271.
- Stewart, M. 1987. „Igaz beszéd” – avagy miért énekelnek az oláh cigányok? *Valóság*, 30(1): 49–64.

- Szili, K. 2007. Az udvariasság pragmatikája. *Magyar Nyelvőr*, 131: 1–17.
- Vékás, D. 2007. *Számítógépes dialektológia*. In: Guttman, M., Molnár, Z. (szerk.) *V. Dialektológiai Szimpozium* (Szombathely, 2007. augusztus 22–24.). Szombathely: Berzsényi Dániel Főiskola. 289–293.
- Wenger, E. 1998. *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.

## 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

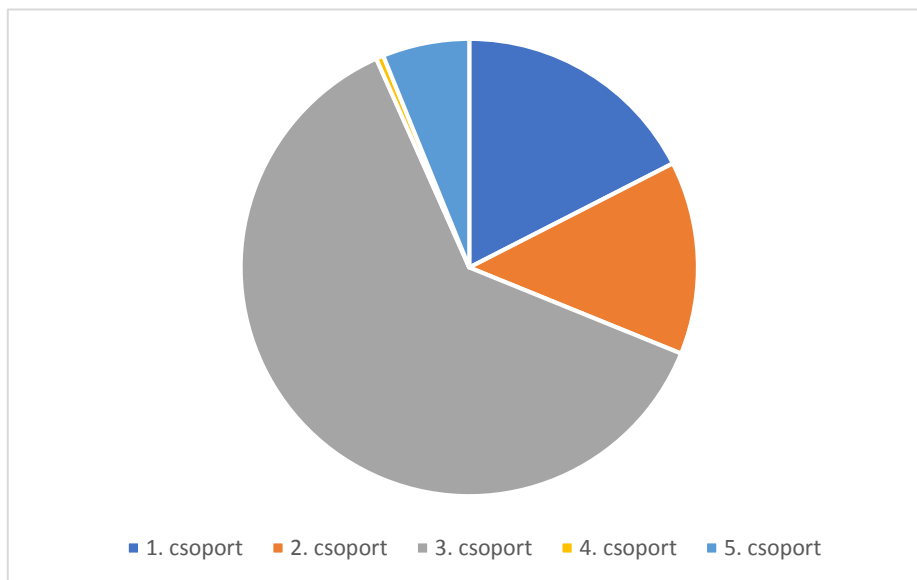
## 5 Az adatbázis

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: 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.

## Magyarul tanuló lengyel hallgatók köztesnyelvről különböző beszédaktusokban<sup>i</sup>

Sitkei Dóra

ELTE BTK Nyelvtudományi Doktori Iskola  
sitkeidora@gmail.com

**Kivonat:** Második nyelv (L2) tanulása során több tényező hathat a nyelvtanuló köztesnyelvére, amelyek egyike az anyanyelvi (L1) transzfer. Jelen kutatás során lengyel L1 hallgatók magyar köztesnyelvi beszédaktus-produkcióiban megnyilvánuló negatív anyanyelvi transzfer vizsgálom egy köztesnyelvi udvariassági vizsgálatokhoz megfelelő módszer megtalálása céljából, a következő udvariassági tényezők alapján: arc(ulat)fenyegetés-arc(ulat)védelem, direktség-indirektség, a szolidaritás kifejezése, az operációs szempontok és a külső és belső módosítók használata. Emellett a köztesnyelvi produkciók közelebbi megismerése céljából megfigyeléseket teszek az udvariassággal átfedésben lévő nyelvi túlbiztosításról is. A vizsgálatokat a következő beszédaktusokban végeztem: az üdvözlés, az ételkínálás, a kérdés, a köszönetnyilvánítás, az elutasítás, az egyet nem értés (véleménykülönbség) kifejezése, a bocsánatkérés, a bókolás és a bókolásra reagálás, a kívánság és a javaslat kifejezése, valamint az ugratás.

### 1 Bevezetés<sup>ii</sup>

A nyelvtanulás kreatív folyamat, melynek során a nyelvtanulók a környezetükkel való interaktív kapcsolat révén létrehozzák a megfigyelt szabályosságok egyfajta belső reprezentációját, amely dinamikusan változik a nyelvtanulási folyamat alatt (Corder 1992). Ezen nyelvi rendszer megnevezésére a köztesnyelv (interlanguage, rövidítve: IL) fogalmát Selinker vezette be 1972-ben. A nyelvtanuló köztesnyelvi megnyilatkozására több tényező van hatással, amelyek egyike az anyanyelvi (L1) transzfer (Selinker 1972).

Jelen kutatás fő célja doktori kutatásomhoz egy – köztesnyelvi udvariassági vizsgálatok számára megfelelő – kutatási módszer megtalálása, amelyhez a köztesnyelvi beszédaktus-produkciók tipikus megnyilatkozásaiban megfigyelhető negatív anyanyelvi transzfer vizsgálom a pragmatikai kutatások általános keretein belül (l. például Szili 2004, Bándli 2015) a következőkben (1.2. pont) bemutatandó udvariassági tényezők alapján.

<sup>i</sup> A tanulmány Az Emberi Erőforrások Minisztériuma Új Nemzeti Kiválóság Programjának támogatásával készült.

<sup>ii</sup> Ezúton is szeretném megköszönni mindazok segítségét, akik valamely módon hozzájárultak a kutatásom megvalósításához.

Mindemellett a köztesnyelvi produkció közelebbi megismerése céljából megfigyeléseket teszek az udvariassággal átfedésben lévő nyelvi túlbiztosításról is. A vizsgálatokat a következő beszédaktusokban végeztem: az üdvözlés, az ételkínálás, a kérés, a köszönetnyilvánítás, az elutasítás, az egyet nem értés (véleménykülönbség) kifejezése, a bocsánatkérés, a bókolás és a bókolásra reagálás, a kívánság és a javaslat kifejezése, valamint az ugratás beszédaktusa.

## 2 Módszertan

Nyílt végű diskurzuskiegészítő tesztre<sup>iii</sup> 80 lengyel anyanyelvű, magyarul tanuló és legalább középfokon tudó egyetemistától kaptam magyar nyelvű válaszokat. A kérdőíven a szituációkat magyarul és zárójelben angolul adtam meg. A szituációk leírása tartalmazza azokat a szociopragmatikai (vertikális/hatalmi és horizontális/familiáris interperszonális) és beszédaktus-specifikus (például a köszönetnyilvánítás beszédaktusában a kapott szívesség mértéke) tényezőket, amelyeknek hatása lehet a megnyilatkozásokra. A válaszokat összevettem 40 magyar anyanyelvű egyetemista magyar válaszával, ami alapján beazonosítottam a lengyel válaszadók – adott esetben – eltérő tipikus válaszstratégiáit. Lengyel anyanyelvű hallgatóktól (30 fő) kapott lengyel nyelvű válaszok alapján (a kérdőív is lengyelül íródott) megfigyeltem, hogy a negatív anyanyelvi transzfer<sup>iv</sup> mennyiben, mely nyelvi szinten (lexikai, szintaktikai vagy pragmatikai) magyaríthatja a magyartól eltérő válaszstratégiákat. A kutatás során tipikusnak tekintettem a legalább 25%-os relatív gyakorisággal megjelenő nyelvi formákat.

A következő udvariassági tényezők alapján vettem össze a lengyel és a magyar anyanyelvű személyek magyar nyelvű tipikus megnyilatkozásait:

- 1) arc(ulat)fenyegetés-arc(ulat)védelem (Brown–Levinson 1978, 1987; Szili 2004),
- 2) direktség-indirektség (Brown–Levinson 1987; Blum-Kulka 1987; Szili 2004; Bándli 2015),
- 3) szolidaritás kifejezése (Brown–Levinson 1987; Jobst 2007; Brown 2015),
- 4) operációs szempontok (point-of-view operations)<sup>v</sup> (Brown–Levinson 1978; Blum-Kulka–Olshtain 1984; Brown–Levinson 1987),

<sup>iii</sup> A diskurzuskiegészítő teszt széles körben használatos, de vannak hiányosságai. Beebe–Cummings (1996) vizsgálatában bebizonyította, hogy elutasítás során az élnyelvi változatokhoz hasonló diskurzuskiegészítő tesztbeli válaszokat adtak az adatközlők az alkalmazott szemantikai formák tekintetében, viszont a módszer negatív oldalaként említik a pszichológiai tényezőt: nem biztos, hogy éles szituációkban is ezen stratégiákat alkalmazzák az adatközlők, a teszt nem tudja hozni az interakciónak a beszélgetésben részt vevők közötti pszichoszociális dinamikáját (Beebe–Cummings 1996).

<sup>iv</sup> Azon köztesnyelvi jelenségeknél, amelyek az anyanyelvi formáknak megfelelnek, de a célnyelvi formáktól különböznek, fennáll a negatív anyanyelvi transzfer lehetősége, viszont ez nem elegendő bizonyíték rá (Beebe–Takahashi–Uliss–Weltz 1990).

<sup>v</sup> Az operációs szempont alapján beszélő-, hallgató-, kettős és személytelen orientáció különböztethető meg (Blum-Kulka–Olshtain 1984, Brown–Levinson 1987), amely beszélőorientált megnyilatkozás esetén a beszélőhöz való lehorgonyzást jelenti (I. Kugler 2013), és így tovább.

- 5) külső és belső módosítók (pl. enyhítő vagy támogató elemek) használata (Blum-Kulka–Olshtain 1984).

### 3 A köztesnyelvi vizsgálat

#### 3.1 A beszédaktusok, melyeknek tipikus produkcióiban nem volt eltérés

A következő beszédaktusok esetén nem volt eltérés a köztesnyelvi és a magyar L1 tipikus megnyilatkozások között a vizsgált udvariassági szempontok alapján: az üdvözlés, a köszönetnyilvánítás és az elutasítás beszédaktusában. A következőkben a három beszédaktus tipikus produkcióit mutatom be röviden.

##### 3.1.1 Az üdvözlés beszédaktusa

A szituáció leírása: a városban találkozol a magyar tanárral, aki (a) egy 35 éves nő, (b) egy 35 éves férfi, (c) az albérleted előtt az 50 éves szomszédasszonnyal, (d) az egyetemen reggel az osztálytársaddal, mit mondasz neki (mit köszönsz, mondasz-e mást is a köszönésen kívül).

A tipikus magyar nyelvű produkciók, mind a magyar L1, mind a köztesnyelvi beszélők részéről: (a, b, c) *Jó napot (kívánok)!* (d) *Szia! (Hogy vagy? / Mi újság?)*. A köztesnyelvi és a magyar L1 tipikus válaszok megegyeztek. Sok esetben mind a lengyel, mind a magyar adatközlők zárójelben megadták a bővített válasz lehetőségét, amelyhez a magyar válaszolók közül néhányan magyarázatot is fűztek.

##### 3.1.2 A köszönetnyilvánítás beszédaktusa

A szituáció leírása: (a) nem tudtál menni egy órára az egyetemen, egy barátod felajánlja, hogy odaadja a jegyzeteit, (b) elkérted a húgodtól egy DVD-jét, és ő odaadta neked, (c) egy tanárod kölcsönad neked egy könyvet, amit nem tudtál könyvtárból kikölcsönözni.

A köztesnyelvi válaszokban is megjelenik az (a) és a (c) szituációhoz kapcsolható nagyobb mértékű szívesség figyelembe vétele, és így egy erősebb megköszönési mód (az illokúciós erőt fokozó elemmel kiegészített performatívum: „köszönöm szépen”) alkalmazása. A köztesnyelvi és a magyar L1 tipikus válaszok a következők: (a) *Kösz! / Köszönöm szépen!*, (b) *Kösz!*, (c) *Köszönöm szépen!*

##### 3.1.3 Az elutasítás beszédaktusa

A szituáció leírása: (a) barátod, (b) húgod kéri egy CD-det, de nem akarsz odaadni, (c) egy barátod moziba hív, de nem tudsz elmenni.

A köztesnyelvi és a magyar L1 tipikus válaszok a következők: (a) *Bocsi / Sajnálom, de <indoklás>*, (b) *Nem (+ rövid indoklás) / Nem (adom oda)*, (c) *Bocs / Bocsi, de <indoklás>*. Az (a) és a (c) szituációkban a magyar L1 válaszok esetén megjelent a *Ne haragudj* bocsánatkérési kifejezés, az (a) szituációban 30%-ban, a (c) szituációban

13%-ban, amely kifejezést nem használták a köztesnyelvi válaszolók. A helyette alkalmazott *bocsánat* magyarázata olvasható a 3.2.3 pont alatt.

### 3.2 A beszédaktusok, melyeknek tipikus produkcióiban volt eltérés

A következő beszédaktusok során a vizsgált udvariassági szempontok alapján különbséget lehetett megállapítani a köztesnyelvi és a magyar L1 tipikus válaszok között: az ételkínálás, a kérés, az egyet nem értés (véleménykülönbség) kifejezése, a bocsánatkérés, a bókolás és a bókolásra reagálás, a kívánság és a javaslat kifejezése, valamint az ugratás beszédaktusában. Ezen szituációkra megnéztem a lengyel L1 megnyilatkozásokat, hogy megvizsgáljam a negatív anyanyelvi transzfer lehetőségét.

#### 3.2.1 Az ételkínálás beszédaktusa

A szituáció leírása: (a) osztálytársadat, (b) húgodat, (c) tanárodát kínálsz meg keksszel.

A köztesnyelvi megnyilatkozásokban megjelent az *akar* (lengyelül: *chcieć*) ige az (a) és a (b) szituációban 30%-os, a (c) szituációban 13%-os relatív gyakorisággal, amely arcfenyegetőbb és direktebb a magyar L1 tipikus válaszokhoz<sup>vi</sup> képest: (a) *Kérsz?*, (b) *Kérsz?*, (c) *Megkínálhatom?*

Tipikus köztesnyelvi válaszok az (a) és a (b) szituációban: IL: *Akarsz (egy kekszet)?*, L1: *Chcesz ciastko?* ('Akarsz kekszet?').<sup>vii</sup> A (c) szituációban megjelent a *szeretne* 25%-os relatív gyakorisággal: IL: *Szeretne egy kekszet?*; L1: *Czy miałby Pan ochotę na ciastko?* ('Szeretne egy kekszet?'), amelyre magyar L1-en nem volt példa.

A (c) szituációban az operációs szempontok tekintetében (Blum-Kulka–Olshstein 1984) a köztesnyelvi változat (*Szeretne egy kekszet?*) hallgatóorientált a beszélőorientált magyar L1 változathoz (*Megkínálhatom?*) képest, amely a hallgatótól történő engedélykérést tartalmaz, nyelviileg azt fejezi ki, hogy a hallgató kontrollálja a cselekvést (Blum-Kulka–Olshstein 1984; Blum-Kulka–House–Kasper 1989).

#### 3.2.2 A kérés beszédaktusa

A szituáció leírása: italautomatánál szeretnél pénzt váltani. Akihez fordulsz: (a) egyetemi hallgató, (b) 5 évvel fiatalabb személy, (c) 10 évvel idősebb személy; illetve egy könyved kéred vissza (d) osztálytársadtól, (e) húgodtól.

A köztesnyelvi produktumokban megjelent a hatóige-képző, amely feltételezhetően a lengyel L1-ben szereplő *móc* 'tud (képes)', '-hat/-het' lexikai/szintaktikai elem átvitelének tekinthető, a magyar L1-en nem volt rá példa:

<sup>vi</sup> Magyar L1-en létezik az *Akarsz ([egy] kekszet)?* forma, de a kutatás során nem adták meg az adatközlők.

<sup>vii</sup> A lengyel nyelvről magyarra történő fordítások a tanulmányban a saját fordításaim, melynek során – a lehetőségekhez mérten – a szó szerinti vagy ahhoz közeli fordításra törekedtem. – S. D.



|  | a   | b   | c   | d   | e   |
|--|-----|-----|-----|-----|-----|
| L1(l.): m3c 'tud (k3pes)', '-hat/-het'     | 93% | 83% | 87% | 73% | 24% |
| IL: 'tud vmit csin3lni'                    | 55% | 52% | 52% | 15% | 2%  |
| L1(m.): 'tud vmit csin3lni'                | 92% | 87% | 87% | 57% | 7%  |
| IL: 'lehet vmit csin3lni', '-hat/-het'     | 27% | 25% | 25% | 25% | 9%  |
| L1(m.): 'lehet vmit csin3lni', '-hat/-het' | 0%  | 0%  | 0%  | 0%  | 0%  |

1. t3bl3zat. A k3r3s szitu3ci3iban a lengyel *m3c* 3s a magyar *tud*, valamint a hat3ige-k3pz3 el3fordul3sa (relat3v gyakoris3gi 3rt3ekkel megadva)

A k3ztesnyelvi k3rd3sekben megjelen3 hat3ige-k3pz3 els3sorban lehet3s3gre k3rd3z r3 (a hat3ige-k3pz3 jelent3seir3l r3szletesen l. Csonka 1998), 3gy a partnernek elutas3t3s eset3n t3bb lehet3s3ge van udvariasan elutas3t3sra a k3r3st, a magyar L1-beli *tud* haszn3lat3hoz (a partner k3pess3ge ir3nti tudakoz3d3shoz) k3pest arcv3d3bbnek tekinthet3 (Leech 2014).

Az (a) 3s a (b) szitu3ci3ok eset3n: IL: *Eln3z3st / Bocsi, felv3lthatsz f3mp3nzre?*, a lengyel L1: *Przepraszam, mozesz mi rozmieni3 10 zl?* ('Eln3z3st k3rek, fel **tudsz** nekem v3ltani / felv3lthatsz nekem 10 zlotyt?') / *Masz moze<sup>viii</sup> rozmieni3?* ('Van [amivel] esetleg felv3ltani?'), a tipikus magyar L1: *Fel tudn3d v3ltani?*

A (c) szitu3ci3iban: IL: *Eln3z3st, felv3lth3t f3mp3nzre?*, a lengyel L1: *Przepraszam, czy moglaby Pani rozmieni3 mi ten banknot?* ('Eln3z3st k3rek, fel **tudn3** v3ltani / felv3lth3tn3 nekem ezt a pap3rp3ntz?') / *Ma moze Pan / Pani rozmieni3?* ('Van 3nnek esetleg [mivel] felv3ltani?'), a magyar L1: *Eln3z3st, fel tudn3d v3ltani?*

A (c) szitu3ci3iban, a 10 3vvvel id3sebb szem3ly megsz3l3t3sakor a k3ztesnyelvi v3laszol3k 20%-ban v3laszoltak tegez3dve, a lengyel L1 v3laszol3k 13%-ban, a magyar L1 v3laszol3k pedig 60%-ban (az egyetemi k3zegben megfigyelhet3 lengyel–magyar tegez3si-mag3z3si ar3nyok k3z3tti k3l3nbs3gekr3l l. Koutny 2004). K3ztesnyelven a lengyel L1-hez k3pest kis m3rt3kben magasabb a tegez3s ar3nya, de nem k3zel3ti meg a magyar L1 ar3nyt.

A (d) szitu3ci3iban megjelen3 hat3ige-k3pz3:

IL: *Visszaadhatod a k3nyvetem? / Visszaadhatn3d a k3nyvetem?*, a lengyel L1: *Mozesz mi odd3c moj3 ksi3zke?* ('Visszaadhatod / Vissza **tudod** adni a k3nyvetem?') / *Moglby3 mi odd3c ksi3zke?* ('Visszaadhatn3d / Vissza **tudn3d** adni a k3nyvetem?'), a magyar L1: *Vissza tudn3d adni a k3nyvetem?*

Az (e) szitu3ci3iban mind a k3ztesnyelvi, mind az L1 besz3l3k tipikus v3laszk3nt felsz3l3t3 m3d3 3gealakot haszn3ltak: IL: *Adj vissza a k3nyvetem!*, a lengyel L1: *Oddaj mi ksi3zke!* ('Add vissza a k3nyvetem!'), a tipikus magyar L1 v3ltozat: *Add vissza a k3nyvetem!*

Az (a), a (b) 3s a (c) szitu3ci3okban ismeretlen szem3lyt sz3l3t meg a besz3l3, ami hat3ssal volt a k3ls3 m3dos3t3k (t3mog3t3 elemek) megjelen3s3re: az egyed3li k3sz3n3si forma, a *Szia!* / *Cze3c!* 3s az eln3z3sk3r3s mint figyelemfelkelt3 elem haszn3lat3t a k3vetkez3 t3bl3zat szeml3lteti (v3. Blum-Kulka–Olshtain 1984).

<sup>viii</sup> A *moze* a *m3c* 'tud (k3pes)', hat3ige-k3pz3 E/3. alakja jelen id3ben, kijelent3 m3dban, emellett 'tal3n', '(meg)lehet' jelent3ssel is bír, a k3r3s illok3ci3s erej3t cs3kkent3 elemk3nt haszn3latos. Jelen mondatban az ut3bbi jelent3sben szerepel, a mondat szerkezet szerint nem ragozott 3gei alakban, hanem hat3roz3sz3k3nt 3ll.

|                 | <b>a</b>  |                |                | <b>b</b>  |                |                | <b>c</b>  |                |                |
|-----------------|-----------|----------------|----------------|-----------|----------------|----------------|-----------|----------------|----------------|
|                 | <b>IL</b> | <b>L1 (l.)</b> | <b>L1 (m.)</b> | <b>IL</b> | <b>L1 (l.)</b> | <b>L1 (m.)</b> | <b>IL</b> | <b>L1 (l.)</b> | <b>L1 (m.)</b> |
| köszönés (Szia) | 10%       | 13%            | 50%            | 19%       | 13%            | 33%            | 4%        | 10%            | 25%            |
| elnézéskérés    | 50%       | 50%            | 38%            | 30%       | 50%            | 50%            | 65%       | 60%            | 70%            |
| mindkettő       | 4%        | 3%             | 25%            | 3%        | 0%             | 25%            | 0%        | 0%             | 20%            |

**2. táblázat.** A kérés (a), (b) és (c) szituációiban a köszönés és az elnézésekérés előfordulása (relatív gyakorisági értékekkel megadva)

A (d) és az (e) szituációban:

|                 | <b>d</b>  |                |                | <b>e</b>  |                |                |
|-----------------|-----------|----------------|----------------|-----------|----------------|----------------|
|                 | <b>IL</b> | <b>L1 (l.)</b> | <b>L1 (m.)</b> | <b>IL</b> | <b>L1 (l.)</b> | <b>L1 (m.)</b> |
| köszönés (Szia) | 14%       | 10%            | 13%            | 4%        | 0%             | 0%             |
| elnézéskérés    | 8%        | 7%             | 0%             | 0%        | 0%             | 0%             |
| mindkettő       | 0%        | 0%             | 0%             | 0%        | 0%             | 0%             |

**3. táblázat.** A kérés (d) és (e) szituációiban a köszönés és az elnézést kérés előfordulása (relatív gyakorisági értékekkel megadva)

Jól látható, hogy a köztesnyelvi megnyilatkozásokban a vizsgált külső módosítók használata a lengyel L1 válaszokban megfigyeltékhez hasonlít.

### 3.2.3 A bocsánatkérés beszédaktusa

A szituáció leírása: a buszon véletlenül rálépsz (a) egy 50 éves nő, (b) egy egyidős lány, (c) egy 10 éves fiú lábára.

Az alábbi táblázat összefoglalja a magyar L1 és a köztesnyelvi bocsánatkérési kifejezéseket. Köztesnyelven a magyar L1-beli használathoz képest hangsúlyosabban megjelenik a *bocsánat/bocsi* alkalmazása, másfelől az *elnézést kérek* és a *ne haragudj* kifejezéseket kevésbé használták a köztesnyelvi adatközlők.

|                       | <b>a</b>       |           | <b>b</b>       |           | <b>c</b>       |           |
|-----------------------|----------------|-----------|----------------|-----------|----------------|-----------|
|                       | <b>L1 (m.)</b> | <b>IL</b> | <b>L1 (m.)</b> | <b>IL</b> | <b>L1 (m.)</b> | <b>IL</b> |
| (1) elnézést (kérek)  | 88%            | 11%       | 15%            | 7%        | 5%             | 4%        |
| (2) ne haragudj(on)   | 40%            | 4%        | 55%            | 7%        | 38%            | 7%        |
| (1) + (2)             | 35%            | 4%        | 5%             | 2%        | 0%             | 0%        |
| (3/a) bocsánat        | 0%             | 90%       | 38%            | 45%       | 43%            | 40%       |
| (3/b) bocsi           | 0%             | 0%        | 20%            | 48%       | 13%            | 55%       |
| (3/c) nagyon bocsánat | 0%             | 4%        | 0%             | 0%        | 0%             | 0%        |
| (3) + (2)             | 0%             | 6%        | 20%            | 6%        | 3%             | 7%        |
| (4) (nagyon) sajnálom | 5%             | 6%        | 5%             | 0%        | 5%             | 0%        |

**4. táblázat.** A bocsánatkérési kifejezések magyar L1-en és köztesnyelven (relatív gyakorisági értékekkel megadva)

A lengyel L1 válaszok kb. 90%-ában a *Przepraszam (bardzo)* '(Nagyon) bocsánatot kérek!' kifejezéssel éltek az adatközlők, illetve a (b) és a (c) szituációnál megadták a *Sorry*-t 13-13%-ban. A *przepraszam* 'bocsánatot kérek' performatív ige a lengyel nyelvben a legnépszerűbb IFID (illocutionary force indicating device) forma bocsánatkérés esetén. A kifejezés széles körű használatát jelzi, hogy különböző hierarchiabeli távolságok, státusz és a sérelem súlyától függetlenül is használható. Az angol *I apologize*-zal összehasonlítva, kevésbé formális vagy távolságtartó, anélkül hogy bizalmas stílusúvá válna, másfelől kevésbé fejez ki nyelvileg szerénységet az egyéb IFID-lehetőségekhez képest (Suszczyńska 1999).

### 3.2.4 Az egyet nem értés beszédaktusa

A szituáció leírása: egy osztálytárs egy ételről rosszat mond, úgy gondolja, hogy nem finom; te úgy gondolsz, hogy finom.

A köztesnyelvi megnyilatkozások 40%-a a *szerintem* formával kezdődött, ami a magyar L1 válaszok esetén 18%-ban volt jelen. A feladatkiírásban szereplő *úgy gondolja, úgy gondolod* kifejezések implikálhatták ezt a kezdést, viszont a lengyel L1 válaszolók közül nem volt olyan adatközlő, akinél szerepelt volna ez a nyelvi forma. A köztesnyelvi változatban kizárva a negatív anyanyelvi transzfert, a sikeres köztesnyelvi beszédstratégia-választás elismerését meghagyva, egyéb magyarázat is állhat a háttérben, például ha a magyar a nyelvtanulónak nem az első idegen nyelve, akkor az L2-kompetenciák, beszédstratégiák, vagy egy általános idegen nyelvi kommunikatív stratégia alkalmazása is befolyásolhatta a produkciót.

Emellett a köztesnyelvben 10%-ban megjelent az ún. „konfrontatív interakciós stílus”, amely pragmatikai transzfernek tekinthető, és arcfenyegetőbb, direkter megoldásokat eredményezett a magyar L1 változathoz képest. A következő köztesnyelvi példák nyújtanak betekintést a jelenségre: *Nem igaz. / Nem értek egyet. Nagyon finom nekem! / Nem egyetértek. Ez jó.* A lengyel L1 válaszok között is találhatunk rá példát: *Nie, nie zgadzam się, jak dla mnie jest pyszne!* ('Nem, nem értek egyet, ami engem illet, [ez] nagyon finom!'). A konfrontatív válaszok esetében a megnyilatkozások háttérét adhatta a lengyel interakciós stílus – adott tényezők megléte mellett – konfrontatív, spontán és kötöttség nélkülinek is tekinthető jellege (hogyan lehet mondani, amit az ember gondol) (l. Wierzbicka 2003). A szolidaritás érzése ellen hat a magyar hallgató számára az egyet nem értés kifejezése ezen konfrontatív formában (vö. Brown 2015).

### 3.2.5 A bókolás beszédaktusa

A szituáció leírása: egy osztálytársadnak új táskája van, ami nagyon tetszik neked.

Tipikus köztesnyelvi megnyilatkozás: *Nagyon szép a táskád! Hol vetted?* (a „hol vetted?” rákérdezés 30%-ban jelenik meg köztesnyelven), a lengyel L1: *Masz bardzo ładną torebkę. Gdzie kupiłaś?* 'Nagyon szép táskád van. Hol vetted?' – a „hol vetted?” rákérdezés külső módosítónak, támogató elemnek tekinthető, lengyel L1-en 33%-ban jelent meg, a magyar L1 válaszokban 17%-ban kérdeztek erről. Ha jelentősnek ítéljük a különbséget, akkor negatív pragmatikai anyanyelvi transzfert állapíthatunk meg.

Érdekes megjegyezni, hogy a magyar válaszolók 8%-a nem nyilatkozott a szituációban, a lengyel L1 beszélők közül egy személy adott kitérő jellegű választ, a köztesnyelvi beszélők minden esetben írtak bókokat.

Köztesnyelven sok esetben több elemből áll a bók, a lengyel L1 vagy a magyar L1 változatokhoz képest. Általában a táská szépsége mellett megemlítik, hogy tetszik, például: *Milyen szép táskája. Nagyon tetszik nekem.* A „Hol vetted?” jellegű kérdést leszámítva a több elemből álló bókok aránya köztesnyelven 15%, magyar L1-n 0%, lengyel L1-n egy hasonló jellegű választ lehetett regisztrálni: *Masz świątną torebkę, bardzo mi się podoba.* 'Nagyszerű táskád van, nagyon tetszik.'

A bókolás a szolidaritás kifejezésének az egyik lehetősége, a „hol vetted?” kérdéssel vagy a többelemű bókokkal fokozottabban lehet nyelvileg kifejezni a szolidaritást, mely mindkét esetben a köztesnyelvi válaszoknál jelentősebb számban fordult elő a magyar L1 válaszokhoz képest, a lengyel L1 válaszok arányához közelebbi értéket felmutatva.

### 3.2.6 A bókolásra reagálás beszédaktusa

A szituáció leírása: (a) szomszédasszonyod, (b) egy osztálytársad azt mondja neked: „Milyen jól tudsz magyarul!” (A magyar és a lengyel L1 kontrollszituációra nem tudtam az eredeti szituáció minden tényezőjének megfelelő helyzetet megadni: a beszélgetőtárs a másik fél angol tudását dicsérte meg.)

A köztesnyelvi tipikus válaszok a következők: (a) *Köszönöm (szépen)* (+ a szerénység nyelvi kifejezése), (b) *Köszönöm/Köszí*, amellyel megegyezők a magyar L1 tipikus válaszok, azzal a kitételrel, hogy a szerénység nyelvi kifejezése köztesnyelven 33%-ban jelent meg, lengyel L1-en 27%-ban, addig magyar L1-en 18%-ban, tehát ez utóbbi a kutatás keretén belül nem mondható tipikusnak. Nézzünk meg egy-egy példát az (a) szituációra, IL: *Köszönöm szépen, de én csak kicsit beszélek magyarul*, lengyel L1: *Dziękuję, mogłoby być lepiej.* 'Köszönöm, lehetne jobb', illetve magyar L1: *Köszönöm a dicséretet. Még fejleszteni kell a tudásomon.*

### 3.2.7 A kívánság kifejezésének beszédaktusa

A szituáció leírása: (a) egy barátoddal, (b) húgoddal mentek az utcán, egy könyvesboltba értek, szeretnél bemenni.

A tipikus magyar L1 megnyilatkozás: (a) szituációban: *Menjünk be! Bemegyünk?*, a (b) szituációban (*Gyere,*) *menjünk be!*, amelyhez képest a köztesnyelvi variációkban megjelenik a *móc* 'tud (képes)', hatóige-képző lexikai/szintaktikai elem, az (a) szituációban 25%-os, a (b) szituációban 3%-os relatív gyakorisággal. Belső módosítóként indirektebbé teszi a megnyilatkozást a magyar L1-formához képest (Leech 2014).

IL: (a) *Bemehetünk a könyvesboltba?*, (b) *Menj be velem!*

L1: (a) *Możemy wejść na chwilę do tej księgarni?* ('Be tudunk menni / Bemehetünk egy pillanatra ebbe a könyvesboltba?'), (b) *Chodź do księgarni!* ('Gyere a könyvesboltba!')

A köztesnyelvi megnyilatkozások másik csoportját alkotja a *szeretnék* kifejezéssel kezdődő kívánságmegfogalmazások halmaza, az (a) szituációban 26%, a (b) szituációban 4% relatív gyakorisággal találkozhattunk vele, viszont a lengyel L1

adatközlők részéről nem volt példa rá. A magyar L1 beszélők az (a) szituációban 13%-ban, a (b) szituációban 7%-ban éltek ezzel a kifejezésmóddal. Például köztesnyelvi válasz az (a) szituációnál: *Szeretnék bemenni a könyvesboltba*, magyar L1: *Szeretnék benézni a könyvesboltba. Bejössz te is?*

Köztesnyelvi vizsgálatok részét képezheti, hogy a köztesnyelvi beszélő mennyire érzi szükségét, hogy lexikailag explicit legyen. A jelen szituációban a *könyvesbolt* mint lexikai elem megjelenését lehetett megnézni a megnyilatkozásokban.

| IL  |     | L1 (l.) |     | L1 (m.) |     |
|-----|-----|---------|-----|---------|-----|
| (a) | (b) | (a)     | (b) | (a)     | (b) |
| 75% | 75% | 50%     | 55% | 25%     | 10% |

5. táblázat. A kívánság kifejezésekor a *könyvesbolt* lexikai elem előfordulása (relatív gyakorisági értékekkel megadva)

Jól látható, hogy mind a lengyel L1, mind a magyar L1 beszélőkhöz képest a köztesnyelvi változatban gyakrabban megadták a „könyvesbolt” szót az adatközlők a rá való utalás helyett. (Elképzelhető, hogy a nyelvtanulók még nem tanulták, vagy nem tudták aktívan használni a deiktikus elemeket.)

### 3.2.8 A javaslattevés beszédaktusa

A szituáció leírása: (a) osztálytársad, (b) húgod nem érzi jól magát, javasold neki, hogy menjen el orvoshoz.

A magyar L1 tipikus változatok az (a) szituációban: *(Szerintem) el kéne menned az orvoshoz!* (25%), *Menj el orvoshoz!* (15%), a (b) szituációban: *Menj el orvoshoz!* (45%). A felszólító mód a köztesnyelvi megnyilatkozásoknál nagy arányban megjelenik mind az (a), mind a (b) szituációnál: *Menj el orvoshoz!*, az (a) szituációban 41%, a (b) szituációban 84%-ban, viszont a lengyel L1 adatközlőknél az (a) szituáció 10%-ában, a (b) szituáció 38%-ában jelent meg *Idź do lekarza!* 'Menj (el) (az) orvoshoz!' alakban. A magyar és a lengyel L1 adatközlők hasonló arányban éltek a felszólító móddal, melyhez képest köztesnyelven mindkét szituációban körülbelül duplájára nőtt a felszólító mód használata, ami az egyéb nyelvi lehetőségekhez képest direkter megnyilvánulásnak tekinthető. A jelenség magyarázatának megadásához további kutatások szükségesek.

| IL  |     | L1 (l.) |     | L1 (m.) |     |
|-----|-----|---------|-----|---------|-----|
| (a) | (b) | (a)     | (b) | (a)     | (b) |
| 41% | 84% | 10%     | 38% | 15%     | 45% |

6. táblázat. Javaslattevéskor a felszólító mód használatának előfordulása (relatív gyakorisági értékekkel megadva)

### 3.2.9 Az ugratás beszédaktusa

A szituáció leírása: egy barátod azt mondja, hogy nem tanult a dolgozatra, de mégis maximális pontszámot ért el (általában nem szoktak jól sikerülni a dolgozatai).

A barát által mondottakat önmagában is lehet ugratási megnyilatkozásnak tekinteni, illetve ugratáskiváltó funkcióban állónak is tarthatjuk.

Tipikus köztesnyelvi válasz: *Gratulálok!* (40%), a lengyel L1: *Gratulacje!* ('Gratulálok!') (18%). A magyar L1 adatközlők 3%-ban fejezték ki gratulációjukat.<sup>ix</sup> Lexikai, és egyúttal pragmatikai transzfernek tekinthető a köztesnyelven megjelenő gratuláció.

A köztesnyelvi megnyilatkozáskor egy anyanyelvi szintű nyelvi kompetenciához képest sok esetben a beszélő nem tudja a mondanivalójának minden részletét, nüanszát közvetíteni a másik felé a köztesnyelven, így a köztesnyelvi kompetenciájának megfelelően, tartalmában, illetve az alkalmazott nyelvi eszköztárnak megfelelően egy redukáltabb megnyilatkozással él.

|                            | IL  | L1 (l.) | L1 (m.) |
|----------------------------|-----|---------|---------|
| A gondolati egységek száma | 1,2 | 1,6     | 2,2     |

7. táblázat. Az ugratás szituációjában a gondolati egységek száma (átlag)

Példaként két gondolati egységből álló válaszokat mutatok be: IL: *Hú, gratulálok! Szerencses vagy! / Grat! Nagyon okos vagy!*, lengyel L1: *Hmm, na pewno się nie uczyła! Gratulacje.* 'Hmm, biztos nem tanultál! Gratulálok!', magyar L1: *Biztos szerencséd volt. Vagy puskáztál?*

A köztesnyelvi válaszolók 5%-a kihúzta a válaszadás helyét, vagy odaírta, hogy nem mondana semmit, a lengyel L1 és a magyar L1 válaszolók minden esetben adtak választ.

A köztesnyelvi válaszolók közül volt, aki visszaugratott: *Mint mindig* ironikus válasszal, illetve a *Hehe, a hülyéknek mindig szerencséje van* reagálással, ez utóbbi kifejezés megjelent a lengyel L1 válaszok között is: *Glupi ma szczęście.* 'Hülyének van szerencséje.' A magyar L1 válaszok között visszaugratásnak elsősorban akkor lehetett értelmezni egy-egy megnyilatkozást, amikor emotikonokkal jelezte a kérdőívkitöltő, hogy játékos, humoros keretben válaszol, mint például a *Jó bemákoladt! :D Puskáztá??* válasz esetén (l. Sitkei 2017).

<sup>ix</sup> A magyar L1 válaszok megoszlása főbb vonalakban a következő szemantikai-pragmatikai kategóriákhoz köthető: szerencse feltételezése (25%), puskázás feltételezése (20%), *Ügyes voltál!* (15%), értetlenkedés (15%), hitetlenkedés (12%), valamely egyéb magyarázat feltételezése (10%). A köztesnyelvi válaszok között a gratuláció mellett megjelent még: a szerencse feltételezése (20%), értetlenkedés (12%), puskázás feltételezése (8%), *Ügyes voltál!* (8%), hitetlenkedés (6%), a hazugsággal vádolás (6%). A lengyel L1 válaszokban: értetlenkedés (20%), szerencse feltételezése (17%), hitetlenkedés (17%), hazugsággal vádolás (14%), *Ügyes voltál!* (14%), puskázás feltételezése (14%), valamely egyéb magyarázat (10%), emellett a biztatás (7%) és a dicséret (7%) kategóriája is megjelent. A válaszok sok esetben több elemből álltak, így a teljes választ egyszerre több kategóriához is be lehetett sorolni.

## 4 Köztesnyelvi megfigyelések

### 4.1 Nyelvi túlbiztosítás

A bókolás beszédaktusában láttuk, hogy a lengyel L1 és a magyar L1 bókokhoz képest köztesnyelven megnő a többemű bókok száma, a szolidaritást a másikkal fokozottabban kifejezve.

A kívánság kifejezésének beszédaktusában megfigyelhettük, hogy a köztesnyelvi beszélő mind a magyar L1-hez, mind az anyanyelvéhez képest explicitebben nyilvánult meg lexikai szinten. Kellerman (1983) szerint a köztesnyelvi megnyilatkozás egyik jellemzője az explicittség (a szisztematikusság és a logikusság mellett).

### 4.2 Az udvariasságról

Az egyes transzferjelenségek célnyelvi udvariassági hatását az említett udvariassági szempontok alapján a köztesnyelvi vizsgálat fejezetben beszédaktusonként röviden bemutattam. Jelen fejezetben összefoglalom az udvariasságot érintő főbb megfigyeléseket.

A kutatás során megfigyelhető volt, hogy egy-egy külső vagy belső módosító (Blum-Kulka-Olshtain 1984) hatására lett arc(ulat)védőbb vagy fenyegetőbb, direkter vagy indirekter, illetve adott esetben szolidárisabb a vizsgált köztesnyelvi jelenség a célnyelvi változathoz képest.

Az operációsszempont-váltások esetében, az ételkínálás harmadik szituációjában, amikor az adatközlők egy tanárukat kínálják meg keksszel, a köztesnyelvi változat hallgatóorientált (Szeretne...?) a beszélőorientált magyar L1 változathoz (Megkínálhatom?) képest. A magyar változat a hallgatótól történő engedélykérést tartalmaz, nyelvileg azt fejezi ki, hogy a hallgató kontrollálja a cselekvést, ami a köztesnyelvi változathoz képest a hallgató számára kevésbé arcfenyegető (Blum-Kulka-Olshtain 1984; Blum-Kulka-House-Kasper 1989).

A köztesnyelvi variációkban megjelenő m<sup>o</sup>c 'tud (képes)' hatóige-képző lexikai/szintaktikai elem hatóige-képzőként való átvitele belső módosítóként hat. A kérés beszédaktusában a köztesnyelvi változatban megjelenő hatóige-képző lehetőségre utal, arra rákérdezve a másiknak – elutasítási szándéka esetén – több mozgásteret van udvariassággal elutasítania a kérést, a magyar L1 tud használatához (a partner képessége iránti tudakozódáshoz) képest arcvédőbbnek, és így udvariasabbnak tekinthető (Leech 2014). A kívánság kifejezésének beszédaktusában a köztesnyelven megjelenő hatóige-képző a tipikus magyar L1 formákhoz (Menjünk be! / Bemegyünk?) képest szintén arcvédőbbé, és így udvariasabbá teszi a megnyilatkozást (Bemehetünk a könyvesboltba?) (vö. Brown 2015). A hatóige-képző egyik használati módja, amelyet magyar nyelvtanulás során viszonylag hamar elsajátítanak a nyelvtanulók, a „szerény, udvariasság kérés”-ként való használata (Csonka 1998), ami hatással lehetett a köztesnyelvi megnyilatkozásokra, udvariasság szándékú megnyilatkozás esetén ezen forma választásának előnyben részesítésével.

A bókolás és az egyet nem értés beszédaktusában megfigyelhető konfrontatív válaszok példát nyújtottak a szolidaritás kifejezéséhez kapcsolódó nyelvi jellegzetességek megfigyelésére. A szolidaritás és az udvariasság kapcsolatát tekintve elmondható, hogy a szolidaritás nyelvi kifejezése a Brown-Levinson-i pozitív

udvariassághoz köthető (Brown–Levinson 1987), így a nyelvileg fokozottabban megjelenő szolidaritás a köztesnyelven a bókolás terén udvariasabb megnyilatkozásnak tekinthető a magyar L1 és a lengyel L1 válaszokhoz képest.

Az udvariassági vizsgálatoknál érdemes lehet szétválasztani azon jelenségeket, amikor az anyanyelvi transzfer magyar L1-beli értékelését végezzük, azon jelenségektől – mint a bókolásnál láttuk –, hogy a lengyel L1 válaszokhoz képest is fokozta a beszélő a bókot, ami nyelvi túlbiztosításnak tekinthető. Az előbbi esethez kapcsolódó vizsgálatok a köztesnyelvi udvariassági vizsgálatok módszertani kidolgozásához járulhatnak hozzá, míg az utóbbi megfigyelések a köztesnyelvi kommunikációs stratégiák megismerését szolgálhatják.

## 5 Zárógondolatok

A kutatás során lengyel L1 beszélők magyar köztesnyelvi tipikus produkcióiban azonosítottam be a magyar L1 változathoz képest a lengyel L1 tipikus megnyilatkozásokhoz hasonló jellemzőket felmutató nyelvi jelenségeket (itt feltételezhető a negatív anyanyelvi transzfer) – az ételkínálás, a kérés, az egyet nem értés (véleménykülönbség) kifejezése, a bocsánatkérés, a bókolás és a bókolásra reagálás, a kívánság és a javaslat kifejezése, valamint az ugratás beszédaktusában. Az egyes jelenségek célnyelvi hatásának jellemzőit udvariassági tényezők alapján néztem meg egy – köztesnyelvi udvariassági vizsgálatokhoz megfelelő – módszer megtalálása céljából. A köztesnyelvi beszélő egy nyelvi forma átvitelekor adott esetben (ha nyelvi kompetenciája megengedi) választhat a célnyelvi változatok között, ezen választásait szeretném a kutatás folytatásaként részletesebben vizsgálni.

A vizsgálati eredmények alapján foglalkoztam a nyelvi túlbiztosítás jelenségével is a köztesnyelvi kommunikációs stratégiák közelebbi megismerése céljából.

## Irodalom

- Bándli, J. 2015. (Túl) Direkt stratégiák a célnyelvi kommunikációban. *THL*, 2(1): 5–14.
- Beebe, L. M., Takahashi T., Uliss-Weltz R. 1990. Pragmatic transfer in ESL refusals. In: Scarcella, R. C., Andersen E., Krashen S. C. (szerk.) *Developing communicative competence in a second language*. New York: Newbury House. 55–73.
- Beebe, L. M., Cummings, M. C. 1996. Natural speech act versus written questionnaire data: How data collection method affects speech act performance. In: Gass, S. M., Neu, J. (szerk.) *Speech Acts across Cultures*. Berlin: Mouton de Gruyter. 65–86.
- Blum-Kulka, S., Olshtain, E. 1984. Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics*, 5(3): 196–213.
- Blum-Kulka, S. 1987. Indirectness and politeness in requests: same or different? *Journal of Pragmatics*, 11: 131–146.
- Blum-Kulka, S., House J., Kasper G. 1989. *Cross-cultural Pragmatics: Requests and Apologies*. Norwood, NJ: Ablex Publishing Corporation.
- Brown, P., Levinson S. C. 1987. *Politeness. Some universals in language use*. Cambridge: Cambridge University Press.
- Brown, P. 2015. Politeness and language. In: Wright, J. D. (szerk.) *The International Encyclopedia of the Social and Behavioural Sciences (IESBS)* (2nd ed.). Amsterdam: Elsevier. 326–330.
- Corder, S. P. 1992. A Role for the Mother Tongue. In: Gass, S. M., Selinker, L. (szerk.) *Language Transfer in Language Learning*. Amsterdam, Philadelphia: John Benjamin Publishing Company. 18–31.



- Csonka, Cs. 1998. A *-hat/-het* képzőről. *Intézeti Szemle*, 20(1): 5–13.
- Jobst, A. 2007. A *mi* mint a hatalom és a szolidaritás névmása. *Magyar Nyelvőr*, 131: 29–46.
- Kellerman, E. 1983. Now You See It, Now You Don't. In: Gass, S., Selinker, L. (szerk.) *Language Transfer in Language Learning*. Rowley, Mass.: Newbury House.
- Koutny, I. 2004. Hungarian Greetings and Forms of Address in a Cross-Cultural Approach. *Modern Filológiai Közlemények*, 6: 5–20.
- Kugler, N. 2013. A szubjektívizáció jelenségének nyelvészeti értelmezései. *Magyar Nyelvőr*, 137: 8–30.
- Leech, G. N. 2014. *The pragmatics of politeness*. New York: Oxford University Press.
- Selinker, L. 1972. Interlanguage. *International Review of Applied Linguistics*, 10: 209–231.
- Sitkei, D. 2017. Hogyan ugratnak más nyelvek beszélői – egy empirikus kutatás tanulságai. In: Bíró, Gy. (szerk.) *Szakkollégiumi füzetek 4. Móra Akadémia szakkollégiumi tanulmánykötet*. Szeged: SZTE Móra Ferenc Szakkollégium. 67–84.
- Suszczyńska, M. 1999. Apologizing in English, Polish and Hungarian: Different languages, different strategies. *Journal of Pragmatics*, 31: 1053–1065.
- Szili, K. 2004. *Tetté vált szavak. A beszédaktusok elmélete és gyakorlata*. Budapest: Tinta.
- Wierzbicka, A. 2003. *Cross-cultural pragmatics: the semantics of human interaction* (2nd edition). Berlin–New York: Mouton de Gruyter.

## A negatív emotív elemek vizsgálata a nemek közötti nyelvhasználati különbségek szempontjából<sup>i</sup>

Szabó Martina Katalin<sup>1,2,3</sup>, Lázár Bernadett<sup>1</sup>, Nyíri Zsófi<sup>1</sup>, Morvay Gergely<sup>1</sup>

<sup>1</sup>Precognox Informatikai Kft

<sup>2</sup>Szegedi Tudományegyetem, Bölcsészettudományi Kar, Szláv Intézet

<sup>3</sup>„Lendület” Research Center for Educational and Network Studies (RECENS)

szabo.martina@lit.u-szeged.hu

(mszabo, blazar, zsnyiri, gmorvay)@precognox.com

**Kivonat:** A dolgozat a negatív emotív szemantikai tartalmú elemeket vizsgálja egy magyar nyelvű Twitter-korpuszban, a női és férfi nyelvhasználati különbségekre fókuszálva. A vizsgálat tárgyát azok a nyelvi elemek képezik, amelyek lexikai szinten negatív emotív szemantikai tartalommal rendelkeznek ugyan, azonban fokozó funkcióban képesek azt részben vagy egészben elveszíteni (pl. *mocskos nagy, rohadjt jól*), sőt bizonyos esetekben arra is, hogy negatív tartalmuk ellenére pozitív értékelést fejezzenek ki (pl. *brutális volt a hétvégi koncert*). A vizsgált jelenség mind elméleti, mind alkalmazott nyelvészeti, különösen nyelvtechnológiai szempontból figyelemre méltó. A probléma kimerítő elméleti nyelvészeti leírása még nem történt meg, hatékony nyelvtechnológiai kezelése pedig egyelőre nem megoldott. A vizsgálati korpuszt, amelyet a kutatás céljainak megfelelően kézzel annotáltunk, női és férfi felhasználók által írt magyar nyelvű Twitter-bejegyzések (tweetek) alkotják. A korpusz létrehozásának fő célja egy olyan adatbázis megalkotása volt, amely lehetővé teszi az adott elemcsoport beható, szemantikai-pragmatikai szempontú vizsgálatát. A dolgozat ismerteti a korpusz létrehozásának elveit, menetét és eszközeit, valamint az annotált korpusz kvantitatív és kvalitatív elemzési eredményeit a férfi és női nyelvhasználati különbségek tekintetében. Mindemellett bemutatjuk azokat az eredményeket is, amelyeket a teljes Twitter-korpuszon kaptunk az adatbázis szótáralapú automatikus emócióelemzése segítségével.

<sup>i</sup> A jelen kutatást Az Emberi Erőforrások Minisztériuma Új Nemzeti Kiválóság Programja, az MTA Társadalomtudományi Kutatóközpont „Lendület” RECENS Kutatócsoportja támogatta, valamint a European Research Council (ERC) az Európai Unió „Horizon 2020” nevű, 648693-es számú kutatási és fejlesztési programja keretében.

## 1 Bevezetés

A dolgozat egy, az elméleti nyelvészetben kevésbé tárgyalt, és a nyelvtechnológia számára is problematikus jelenséget, a negatív emotív szemantikai tartalmú elemeket veszi górcső alá, elsősorban a női és férfi nyelvhasználati különbségek szempontjából.

A vizsgálat tárgyát azok az elemek képezik, amelyek lexikai szinten negatív emotív szemantikai tartalommal rendelkeznek ugyan, azonban fokozó funkcióban képesek azt részben vagy egészben elveszíteni (pl. *mocskos nagy*, *rohadt jól*), sőt bizonyos esetekben arra is, hogy negatív tartalmuk ellenére pozitív értékelést fejezzenek ki (pl. *brutális volt a hétvégi koncert*) (vö. Szabó 2015). Az előbbi esetben a vizsgált elemek más, értékkelő kifejezések jelentését fokozzák (ezt a jelenséget nevezzük **értékvesztésnek**), az utóbbi esetben pedig maguk töltik be értékkelő kifejezés (szentimentkifejezés) funkcióját (ezt a jelenséget nevezzük **értékváltásnak**) (vö. Szabó et al. 2017).

A vizsgált jelenség a szentimentelemzés egy kardinális részproblémájához, a szentimentérték módosulásához tartozik. A szentimentérték módosulásának azt a jelenséget nevezzük, amikor egy adott nyelvi elem lexikai szintű szentimentértéke nem azonos, vagy nem teljes mértékben azonos az őt magában foglaló teljes megnyilatkozás szentimentértékével (vö. Szabó 2015). Egy szentimentkifejezés lexikai szintű értéke számos okból kifolyólag eltérhet a bennfoglaló megnyilatkozás értékétől. Így például egy kifejezés értéke negálható (pl. *nem jó*), bizonytalanná tehető (pl. *kevésbé jó*), vagy az ironia eszközével az ellentétére fordítható (pl. *ezt jól megcsináltad!*).

A jelen dolgozat a szentimentérték módosulásának fent már említett két speciális típusát vizsgálja: az értékváltást és az értékvesztést. A jelenség mind elméleti, mind alkalmazott nyelvészeti, különösen nyelvtechnológiai szempontból figyelemre méltó. Kimerítő elméleti nyelvészeti leírása még nem történt meg, hatékony nyelvtechnológiai kezelése pedig egyelőre nem megoldott (vö. pl. Andor 2011; Kugler 2014; Szabó 2015; Drávucz–Szabó–Vincze 2017). Ezekre az elemekre ugyanis a szótáralapú automatikus szentiment- és emócióelemzés – a lexikai szintű polaritásuktól való eltérés miatt – téves elemzési eredményt ad. Bízunk abban, hogy a kutatás során kapott eredményeket fel tudjuk használni egy olyan elemző szoftver építésében, amely képes lesz a vizsgált, a szentimentelemzés számára problémás jelenséget hatékonyan kezelni.

A dolgozatban a következőket tárgyaljuk részletesen: 1) bemutatjuk a vizsgált nyelvi jelenséget, és a szakirodalom eddigi vonatkozó megállapításait; 2) ismertetjük a vizsgálati korpusz létrehozásának menetét, módszerét és eszközeit; 3) bemutatjuk az elemzés részletes, kvalitatív és kvantitatív eredményeit, a korpuszból vett példákkal szemléltetve a kutatás legfontosabb tapasztalatait. Mivel a saját emóciószótárainkkal (vö. Szabó–Morvay 2015, Szabó–Vincze–Morvay 2016) lehetőség nyílt a teljes Twitter-korpusz automatikus emócióelemzésére is, és a kutatás alapproblémája a negatív emotív szemantikai tartalmú elemek nyelvhasználati sajátosságai, elvégezzük a korpusz szótáralapú automatikus érzelemelemzését is, majd az eredményeket a nemek közötti nyelvhasználati eltérések szempontjából elemezzük.

## 2 Szakirodalmi áttekintés

A szakirodalom vonatkozó eredményei alapján a következő legfontosabb megállapításokat tehetjük a negatív emotív fokozó elemek nemek közötti nyelvhasználati eltéréseit illetően:

Székely (2006) amellett érvel, hogy a magyar nyelv női és férfi beszélői a fokozásra nem azonos elemeket használnak. Véleménye szerint ugyanis a nők gyakrabban élnek az ún. „illogikus fokozó értelmű” kapcsolatokkal, mint amilyen például az *eszméletlen módon örülnek nekem, a borzalmasan imádom a görögöket, az iszonyú jó érzés* vagy a *borzasztóan élveztem*. Ezekben a kifejezésekben a szókapcsolat feje pozitív szemantikai tartalommal rendelkezik, míg annak fokozója negatív elsődleges polaritást hordoz. Székely (2006) megállapításaira reflektálva Balogh (2009) hozzát teszi, hogy bár a férfiak nyelvhasználatában is előfordulnak hasonló kifejezések, „mégis inkább a nőknél gyakoribbak”.

Székely (2006) véleménye szerint a fokozó értelmű kifejezések közül a durvább, vulgárisabb elemek inkább a férfiak nyelvhasználatát jellemzik (pl. *rohadtul, piszkosul, mocskosul* stb.). Balogh (2009) ezzel a megállapítással egyetért, és kiemeli, hogy az eltérést statisztikai adatokkal igazolni lehetne. Ugyanakkor arról is említést tesz, hogy a nemek közötti nyelvhasználati eltérés folyamatosan csökken, és mára számos fokozó elem már uniszexszé vált.

Jing-Schmidt (2007: 432) ugyancsak megállapítja a vulgáris fokozó elemeket érintő nemek közötti nyelvhasználati eltéréseket. Dolgozata alapján a német anyanyelvű beszélők elfogadhatóbbnak találják, amikor férfi beszélő használja a német *verdammst* ('átkozott') vulgáris emotív fokozó elemet, mintha nő teszi ezt. Ugyanerről számolnak be az angol anyanyelvű adatközlők is a *damn(ed)* ('átkozott, átkozottul, rohadt') vonatkozásában. Jing-Schmidt (2007: 432) amellett érvel, hogy a fenti megfigyelések, még ha csupán indirekt módon is, de azzal a kulturális sajátossággal vannak kapcsolatban, hogy míg a férfi beszélőktől a harag érzelem nyelvi kifejezései elfogadottabbak, addig a női beszélőktől kevésbé. Másképpen: a harag emóció verbális kifejeződését a két nem részéről a társadalom nem egyformán tolerálja. A szerző felhívja a figyelmet azokra a szociológiai tárgyú kutatásokra is, amelyek a harag kifejezésére irányulnak, és a nemek között aszimmetriát állapítanak meg. Véleménye szerint ez a szocializációs eltérés összefüggésben állhat a harag érzelmét kifejező fokozó elemek használati sajátosságaival is.

## 3 A nyers korpusz bemutatása és elemzési eredményei<sup>ii</sup>

A nyers korpusz, amelyből a kutatáshoz szükséges anyagot gyűjtöttük, összesen 37 818 magyar nyelvű tweetből áll, ebből 19 559 tweet női, 18 236 pedig férfi szerzőségű. Az 1. táblázat bemutatja a korpusz alapvető statisztikai adatait.

---

<sup>ii</sup> Nagyon köszönjük Ilyés Virágnak a dolgozat végső változatának elkészítése során nyújtott hathatós segítségét.

|   | Nő    | Férfi | Összes |
|---|-------|-------|--------|
| A tweetekben lévő mondatok átlagos száma (eszköz: magyarlanc, Zsibrita et al. 2013) | 1,47  | 1,58  | 1,53   |
| A tweetekben lévő szavak átlagos száma (alfanumerikusan)                            | 10,97 | 10,89 | 10,93  |
| A tweetekben lévő karakterek átlagos száma  | 79,29 | 81,93 | 80,56  |

**1. táblázat.** A nyers korpusz alapvető statisztikai adatai

A fenti statisztikai adatok alapján a nők és a férfiak korpusza az alapvető kvantitatív sajátságokat tekintve nem mutat érdemi eltérést. Ez a tapasztalat a további vizsgálatok (az emóciók detektálása és a negatív emotív elemek elemzése) szempontjából is fontos, hiszen így effektíve nincs nagyobb „esélye” egyik nemnek sem a nagyobb szószám által erősebb emóciópontszám elérésére a szótáras elemzésben, emellett a negatív emotív fokozó elemek gyakoribb használatára sem.

Szerettük volna feltárni, hogy a férfi és női szerzőségű tweetek milyen számszerű eltérést mutatnak a különböző emotív tartalmak tekintetében. Azt kívántuk tehát megvizsgálni, hogy a szótárillesztéses megoldással kapott emóciópontszámokban van-e szignifikáns eltérés a férfi és a női szerzőségű tweetek között. Ennek céljából a két teljes, nyers alkorpuszt automatikus módszerrel elemeztük a 6 és a 8 érzelmtípus nyelvi elemeit tartalmazó emóciólexikonunkkal.

Az emóciószótáraink létrehozásakor Ekman és Friesen (1969) érzelmkategorizálási rendszerét követtük, tehát azt a hat alapérzelmet vettük figyelembe, amelyek arckifejezéseit a kutatások alapján kultúrafüggetlenül azonos módon produkáljuk és azonosítjuk (vö. Szabó–Morvay 2015). Az alapérzelmek, amelyekre építve tehát a szótárunk kifejezéseit kategorizáltuk, a következők: BÁNAT, DÜH, FÉLELEM, MEGLEPŐDÉS, ÖRÖM, UNDOR. Az elkészített szótár a hat kategóriában összesen 1798 emóciókifejezést tartalmaz.

Egy későbbi kutatásunkban, egy emóciókorpusz annotálása során azt tapasztaltuk, hogy a hat kategóriát további két érzelmtípussal szükséges kibővítenünk: a FESZÜLTÉG és a VONZALOM kategóriákkal (vö. Szabó–Vincze–Morvay 2016). A meglévő rendszerrel ugyanis gyakran a korpusz kifejezéseit nem tudtuk lefedni a feldolgozó munka során. Az ekmani 6 érzelem mellé felvettük tehát e két új emóciót is, és a nyolc kategóriával átdolgoztuk a meglévő teljes emóciólexikont. Az így kialakított új szótár összesen 1842 elemet tartalmaz.

A két szótárral elvégzett elemzések eredményeit az alábbi táblázatokban közöljük:

| Emóció            | Férfiak<br>átlaga | Nők<br>átlaga | Különb-<br>ség | Standard<br>hiba | t-érték<br>(szabadság-<br>fok) | Szórás      | p-érték |
|-------------------|-------------------|---------------|----------------|------------------|--------------------------------|-------------|---------|
| <b>bánat</b>      | 0,00189           | 0,00207       | -0,0001<br>8   | 0,00017          | -1,07812<br>(36543)            | 0,0233<br>6 | 0,28099 |
| <b>düh</b>        | 0,00098           | 0,00111       | -0,0001<br>3   | 0,00014          | -0,97661<br>(36716)            | 0,0185<br>6 | 0,32877 |
| <b>félelem</b>    | 0,0006            | 0,00079       | -0,0001<br>9   | 0,0001           | -1,8909<br>(32144)             | 0,0141<br>2 | 0,05865 |
| <b>meglepődés</b> | 0,00047           | 0,00076       | -0,0002<br>9   | 0,00012          | -2,49625<br>(36667)            | 0,0159      | 0,01256 |
| <b>öröm</b>       | 0,00801           | 0,00685       | 0,00116        | 0,00038          | 3,07792<br>(37360)             | 0,0520<br>6 | 0,00209 |
| <b>undor</b>      | 0,00155           | 0,00182       | -0,0002<br>7   | 0,00016          | -1,71795<br>(34442)            | 0,0220<br>1 | 0,08582 |

2. táblázat. A 6 szólistával végzett elemzés eredményei<sup>iii</sup>

| Emóció            | Férfiak<br>átlaga | Nők<br>átlaga | Különbség | Standard<br>hiba | t-érték<br>(szabadságfok) | Szórás  | p-érték |
|-------------------|-------------------|---------------|-----------|------------------|---------------------------|---------|---------|
| <b>bánat</b>      | 0,00175           | 0,00152       | 0,00023   | 0,00015          | 1,48971<br>(37691)        | 0,02069 | 0,13631 |
| <b>düh</b>        | 0,00033           | 0,00019       | 0,00014   | 0,00006          | 2,22208<br>(31482)        | 0,00871 | 0,02629 |
| <b>félelem</b>    | 0,00038           | 0,00024       | 0,00014   | 0,00006          | 2,29052<br>(37622)        | 0,00844 | 0,022   |
| <b>meglepődés</b> | 0,00034           | 0,00022       | 0,00012   | 0,00008          | 1,44357<br>(32549)        | 0,01158 | 0,14887 |
| <b>öröm</b>       | 0,00708           | 0,00579       | 0,00129   | 0,00036          | 3,64017<br>(37086)        | 0,04905 | 0,00027 |
| <b>undor</b>      | 0,00141           | 0,00128       | 0,00013   | 0,00014          | 0,95721<br>(36853)        | 0,01914 | 0,33847 |
| <b>feszültség</b> | 0,001             | 0,00068       | 0,00032   | 0,00011          | 2,88852<br>(37162)        | 0,01536 | 0,00387 |
| <b>vonzalom</b>   | 0,00076           | 0,00048       | 0,00028   | 0,0001           | 2,78117<br>(35838)        | 0,0139  | 0,00542 |

3. táblázat. A 8 szólistával végzett elemzés eredményei<sup>iv</sup>

<sup>iii</sup> A férfítweetek száma (N) 18236, a női tweetek száma 19559, az összevetés kétoldalú, kétmintás t-próbával készült.

<sup>iv</sup> A férfítweetek száma (N) 18236, a női tweetek száma 19559, az összevetés kétoldalú, kétmintás t-próbával készült.

A két nem közötti különbségre kétoldali t-próbát számoltunk, hogy lássuk, hogy statisztikailag szignifikánsak-e az eltérések. A táblázatok utolsó oszlopa tartalmazza az erre vonatkozó két értéket: az első a t-próba értéke, a második a p-érték. A p-érték alapján tudjuk meghatározni, hogy az eltérés szignifikáns-e.

A jelen vizsgálat során először azt a hipotézist teszteltük, amely szerint a nők és a férfiak tweetjeiben megjelenő emóciók mennyisége egyenlő. Ezt a nullhipotézist a 6 elemből álló szótár esetében az **ÖRÖM** illetve **MEGLEPŐDÉS** emócióknál el is vethettük, hiszen ezen emóciók esetében a p-érték (bizonytalansági faktor, amely minél kisebb, annál nagyobb eltérésre utal) kisebb volt, mint 0,05, tehát a két csoport közötti eltérés e két emóció esetében szignifikáns. Az elemzés rámutatott továbbá arra is, hogy a **MEGLEPŐDÉS** emóció a férfiak tweetjeiben, míg az **ÖRÖM** emóció a nők tweetjeiben jelent meg többször. Amint az az adatokból ugyancsak látszik, bár a többi emóció esetében is megfigyelhető a két nem közötti eltérés, azok nem jelentős mértékűek.

A 8 elemből álló szótár eredményei alapján a **DÜH**, a **FÉLELEM**, a **FESZÜLTSG**, az **ÖRÖM** és a **VONZALOM** emóciók esetében volt a p-érték kisebb, mint 0,05, tehát a két nem tweetjeiben ezen emóciók frekvenciája mutatott érdemi eltérést. Ezek az emóciók a nők tweetjeiben jelentek meg gyakrabban.

#### 4 A szemantikai-pragmatikai szempontból annotált vizsgálati korpusz létrehozása

A kézzel annotált korpusz létrehozásához a nyers korpusz azon nyelvi adataira volt szükségünk, amelyek tartalmaztak legalább egy, értékváltásra vagy értékvesztésre képes elemet. A feladathoz kézzel összeállítottunk egy listát, amely összesen 109 értékváltásra és/vagy értékvesztésre képes elemet tartalmaz (vö. Szabó et al. 2017). A lista alapján végzett szűréssel összesen 610 tweetet nyertünk ki a nyers korpuszból. A nyers korpusz kézi feldolgozásához a brat nevű, online elérhető annotálóprogramot használtuk.<sup>v</sup> (A munkához az annotálni kívántak alapján az eszközt magunk állítottuk be (az eszközről részletesebben l. Szabó et al. 2017).

Az adatbázis feldolgozását a következő annotációs szempontok alapján végeztük el manuálisan: először eltérő tagekkel bejelöltük azokat az elemeket, amelyek az aktuális kontextusban értékváltást vagy értékvesztést mutattak. Ennek megfelelően azokat az eseteket, ahol a vizsgált elem aktuális szentimentértéke megegyezett annak elsődleges negatív értékével, nem annotáltuk, pl.

[1] Egy **rohadt** kukásautó miatt áll már vagy tíz perce a trolis.

Az értékváltás esetében a targetet [2], az értékvesztés esetében pedig a vizsgált elem által módosított szerkezeti alaptagot is annotáltuk [3]. (A vizsgált elemeket ebben az esetben is kövér szedéssel, míg a további annotált kifejezéseket aláhúzással jelöljük.)

[2] ismét **brutális koncertet** adott az énekes

[3] Valljátok be, **rohadt jó** az időérzésem!

<sup>v</sup> Elérhető: <http://brat.nlplab.org>. Letöltve: 2017. október 6.

Az annotálásnak az volt a célja, hogy segítségével vizsgálni lehessen egyrészt a fokozó szerepű, értékvesztésre képes elemek és az általuk módosított elemek, valamint az értékváltásra képes elemek és azok targeteinek a kapcsolatait.

Az értékvesztésre vonatkozóan további tageket is felvittünk az annotált elemek szemantikai–pragmatikai viselkedése alapján. E megoldás alkalmazása mellett korábbi vizsgálati eredményeink alapján döntöttünk: megfigyeltük, hogy e fokozó elemek nem deszemantizálódnak minden esetben (részletesen l. Szabó 2017, kézirat), pl. (A példákban a vizsgált elemet kövérrel, az alaptagot aláhúzással jelöljük.)

[4] 166 centi 46 kiló az milyen egy 14 éves lánynak? Nem vagyok **rohadt magas?** (vö. *túl*)

A fenti példában meglátásunk szerint éppen a vizsgált elem adja hozzá a negatív értékelést a szöveghez. Annak céljából, hogy ezt a jelenséget a korpusz alapján vizsgálni lehessen, mind a fokozó elemek, mind a módosított alaptagok aktuális szemantikai sajátságát jelöltük (részletesebben l. Szabó et al. 2017).

Azért, hogy a nem várt szemantikai–pragmatikai sajátságokat is annotálni tudjunk, minden, szemantikai–pragmatikai sajátságot jelölő tag számára létrehoztuk annak „egyéb” párját is, és az előzetesen kidolgozott rendszerbe nem illeszkedő jelenségeket ezekkel jelöltük.

A bemutatott feldolgozási megoldást az annotátorok egy részletes annotálási útmutatón keresztül ismerték meg, amelyben az egyes jelenségeket példák segítségével prezentáltuk.

Az 1. ábra egy részletet közöl a korpusz annotációjából a brat programban.



1. ábra. Részlet a korpusz annotációjából a brat programban

Ahhoz, hogy az annotátorok közötti konszenzust mérni tudjunk, a teljes korpusz feldolgozása előtt elvégeztünk egy egyetértésmérést a korpusz egy kisebb részletén (vö. Szabó et al. 2017).

Az egyetértésmérés a számítógépes nyelvészetben a korpuszok kézi annotálása során bevett minőségellenőrzési módszer. Ennek elvégzéséhez a korpusz egy bizonyos részét (lehetőség szerint a 10%-át) az összes annotátor feldolgozza, majd az annotációkat valamely, erre a célra fejlesztett algoritmus segítségével összevetik egymással. Ilyen mérési eszköz például az ún. Cohen's Kappa-statisztika (angolul: *Cohen's Kappa statistic*) (vö. Cohen 1960). A méréssel kapott értékek (annotátorok



közötti egyetértési ráta, angolul: *inter-annotator agreement rate*) megmutatja, hogy az annotátorok hány esetben látták el annotációs taggal a korpusz bizonyos egységeit, valamint azt is, hogy hány alkalommal használták rájuk ugyanazt a taget. A Cohen's Kappa-statisztika a hagyományosabbnak tekinthető egyszerű százalékos egyetértés mérése mellett azt is megmutatja, hogy az egyetértési ráta mennyiben módosul a véletlenszerű egyezések lehetőségének a figyelembe vételével. Mindezek következtében a Cohen's Kappa-statisztika komplex, a minőség ellenőrzésére jól alkalmazható mutatónak tekinthető.

Az egyetértésméréshez összesen 100 tweetet annotáltunk, tekintettel arra, hogy már ez a mennyiség is a teljes korpusz hatod részét tette ki. A kapott adatok felhasználásával az annotálást követően Cohen's Kappa-statisztikát készítettünk. A mérési eredményeket a következő táblázatnak megfelelően értelmezhetjük:

| Értelmezés | -tól | -ig  |
|------------|------|------|
| gyenge     | 0,00 | 0,40 |
| közepes    | 0,41 | 0,60 |
| jó         | 0,61 | 0,80 |
| kiváló     | 0,81 | 1,00 |

4. táblázat. A Cohes's kappa-statisztika eredményeinek értelmezési tartományai

Az előzetesen elvégzett, ún. pilot annotálás megmutatta, hogy az annotátorok a Kappa-érték szerint, összesítve 0,489, azaz a Kappa-sávok alapján közepes szintű átlagos egyetértéssel dolgoztak. Az annotáció részletes vizsgálata alapján kiderült, hogy a közepes szintű eredmény oka egyrészt technikai jellegű volt (az annotátorok a munka során nem azonos kijelölési megoldásokat alkalmaztak), másrészt összefüggést mutatott az annotálási feladat tartalmi vonatkozásaival is.

A technikai megoldások egységesítését, valamint az annotálási alapelvek pontosítását követően újraannotáltuk az egyetértésméréshez használt részkorpuszt, majd megismételtük az annotátorok közötti egyetértés mérését. Az újabb adatok alapján, a Kappa-sávok szerint az annotátorok közötti egyetértés a 0,597 értékre javult. Ebből az annotálás eredményessége a lokációt illetően 0,647, a relációkra vonatkozóan pedig 0,546 értéket mutatott. A Kappa-sávok alapján az előbbi jó, az utóbbi közepes szintű teljesítményt jelent.

A pilot annotálás megismétlése arra is lehetőséget adott, hogy megvizsgáljuk, vajon egy bizonyos idő elteltével az egyes annotátorok hogyan annotálnák ugyanazokat a nyelvi adatokat. Ezt a mérési módszert nevezi az angol nyelvtechnológiai irodalom *intra-annotator agreement rate*-nek (vö. Craggs–Wood 2004: 48). A korábbi és az újabb munka adatait összevetve az annotátorok átlagosan 0,741 értékű egyetértéssel dolgoztak, ami a Kappa-sávok alapján nagyon jó eredménynek tekinthető.

## 5 Az annotált korpusz elemzésének eredményei

A 610 tweetből összesen 289-et annotáltunk, ebben a 289 tweetben pedig összesen 383 lexikai szintű értékváltást vagy értékvesztést jelöltünk be. A maradék 321 tweetet azért nem annotáltuk, mert nem tartalmaztak egyetlen olyan elemet sem, amely értékváltást vagy értékvesztést mutatott volna (vö. Szabó et al. 2017). Ezekben az esetekben

ugyanis a vizsgált elem megőrizte lexikai szintű negatív értékét az aktuális kontextusban is.

A 383 annotált esetből összesen 75-ben jelöltünk értékvtást, és 308 esetben értékvesztést, tehát fokozó funkciót.

Megvizsgálva az értékvtás eseteit, a következő megállapításokat tehetjük: az összesen 75 értékvtás közül 40 (53,33%) a férfi szerzőségű tweetekben, 35 (46,66%) pedig a női szerzőségűekben található. Amennyiben ezt az előfordulást az összes férfi és női szerzőségű tweet számára vetítjük, a férfiak esetében 0,21%-os, a nők esetében 0,17%-os gyakoriságot kapunk. Úgy véljük, ez az eredmény arra mutat, hogy bár a nőknél valamelyest ritkábban fordul elő értékvtás, jelentősebb eltérésről nem beszélhetünk.

Az egyes elemek vizsgálata alapján ugyancsak az látszik, hogy a két nem között jelentős nyelvhasználati különbség ebben a tekintetben nincs. A leggyakoribb elemek mindkét nem esetében a *durva* és a *kemény*; a férfiaknál az előbbi 15, az utóbbi 10 esetben, a nőknél az előbbi 13, az utóbbi pedig 7 esetben jelenik meg értékvtó funkcióban. Érdeemi különbség csupán a *brutális* elemnél mutatkozik. Ez ugyanis a férfiak korpuszában három alkalommal megjelenik [5], míg a nőkében egyszer sem.

[5] Ezek a John Deere-ek forgácsolják szét a kivágott fák föld alatti részeit, **brutális** fűrőfejekkel.

Az eredmény összevág azzal a korábbi megfigyelésünkkel, hogy a *brutális* elem értékvtó funkcióban viszonylag gyakran szerepel technológiai tárgyú szövegekben, pl. *brutális választék, durván jó árak; várok egy nagyon-nagyon brutális iPhone 7-et* stb. (vö. Szabó et al. 2017).

Számos további elemet is annotáltunk még értékvtóként, azonban ezek lexémánként átlagosan mindössze egy vagy két alkalommal fordultak elő.

A 308 értékvesztési eset a következő megoszlási arányokat mutatja: Ahogyan azt a feldolgozás menete kapcsán is ismertettük (l. fentebb, 4.), azokban az esetekben, ahol a vizsgált elem fokozó funkciót töltött be, a fokozót és az általa módosított elemet további szemantikai–pragmatikai annotációval láttuk el, és a korpusz felhasználása során az így annotált sajtóságokat is lekérdeztük (vö. Szabó et al. 2017). Az adatokat a 3. táblázat mutatja be.

|                                       |         | A vizsgált elem<br>aktuális szemantikai tartalma |               |             |             | ÖSSZESEN: |
|---------------------------------------|---------|--|---------------|-------------|-------------|-----------|
|                                       |         | deszem   | negatív       | pozitív     | egyéb       |           |
| Az alaptag<br>szemantikai<br>tartalma | alappoz | 128  | 1             | 1           | 1           | 131       |
|                                       | alapneg | 86   | 1             | –           | 8           | 95        |
|                                       | alapsem | 42   | 37            | 3           | –           | 82        |
|                                       | egyéb   | –  | –             | –           | –           | –         |
| ÖSSZESEN:                             |         | 256:<br>83,11%                                   | 39:<br>12,66% | 4:<br>1,29% | 9:<br>2,92% | 308       |

5. táblázat. Az annotált értékvesztési esetek részletes statisztikai adatai.

Az eredmények alapján a következő megállapításokat tehetjük: A korpusz annotátorai az összesen 308 esetből 256 alkalommal vélték úgy, hogy a vizsgált elem teljesen elvesztette elsődleges negatív szemantikai tartalmát, és az általa módosított elem mellett pusztán fokozó szerepet töltött be. Ez az összes eset 83,11%-át tette ki.

Az elemek az értékvesztést illetően a pozitív és a negatív alaptagok mellett mutatkoztak a legegységesebbnek (vö. Szabó et al. 2017). Ezek a lexémák lexikai szintű pozitív vagy negatív értéket hordozó alaptagok módosítóiként ugyanis szinte kivétel nélkül elvesztették lexikai szintű szemantikai tartalmukat (l. pl. [3] fent). Ugyanakkor semleges alaptagok módosítóiként jelentős változatosságot mutattak a szemantikai–pragmatikai viselkedésük tekintetében. Az összesen 82 esetnek megközelítőleg csupán a felében (51,21%) jelöltek az annotátorok deszemantizáltságot, pl. (a módosított elemeket itt is aláhúzással jelöljük):

[6] Ja és Colin Farrell egy kibaszott nagy színészőriás

A többi esetben úgy ítélték, hogy a vizsgált elem a lexikai szintű értéket illetően nem mutat teljes deszemantizáltságot: vagy negatív (37: 45,12%), vagy – nagyon ritkán (3: 3,65%) – lexikai szintű negatív tartalmuk ellenére pozitív értéket ad a megnyilatkozáshoz, pl.

[7] Szörnyen meleg van még így az éjszaka közepén is.

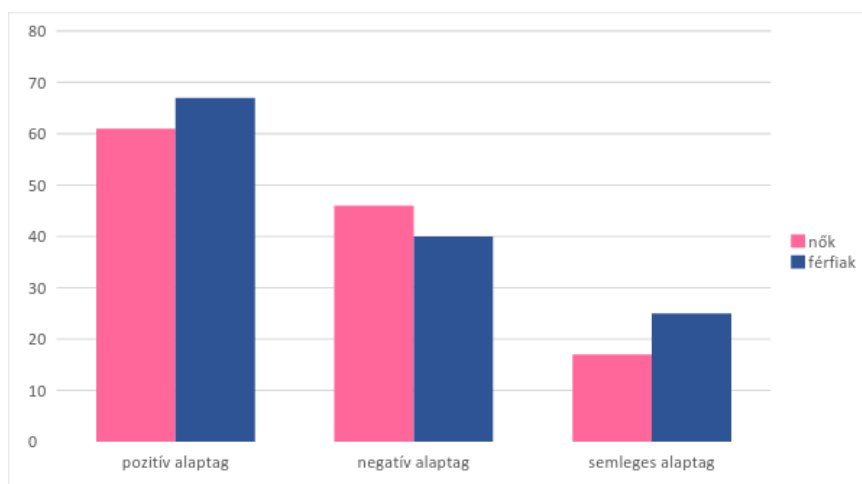
[8] Exportáltam pár képet erről a **brutálisan sűrű** és mozgalmas hétről, hogy legyen mit tolni instára, persze hogy minden az iMac-en maradt

Ezekben az esetekben tehát, annak ellenére, hogy az aktuális kontextusban a vizsgált elem fokozó funkciót tölt be, a szemantikai tartalma nem merült ki ebben a funkcióban. A vizsgálat fenti tanulságai figyelemre méltóak, hiszen számos, a negatív emotív szemantikai tartalmú fokozó elemekkel foglalkozó dolgozat amellet érvel, hogy azok szemantikailag rendre kiüresednek, elveszítik lexikai szintű negatív tartalmukat (vö. Szabó et al. 2017). Balogh (2009) például úgy gondolja, hogy amennyiben „az ilyen, másodlagos fokozóelemeket egy-egy megfelelő kulcsszóhoz kapcsoljuk, elveszítik elsődleges, azaz lexikális jelentésüket és átveszik a fokozó értelmű »nagyon«

adverbium jelentését.” Hozzá hasonlóan érvel Jing-Schmidt (2007), aki szerint például a FÉLELEM érzellemmel kapcsolatos negatív emóciókifejezések esetében, fokozó szerepben a FÉLELEM szemantikai tartalom metonimikusan a magas emotív intenzitásra redukálódik.

A nemek közötti eltérések vonatkozásában a negatív emotív fokozó elemek használatáról a következő megfigyeléseket tehetjük:

A negatív emotív fokozókat a férfiak kicsit gyakrabban alkalmazzák, mint a nők (férfiak: 160, nők: 148). Megvizsgáltuk, hogy a deszemantizált fokozó elemek mutatnak-e gyakorisági eltérést abban a tekintetben, hogy milyen szemantikai tartalmú alaptag mellett használják őket. Azt tapasztaltuk, hogy a nők és a férfiak nem azonos gyakorisággal alkalmazzák ezeket az elemeket deszemantizált fokozóként a pozitív, a negatív és a semleges alaptagok esetében. Az eredményeket a 2. ábrán mutatjuk be.



2. ábra. A deszemantizált fokozó elemek gyakorisága a két nemnél a különböző szemantikai tartalmú alaptagok mellett

Az ábra azt mutatja tehát, hogy bár mindkét nemnél hasonlóak az arányok az egyes szemantikai sajátságokkal rendelkező alaptagok melletti fokozók frekvenciáját illetően (a pozitív alaptagok módosítóiként a leggyakoribbak, utána a negatív, majd a semleges alaptagok következnek), egymáshoz viszonyítva nem azonos gyakorisággal alkalmazzák őket a különböző alaptagok mellett a férfi és a női szerzők. Amíg a férfiak a pozitív [9] és a semleges [10] alaptagok módosítójaként gyakrabban élnek a vizsgált elemmel, addig a nők a negatív alaptagok [11,12] melletti negatív emotív fokozókat preferálják jobban.

[9] Azért veszik sokan a DW-t, mert **rohadt jó** kis óra! Én arra szavaznék

[10] Hát így azért olyan **kurva sok** oroszlán nem futkosna.

[11] Ma **rettentően haszontalan** vagyok, nem mentem messzebb a konyhánál és oda is csak kv-t főzni

[12] az amúgy honnan jött? **iszonyú idegesítő** szó



Az eredmények alapján a következő megállapításokat tehetjük: Amíg a nők 31, addig a férfiak 24 különböző elemet használnak a korpuszban ebben a pozícióban. Emellett amíg a férfiak korpuszában az öt leggyakoribb (kurva[-ra], rohad[-u]), baromi[-ra], [ki-]baszott, iszonyatos[-an]/iszonyú[-an]) kiteszi az összes előfordulás 78%-át, addig a nők korpuszában az egyes elemek gyakorisága sokkal kiegyenlítettebb képet mutat; náluk az öt leggyakoribb elem (melyek ugyanazok, mint a férfiaknál) csupán az előfordulások 63%-át teszik ki. Érdeemes felhívni a figyelmet továbbá arra is, hogy a leggyakrabban használt negatív emotív fokozók mindkét nem esetében szitokszavak. Ami a nemek közötti megoszlási arányt illeti, a női korpusz összes adatának az 54,72%-a, a férfiak korpuszában pedig a 63,12%-a szitokszó. Úgy véljük, hogy ezek az adatok összhangban állnak Székely (2006), Jing-Schmidt (2007) és Balogh (2009) megállapításával, miszerint a vulgáris fokozó elemek használatában a nemek között eltérések mutatkoznak, ugyanis a fokozó értelmű kifejezések közül a durvább, vulgárisabb elemek inkább a férfiak nyelvhasználatát jellemzik (1. fentebb, 2. fejezet). Ugyanakkor az eredmények arra is rámutatnak, hogy a szitokszavak fokozóként való használati gyakorisága alapján – a vizsgált szövegtípusban legalábbis – nincs kimagasló eltérés a két nem között.

## 6 Összegzés

A dolgozatban egy az elméleti nyelvészetben kevésbé tárgyalt, és a nyelvtechnológia számára problematikus jelenséget, a negatív emotív szemantikai tartalmú elemeket vizsgáltuk részletesen, a női és a férfi nyelvhasználati különbségek szempontjából. A vizsgálat tárgyát azok az elemek képezték, amelyek lexikai szinten negatív emotív szemantikai tartalommal rendelkeznek ugyan, azonban fokozó funkciójában képesek azt részben vagy egészen elveszíteni (pl. *mocskos nagy*, *rohadt jól*), sőt bizonyos esetekben arra is, hogy negatív tartalmuk ellenére pozitív értékelést fejezzenek ki (pl. *brutális volt a hétfői koncert*) (vö. Szabó 2015). Az előbbit értékvesztésnek, az utóbbit értékváltásnak neveztük.

A dolgozatban bemutattuk a vizsgált nyelvi jelenséget, és az irodalom eddigi legfontosabb megállapításait, ismertettük a vizsgálati korpusz létrehozásának a menetét, módszerét és eszközeit, majd bemutattuk az elemzés részletes, kvalitatív és kvantitatív eredményeit, a korpuszból vett példákkal szemléltetve a legfontosabb kutatási tapasztalatokat. Emellett, mivel emóciószótárainkkal lehetőség nyílt a teljes, nyers Twitter-korpusz automatikus emócióelemzésére, és a kutatás alapproblémája a negatív emotív szemantikai tartalmú elemek nyelvhasználati sajátosságai voltak, elvégeztük a teljes korpusz szótáralapú automatikus érzelemelemzését is. Ezt követően az eredményeket a nemek közötti nyelvhasználati eltérések szempontjából elemeztük.

Az annotált korpusz elemzési eredményei alapján a következő legfontosabb megállapításokat tettük: A szólistánk alapján szűrt Twitter-korpusz összesen 383 lexikai szintű értékváltást vagy értékvesztést tartalmazott. A maradék 321 tweetet azért nem annotáltuk, mert nem tartalmaztak olyan elemet, amely értékváltást vagy értékvesztést mutatott volna (vö. Szabó et al. 2017). A 383 annotált esetből összesen 75 esetben jelöltünk értékváltást és 308 esetben értékvesztést, tehát fokozó funkciót.

Az értékváltás jelenségét illetően azt láttuk, hogy sem frekvenciájában, sem az egyes elemek használatában a nemek között nem mutatkozott jelentős eltérés. A leggyakoribb elemek mindkét nem esetében a *durva* és a *kemény* voltak. Érdeemi különbség egyedül

a *brutális* elemnél mutatkozott, ami a férfiak korpuszában három alkalommal jelent meg ebben a funkcióban, míg a nőkben egyszer sem.

Az értékvesztőként annotált elemek a pozitív és a negatív alaptagok mellett mutatták a legegységesebb szemantikai-pragmatikai viselkedést mindkét nem esetében (vö. Szabó et al. 2017). A lexikai szinten szentimentértékkel rendelkező alaptagok módosítóiként ugyanis szinte kivétel nélkül elveszítették lexikai szintű szemantikai tartalmukat. Ugyanakkor semleges alaptagok módosítóiként a vizsgált elemek jelentős változatosságot mutattak: deszemantizálódtak (51,21%), megőrizték negatív tartalmukat (45,12%), vagy – nagyon ritkán (3,65%) – lexikai szintű negatív tartalmuk ellenére pozitív értéket adtak a megnyilatkozáshoz.

Az értékvesztést illetően, a nemek közötti eltérések szempontjából a következőket állapítottuk meg: Egyrészt, a negatív emotív fokozókat a férfiak kicsivel gyakrabban alkalmazzák, mint a nők. Másrészt, a nők és a férfiak nem azonos gyakorisággal alkalmazzák ezeket az elemeket deszemantizált fokozóként, a pozitív, a negatív és a semleges alaptagok módosítóiként. Amíg ugyanis a férfiak a pozitív és a semleges alaptagok módosítójaként gyakrabban élnek a vizsgált elemekkel, addig a nők a negatív alaptagok mellettieket preferálják jobban. A vizsgálati tapasztalat véleményünk szerint arra mutat, hogy a férfiak szívesebben aknázzák ki e kifejezések pragmatikai értékét tartalmát pusztá hatáskeltési céllal, míg a nők inkább olyan kontextusokban élnek velük, ahol eleve a negatív értékítéletüket szeretnék kifejezni, és a vizsgált elemekkel ezt a negatív tartalmat kívánják erősíteni.

Végezetül megvizsgáltuk azt is, hogy mutatkozik-e eltérés a férfiak és a nők korpuszában annak tekintetében, hogy a két nem milyen konkrét fokozó elemeket használ. Megállapítottuk, hogy amíg a nők 31, addig a férfiak 24 különféle elemmel éltek ebben a pozícióban. Emellett amíg a férfiak korpuszában az öt leggyakoribb elem (*kurva[-ra]*, *rohadt[-u]*, *baromi[-ra]*, *[ki-]baszott*, *iszonyatos[-an]/iszonyú[-an]*) kiteszi az összes előfordulás 78%-át, addig a nők korpuszában az öt leggyakoribb elem (amely a két nem esetében egybeesik) csupán a 63%-át. Felhívtuk a figyelmet továbbá arra is, hogy a legnagyobb gyakorisággal rendelkező negatív emotív fokozók mindkét nem esetében szitokszavak. Ami a nemek közötti megoszlási arányt illeti, a női korpusz összes adatának 54,72%-a, a férfiak korpuszában pedig a 63,12%-a szitokszó. Úgy véltük, hogy ezek az adatok összhangban állnak Székely (2006), Jing-Schmidt (2007) és Balogh (2009) megállapításával, miszerint a fokozó értelmű kifejezések közül a durvább, vulgárisabb elemek inkább a férfiak nyelvhasználatát jellemzik. Ugyanakkor az eredmények arra is rámutatnak, hogy a szitokszavak fokozóként való használatát illetően a vizsgált szövegtípusban nincs kimagasló eltérés a két nem között.

A szótáralapú automatikus emócióelemzés segítségével feltártuk, hogy a 6 elemből álló szótár esetében az ÖRÖM, illetve MEGLEPŐDÉS emócióknál a két nem között szignifikáns eltérés volt: a MEGLEPŐDÉS emóció a férfiak tweetjeiben, míg az ÖRÖM a nők tweetjeiben jelenik meg többször. Bár a többi emóció esetében is megfigyelhető volt a két nem közötti eltérés, azok nem bizonyultak jelentős mértékűnek. A 8 elemből álló szótár elemzési eredményei alapján a DÜH, a FÉLELEM, a FESZÜLTség, az ÖRÖM és a VONZALOM emóciók a két csoport között statisztikailag szignifikáns különbséget mutattak: ezek az érzelmek a nők tweetjeiben jelentek meg nagyobb mértékben.

## 7 Kitekintés

Ahogy az a dolgozat bevezetőjében (l. 2. fejezet) is hangsúlyoztuk, a lexikai szintű értékértékelés és az értékvesztés mind elméleti, mind alkalmazott nyelvészeti, különösen nyelvtechnológiai szempontból figyelemre méltó. Kimerítő elméleti nyelvészeti leírása még nem történt meg, hatékony nyelvtechnológiai kezelése pedig egyelőre még nem megoldott (vö. pl. Andor 2011; Kugler 2014; Szabó 2015; Szabó et al. 2017; Drávucz–Szabó–Vincze 2017). Ezekre az elemekre ugyanis a szótáralapú automatikus szentiment- és emócióelemzés – a lexikai szintű értékük miatt – téves elemzési eredményt ad.

Amint arra a jelen vizsgálat eredményei is rámutattak, a semleges alaptagok módosítóiként a vizsgált elemek jelentős szemantikai változatosságot mutatnak, és csupán az eseteknek valamivel több mint a felében deszemantizálódnak. Az eredmény figyelemre méltó, hiszen számos, a negatív emotív szemantikai tartalmú fokozó elemekkel foglalkozó dolgozat amellett érvel, hogy azok szemantikailag rendre kiüresednek, elveszítik lexikai szintű negatív tartalmukat (vö. Szabó et al. 2017). Balogh (2009) például úgy gondolja, hogy amennyiben „az ilyen, másodlagos fokozóelemeket egy-egy megfelelő kulcsszóhoz kapcsoljuk, elveszítik elsődleges, azaz lexikális jelentésüket és átveszik a fokozó értelmű »nagyon« adverbium jelentését”. Jing-Schmidt (2007: 434) szerint ezek az elemek az elsődleges szemantikai tartalmuktól teljesen eltávolodnak, és az a pusztán intenzitásra redukálódnak. Ezzel szemben azonban a kapott elemzési eredmények e fokozó elemek sokkal árnyaltabb nyelvi viselkedésére mutatnak rá.

A dolgozatban bemutatott vizsgálat további fontos elméleti hozadékának tekinthető az, hogy kvantitatív és kvalitatív hasonlóságokat, illetve eltéréseket tárt fel a férfi és a női korpusz között a kutatási kérdés tekintetében. Amint azt a szakirodalmi áttekintésben (l. fentebb, 2) megmutattuk, a vonatkozó megállapítások száma csekély, és azok sem nagyobb mennyiségű valós nyelvi adat elemzésén alapulnak. Székely (2006) és Balogh (2009) véleménye szerint a nők gyakrabban élnek az ún. „illogikus fokozó értelmű” kapcsolatokkal. Emellett Székely (2006) és Jing-Schmidt (2007: 432) véleménye szerint a fokozó értelmű kifejezések közül a durvább, vulgárisabb elemek inkább a férfiak nyelvhasználatát jellemzik. A jelen kutatásban tett megállapításaink amellett, hogy számszerű adatok segítségével, valós nyelvi példák alapján pontosabb képet adnak a jelenségről, a szakirodalmi megállapítások egy részét is cáfolják.

A nyelvtechnológiai kezelés problémáit illetően a következő megállapításokat tehetjük: Tekintettel arra, hogy a vizsgált elemek fokozó funkcióban, negatív és pozitív alaptagok módosítóiként szinte kivétel nélkül elveszítették lexikai szintű szemantikai tartalmukat, helytálló automatikus kezelésüket egyszerű pozíciós szabály bevezetésével megvalósíthatónak tartjuk a szótár- és szabályalapú szentimentelemzés során. Ezt a megoldást a jövőben tesztelni kívánjuk úgy, hogy egy kézzel annotált szentimentkorpusz értékeit összevetjük a fenti szabályt nem alkalmazó, valamint az azt alkalmazó automatikus elemzés eredményeivel. Az elmondottakon túl azt is megállapítottuk, hogy a semleges alaptagok melletti, valamint a szentimentkifejezés funkciójában álló negatív emotív elemek ilyen jellegű kezelése nem lehetséges. A jövőben azt is meg szeretnénk vizsgálni, hogy milyen egyéb nyelvi sajátosságok (pl. az emojik és az emotikonok figyelembe vétele, vö. Carvalho et al. 2015) tudnák az értékértékelés automatikus kezelését hatékonyan támogatni.



## Irodalom

- Andor J. 2011. De durva ez a téma! – Megfigyelések a melléknévi polaritásváltásról. *Hungarológiai Évkönyv*, 12: 33–42.
- Andor J. 2003. Functional Studies in the Polarity and Gradation of Amplifier Adjectives and Adverbs in English. In: Andor, J., Horváth J., Nikolov M. (szerk.) *Studies in English Theoretical and Applied Linguistics*. Pécs: Lingua Franca Csoport. 43–59.
- Balogh Péter 2009. Gender-markerek a nyelvben. *WEBFU [Wiener elektronische Beiträge des Instituts für Finno-Ugristik]*. Elérhető: <http://webfu.univie.ac.at/texte/3Balogh.pdf>. Letöltve: 2017. október 16.
- Carvalho, P., Sarmiento L., Silva M. J., de Oliveira E. 2015. Clues for Detecting Irony in User-Generated Contents: Oh...!! It's "so easy";-). In: *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*. Hong Kong: ACM. 53–56.
- Cohen, J. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20: 37–46.
- Craggs, R., Wood, M. M. 2004. A two-dimensional annotation scheme for emotion in dialogue. In: *Papers from the 2004 AAAI Spring Symposium. Exploring Attitude and Affect in Text: Theories and Applications*. Menlo Park: AAAI Press. 44–49.
- Drávucz, F., Szabó M. K., Vincze V. 2017. Szentiment- és emóciósótárak eredményességének mérése emóció- és szentimentkorpuszokon. In: Vincze, V. (szerk.) *XIII. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY 2017*. Szeged: JATEPress. 228–239.
- Ekman, P., Friesen, W. V. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1: 49–98.
- Jing-Schmidt, Z. 2007. Negativity bias in language: A cognitive-affective model of emotive intensifiers. *Cognitive Linguistics*, 18(3): 417–443.
- Kugler N. 2014. A nyelvi polaritás kifejezésének egy mintázata, avagy milyen a félelmetesen jó? *Magyar Nyelvőr*, 138(2): 129–139.
- Szabó M. K. 2015. A polaritásváltás és -változás kezelési lehetőségei a szentimentelemzésben. In: Keresztes Gábor (szerk.) *Tavaszi Szél konferenciakötet*. III. kötet. Eger–Budapest: Líceum Kiadó –Doktoranduszok Országos Szövetsége. 629–643.
- Szabó, M. K. 2017. The usage of elements with emotive semantic content from a gender point of view. Kézirat.
- Szabó, M. K., Morvay, G. 2015. Emócióelemzés magyar nyelvű szövegeken. In: Gecső, T., Sárdi, Cs. (szerk.) *Nyelv, kultúra, társadalom. Segédkönyvek a nyelvészet tanulmányozásához* 177. Budapest: Tinta. 286–292.
- Szabó, M. K., Vincze, V. 2015. Egy magyar nyelvű szentimentkorpusz létrehozásának tapasztalatai. In: Tanács, A., Varga V., Vincze V. (szerk.) *XI. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY2015*. Szeged: JATEPress. 219–226.
- Szabó, M. K., Vincze V., Morvay G. 2016. Magyar nyelvű szövegek emócióelemzésének elméleti nyelvészeti és nyelvtechnológiai problémái. In: Reményi, A. A., Sárdi Cs., Tóth Zs. (szerk.) *Távlatok a mai magyar alkalmazott nyelvészetben*. Budapest: Tinta. 282–292.
- Szabó, M. K., Nyíri Zs., Morvay G., Lázár B. 2017. A szentimentérték módosulásának vizsgálata szemantikai–pragmatikai szempontból annotált korpuszon. In: Vincze, V. (szerk.) *XIII. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY 2017*. Szeged: JATEPress. 251–262.
- Székely, G. 2007. *Egy sajátos nyelvi jelenség, a fokozás. Segédkönyvek a nyelvészet tanulmányozásához* 66. Budapest: Tinta.
- Tolcsvai Nagy, G. 1988. A mai magyar nyelv normarendszerének egy jelentős változásáról az „ifjúsági nyelv” kapcsán. *Magyar Nyelvőr*, 112(4): 398–406.
- Wierzbicka, A. 2002. Australian cultural scripts – bloody revisited. *Journal of Pragmatics*, 34(9): 1167–1209.

## Negatív emotív szemantikai tartalmú fokozó elemek vizsgálata az angol–orosz és orosz–angol fordítás tükrében<sup>i</sup>

Szabó Martina Katalin<sup>1,2,3</sup>, Nyíri Zsófi<sup>1</sup>, Lázár Bernadett<sup>1</sup>

<sup>1</sup>PrecognoX Informatikai Kft.

<sup>2</sup>Szegedi Tudományegyetem, Bölcsészettudományi Kar, Szláv Intézet

<sup>3</sup>„Lendület” – Research Center for Educational and Network Studies (RECENS)

szabo.martina@lit.u-szeged.hu

{mszabo, zsnnyiri, blazar}@precognoX.de

**Kivonat:** A dolgozatban a negatív emotív szemantikai tartalmú fokozó elemeket vizsgáljuk, az angolról oroszra és oroszról angolra történő fordítás szempontjából. Negatív emotív tartalmú fokozóknak nevezzük azokat az elemeket, amelyek lexikai szinten negatív emotív szemantikai tartalommal rendelkeznek ugyan, de különböző értékelő kifejezések mellett fokozó funkciót tölthetnek be (pl. mocskos nagy, durván jó). A jelenség mind elméleti, mind alkalmazott nyelvészeti szempontból figyelemre méltó. Pontos és kimerítő elméleti leírása még nem történt meg, hatékony nyelvtechnológiai kezelése pedig egyelőre nem megoldott. A jelen vizsgálathoz Az orosz nyelv nemzeti korpuszának (НКРЯ) angol–orosz párhuzamos alkorpuszát használtuk. Kiválasztottunk nyelvenként 2-2 olyan, lexikai szinten negatív emotív elemet, amely képes fokozó funkciót betölteni, és gyakorisága viszonylag magas (*чертовски, ужасно, damn(ed), awfully*), majd lekérdeztük a korpuszból a kiválasztott elemeket. Az elemzés során megvizsgáltunk és rendszereztünk minden elemet külön-külön, a fordítási irányok figyelembevételével, kvalitatív és kvantitatív szempontból egyaránt. Végül az eredményeket összevetettük egymással, valamint a kutatás kezdeti hipotéziseivel.

### 1 A dolgozat tárgya, a témaválasztás indoklása

A dolgozatban a negatív emotív szemantikai tartalmú fokozó elemeket vizsgáljuk, az angolról oroszra és oroszról angolra történő fordítás szempontjából.

---

<sup>i</sup> A jelen kutatást Az Emberi Erőforrások Minisztériuma Új Nemzeti Kiválóság Programja, az MTA Társadalomtudományi Kutatóközpont „Lendület” RECENS Kutatócsoportja támogatta, valamint a European Research Council (ERC) az Európai Unió „Horizon 2020” nevű, 648693-es számú kutatási és fejlesztési programja keretében. Köszönjük továbbá a dolgozat névtelen bírálójának, hogy az észrevételeivel, javaslataival és hivatkozásaival értékes hozzájárulást tett a munkánkhoz.

Negatív emotív tartalmú fokozóknak nevezzük azokat az elemeket, amelyek lexikai szinten negatív emotív szemantikai tartalommal rendelkeznek ugyan, de különböző értékkelő kifejezések mellett fokozó funkciót tölthetnek be (A példákban a vizsgált elemet dőlt szedéssel emeljük ki):

- a. *kegyetlen* rosszul (főz)
- b. *mocskos* nagy (halat fogott)
- c. *durván* jó (árak)

A vizsgált lexémák fokozó elemek pozíciójában részben vagy egészben elveszítik elsődleges, azaz lexikai szintű szemantikai tartalmukat (vö. Szabó et al. 2017). Említsük meg, hogy a jelenség nem feltétlenül fokozó elem pozíciójában képzelhető el: a negatív emotív elemek olykor önmagukban, értékkelő elem funkciójában is előfordulnak, pozitív értékelés kifejezőiként, pl.

- a. *brutális* választék
- b. *durván* néz ki a hajad

A jelen dolgozatban ugyanakkor csupán azokat az elemeket tesszük vizsgálat tárgyává, amelyek fokozó funkciót töltenek be.

A negatív emotív szemantikai tartalmú elemek problémája mind elméleti, mind alkalmazott nyelvészeti szempontból figyelemre méltó. Pontos és kimerítő elméleti leírása még nem történt meg, ennek ellenére a jelenséggel kevés dolgozat foglalkozik (vö. pl. Wierzbicka 2002; Jing-Schmidt 2007; Andor 2011; Kugler 2014; Szabó 2015a). Az alkalmazott kutatások oldaláról emeljük ki a hatékony nyelvtechnológiai kezelés problémáját. A szentiment- és az emócióelemzés a nyelvtechnológia részfeladata: az előbbi a nyelvi értékkelő tartalmak automatikus kivonatolására, az utóbbi a szövegekben megbúvó érzelmek automatikus detektálására törekszik. Mindkét nyelvtechnológiai feladat számára kihívást jelent a negatív emotív elemek hatékony kezelése. Ezekben a feladatokban ugyanis az elemzés alapját gyakorta szentiment- és emóciószótárak képezik, amelyben a lexémákat azok elsődleges szemantikai tartalma alapján szerepeltetik a szótár fejlesztői a különböző listákon: a szentimentszótár esetében a pozitív és a negatív, az emóciószótár esetében pedig valamely konkrét érzelmek listáján. Ugyanakkor, amint az a negatív emotív fokozó elemek példájából is látszik, ezek az elemek nem minden esetben utalnak a lexikai szintű szemantikai tartalmuknak megfelelő értékelésre, illetve érzelmekre. Szótáralapú automatikus elemzésük tehát téves következtetéseknek engedhet teret (vö. például Andor 2011; Kugler 2014; Szabó 2015a, 2015b; Drávucz–Szabó–Vincze 2017, Szabó et al. 2017).

Tekintettel arra, hogy az általunk vizsgált elemek lexikai szintű szemantikai tartalma emotív értékkel rendelkezik, és ez az affektivitás magas intenzitású, azt feltételeztük, hogy a negatív emotív fokozó elemek vizsgálatra érdemesek a fordítási gyakorlat szempontjából is. A jelen dolgozatban ezért a jelenséget a fordítási műveletek szempontjából vizsgáljuk meg. Az elvégzett kutatás hiánypótló: nincs tudomásunk más olyan dolgozatról, amely a negatív emotív fokozó elemeket a fordítás szempontjából vizsgálná, annak ellenére, hogy Jing-Schmidt (2007) és Andor (2011) alapján a jelenség valószínűleg nyelvfüggetlen sajátság.

## 2 A vizsgálat módszere és eszköze

A jelen vizsgálathoz Az orosz nyelv nemzeti korpuszának (*Национальный корпус русского языка, НКРЯ*) angol–orosz párhuzamos alkorpuszát használtuk. Az alkorpusz az orosz–angol fordítási irányban összesen 6 591 353, az angol–orosz fordítási irányban 18 086 285 szövegszót tartalmaz.

A korpusz lekérdezéséhez a következő módszert alkalmaztuk: kiválasztottunk nyelvenként 2-2 olyan, lexikai szinten negatív emotív elemet, amely képes fokozó funkciót betölteni, és gyakorisága viszonylag magas. A választásunk az orosz nyelv esetében a *чертовски* ('szörnyen, pokolian, átkozottul') és az *ужасно* ('borzasztóan, szörnyen'), az angol esetében pedig a *damn(ed)* ('átkozott, átkozottul, rohadt') és az *awfully* ('borzasztóan, szörnyen, rémesen') elemekre esett.<sup>ii</sup> Ezt követően lekérdeztük a korpuszból a kiválasztott elemeket, azzal a beállított feltétellel, hogy a keresett elemet a korpuszban melléknév vagy határozószó kövesse közvetlenül. Erre a megkötésre azért volt szükség, mert ezáltal a számunkra a jelen esetben nem releváns találatoknak a nagy részét kizárhattuk a találati listából (pl. azt az esetet, ha a vizsgált elem közvetlenül egy főnevet minősít, és nem fokozó funkciót tölt be). A kitétel tehát növelte annak az esélyét, hogy az eredményül kapott fordítási egységekben a vizsgált elem egy jelző vagy egy határozó módosítójának a funkcióját tölti be.

A lekérdezéssel kapott találatok két típusú fordítási egységet tartalmaztak: egyrészt azokat, amelyekben a vizsgált elem a forrásnyelvi oldalon jelent meg, tehát az eredeti szövegben, másrészt olyanokat, ahol a célnyelvi oldalon, azaz fordítási megoldásként. A vizsgálati korpusz feldolgozási módja lehetővé tette, hogy e két típust egymástól elkülönítve vizsgálhassuk.

Az elemzés során megvizsgáltunk és rendszereztük minden lekérdezett elemet külön-külön, a fordítási irányok figyelembevételével, kvalitatív és kvantitatív szempontból egyaránt. Az elemzéshez Klaudy Kinga (1997) osztályozását vettük alapul.

A vizsgált fordítási egység körébe művenként maximálisan tíz adatot vontunk be, ezzel is törekedve a kapott eredmények reprezentativitására.

## 3 A vizsgálat fő hipotézisei

A vizsgálat eredményeit illetően a következő legfontosabb hipotéziseink voltak: Egyrészt, ha a vizsgált elem a forrásszövegben jelenik meg, a fordítók – annak lexikai szintű affektív tartalma, stiláris értéke miatt – gyakran kerülnek a célnyelvi ekvivalens használatát. Ugyanakkor a keletkező veszteséget különböző kompenzációs megoldásokkal igyekeznek pótolni a célnyelvi szövegben. Másrészt a vizsgált elem – a fentebb írt affektív tartalmi ok miatt – a célnyelvi oldalon csak akkor jelenik meg, ha hasonló stiláris értékkel rendelkező elemet találunk a forrásnyelvi oldalon is. Mindemellett mindkét tendenciát a *damn(-ed)* elem esetében feltételeztük a legjellemzőbbnek, annak lexikai szemantikai tartalma okán.

---

<sup>ii</sup> Mind az orosz, mind az angol nyelvű elemek eltérő stílusréteghez tartoznak, ami nyilvánvalóan hatással van a vizsgálat eredményeire. Köszönjük a dolgozat névtelen bírálójának, hogy felhívta erre a figyelmünket. (A probléma kapcsán l. még lentebb, 5.3).

## 4 A vizsgálat eredményei

### 4.1 A *damn(-ed)* és az *awfully* elemek vizsgálati eredményei

A *damn(-ed)* lekérdezésével 269 fordítási egységet, míg az *awfully* lekérdezésével 307 fordítási egységet nyertünk ki a korpuszból. A találatokból egyik elem esetében sem volt minden adat a kutatási célunknak megfelelő. Például azokban az esetekben, ahol a vizsgált elem nem az azt követő elem módosítójának a szerepét töltötte be, pl. *damned white trash, it has been troubling me awfully ever since* stb. Emellett a korpusz tartalmazott néhány párhuzamosítási hibát is, amely számunkra téves találatokat eredményezett. Tovább szűkítette a vizsgált elemek körét az az alapelvünk, miszerint művenként csupán maximálisan tíz egységet vettünk figyelembe, reprezentativitási céllal. A szűrést követően a *damn(-ed)* esetében 141, az *awfully* esetében 236 adat maradt. Ezek fordítási irányonkénti megoszlását az alábbi táblázatban mutatjuk be:

| Vizsgált elem    | Fordítási irány |       | Összesen: |
|------------------|-----------------|-------|-----------|
|                  | A → O           | O → A |           |
| <i>damn(-ed)</i> | 129             | 12    | 141       |
| <i>awfully</i>   | 142             | 94    | 236       |

1. táblázat. A *damn(-ed)* és az *awfully* gyakorisági megoszlása

A kapott eredmény összhangban áll kezdeti hipotézisünkkel, miszerint a vizsgált elem – annak stílárius értéke okán – a célnyelvi oldalon csak akkor jelenik meg, ha hasonló stílárius értékkel rendelkező elemet találunk a forrásnyelvi oldalon is, és ezt a tendenciát különösen jellemzőnek feltételeztük a *damn(-ed)* esetében. A gyakorisági megoszlás alapján azt látjuk, hogy bár a *damn(-ed)* és az *awfully* gyakorisága az eredeti angol fordítási egységek oldalán nem mutat kimagasló eltérést (129, 142), a fordított irányban jelentős eltérés mutatkozik: amíg az *awfully* fordítási megoldásként is gyakori, addig a *damn(-ed)* elemmel nagyon ritkán él a fordító az oroszról angolra történő átültetés során.

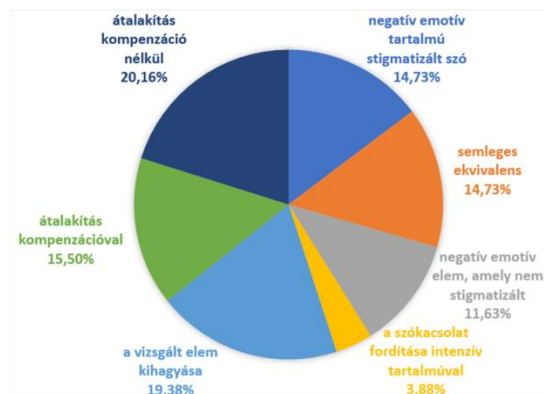
Először azokat a fordítási egységeket vesszük górcső alá, amelyekben a vizsgált elemek a forrásnyelvi oldalon találhatóak.

A *damn(-ed)* fordítási megoldásait az alábbiakban, példák segítségével mutatjuk be.

- [1] Negatív emotív tartalmú, stílárius értékét tekintve hasonló fokozó szóval, pl. *Damned old* → *проклятый старый*
- [2] Semlegesre cserélte, pl. *Damn hard* → *очень трудно*
- [3] Negatív emotív elemre cserélte, amely azonban nem vulgáris elem, pl. *Damn hot* → *страшная жара*
- [4] A vizsgált elem és az általa módosított elem kapcsolatát együtt egy nagyon intenzív pozitív elemre cserélte, pl. *She was a damned nice woman too* → *К тому же как женщина она была замечательна*
- [5] Kihagyás, pl. *You have a damn good sense of humor* → *Оказывается, у тебя есть чувство юмора*
- [6] Teljes átalakítás esetén:

- a) törekszik az elvesző pragmatikai érték pótlására: a szitokszó stílusértékéhez hasonló eszköz (a), más eszköz (pl. irónia, frazéma, írásjel, szlen b-c):
- i. *You knew damn well* → *Но вы же, дьявол вас побери, прекрасно*
  - ii. *Damn poor order* → *хороший порядок!*
  - iii. *damn right* → *что верно, то верно*
- b) nincs kompenzációs törekvés, pl.
- i. *That was a damn good performance you put on out there tonight* → *ты сегодня был в ударе*

A fentebbi fordítási megoldások gyakorisági megoszlását az 1. ábra mutatja be:

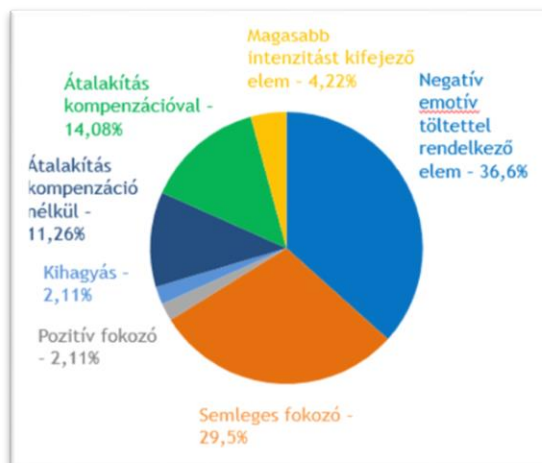


1. ábra. A *damn(-ed)* fordítási megoldásainak gyakorisági megoszlása

Az ábra adatai alapján a következő legfontosabb megállapításokat tehetjük: Tartalmi és stilisztikai szempontból egyenértékű ekvivalenst csupán az esetek 14,73%-ában alkalmaz a fordító. Ilyenkor a vizsgált elemet leggyakrabban a *чертовски* ekvivalenssel ülteti a célnyelvre. Kevés veszteség éri a szöveget, ha a fordító – bár átalakítja a forrásnyelvi kifejezést – kompenzációt alkalmaz (15,5%), vagy negatív emotív tartalmú fokozóval fordítja a forrásnyelvi elemet (11,63%). A fordítási egységek 58,14%-ában azonban komolyabb veszteség éri a szöveget a fordítás során. Ezek azok az esetek, amikor a fordító semleges fokozót alkalmaz (14,73%), kihagyja a vizsgált forrásnyelvi elemet (19,38%), átalakítást végez, azonban azt kompenzáció nélkül teszi (20,16%), vagy pedig a teljes szókapcsolatot egy magasabb intenzitást kifejező, pozitív polaritású elemre cseréli (3,88%). Ezeknél a megoldásoknál a forrásnyelvi stílári és pragmatikai érték nyilvánvalóan komoly veszteséget szenved.

Az *awfully* esetében hasonló fordítási megoldásokkal találkozunk, mint amelyeket a *damn(-ed)* esetében találtunk. Ezeket ezért a dolgozat keretei miatt egyenként nem prezentáljuk. Két eltérést kell mindössze megemlítenünk: Egyrészt az *awfully* esetében – ellentétben a *damn(-ed)* fordítási megoldásaival – nem találtunk olyan esetet, ahol a fordító vulgáris kifejezéssel fordított volna. Másrészt – ugyancsak a *damn(-ed)* fordítási megoldásaival ellentétben – az *awfully* esetében előfordult az a megoldás is, hogy a fordító a vizsgált fokozó elemet pozitív polaritású elemre cserélte, pl. *awfully good* →

*верно хороши*. Az *awfully* fordítási megoldásainak gyakorisági megoszlását az 2. ábra mutatja be:



2. ábra. Az *awfully* fordítási megoldásainak gyakorisági megoszlása

Az ábra adatai alapján a következő legfontosabb megállapításokat tehetjük: Tartalmi és stilisztikai szempontból egyenértékű ekvivalenst jelentősen gyakrabban alkalmaz a fordító a *damn(-ed)* szónál: 36,6%-ban. Ilyenkora vizsgált elemet leggyakrabban az *ужасно* ekvivalenssel ülteti a célnyelvre. Kevés veszteség éri a szöveget, ha a fordító, bár átalakítja a forrásnyelvi kifejezést, kompenzációt alkalmaz (14,08%). A fordítási egységek 49,32%-ában éri a szöveget komolyabb veszteség. Ezek azok az esetek, amikor a fordító semleges (29,5%) vagy pozitív fokozót (2,11%) alkalmaz, kihagyja a vizsgált forrásnyelvi elemet (2,11%), átalakítást végez, azonban azt kompenzáció nélkül teszi (11,26%), vagy pedig a teljes szókapcsolatot egy magasabb intenzitást kifejező, pozitív polaritású elemre cseréli (4,22%).

Összevetve a két vizsgált vizsgált elem fordítási megoldásainak gyakorisági megoszlását a következőket állapíthatjuk meg: amíg a *damn(-ed)* esetében tartalmi és stilisztikai szempontból egyenértékű ekvivalenst csupán 14,73%-ban alkalmazott a fordító, addig ugyanez az arány az *awfully* esetében jóval magasabb, 36,6% volt. Ez az eredmény összhangban áll kezdeti hipotézisünkkel, miszerint a *damn(-ed)* esetében – annak vulgáris tartalma okán – ritkán él a fordító a célnyelvi ekvivalenssel. Ezzel összefüggésben azt is láttuk, hogy amíg az *awfully* esetében 49,32%-ban érte komolyabb veszteség a szöveget, addig ez a *damn(-ed)* esetében jóval gyakrabban, 58,14%-ban következett be. Az elmondottakon túl érdemes felhívni a figyelmet arra, hogy amíg az *awfully* esetében a semleges fokozóval történő fordítás igen gyakori megoldás (29,5%), *damn(-ed)* esetében jóval ritkább (14,73%). Ugyanakkor a kihagyás művelete jelentősen gyakrabban mutatkozik meg a *damn(-ed)* ekvivalensnél (19,38%), amely az *awfully* fokozónál alig fordul elő (2,11%).

A továbbiakban azokat az eseteket elemezzük, amelyekben a vizsgált elemek a fordítási egység célnyelvi oldalán, azaz fordítási megoldásként jelentek meg.

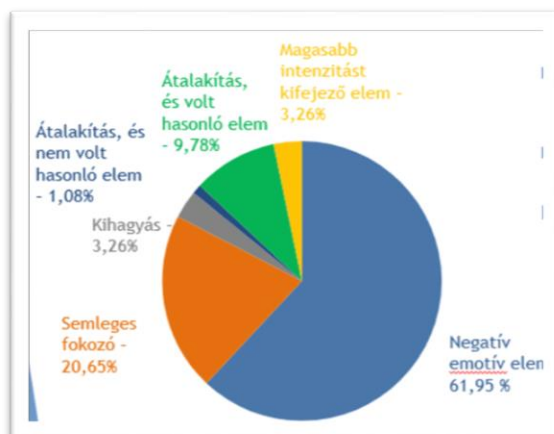
Amint azt a fejezet elején ismertettük, a *damn(-ed)* elem a célnyelvi oldalon, tehát fordítási megoldásként mindössze 12 alkalommal jelent meg. Mivel ez az előfordulás

rendkívül csekély, azokat egyenként nem elemezzük és bővebben nem részletezzük. Az *awfully* szóról a következő megállapításokat tehetjük: Összesen 94 olyan fordítási egységet találtunk a korpuszban, amelyben az *awfully* fordítási megoldásként, azaz a célnyelvi oldalon jelent meg a fordítási egységben. A fordítási egységek egyenkénti vizsgálata alapján megállapítható, hogy hasonló megoldásokat alkalmaznak a fordítók, mint amelyeket az angol–oroszlány irány esetében láttunk. Egyetlen eltérés mutatkozik csupán, mégpedig az, hogy amíg az *awfully* intenzifikálót néhány esetben pozitív emotív fokozó elemmel fordították oroszra (l. fentebb), addig pozitív emotív fokozót sohasem fordítottak az *awfully* ekvivalenssel a fordítók. A fordítási megoldásokat az alábbiakban, példák segítségével prezentáljuk:

- [1] A forrásnyelvben negatív emotív tartalmú fokozót találunk, pl.  
*Ужасно (...) тоненькая* → *awfully thin*
- [2] A forrásnyelvben semleges fokozó elem van, pl.  
*Сыро (...) очень* → *awfully damp*
- [3] A forrásnyelvi szerkezetből hiányzik a fokozó elem, pl.  
*Должно быть далеко* → *It must be awfully far away*
- [4] Teljes átalakítás:
- A fordító, bár teljes átalakítást végez, a forrásnyelvi szövegrész stilisztikailag hasonló elemet is tartalmazott, pl. *До ужаса ловко сперли* – *It was an awfully deft snatch.*
  - A fordító teljes átalakítást végez, azonban a forrásnyelvi szövegrész nem tartalmazott stilisztikailag hasonló elemet, pl. *Видишь ли: я об этом, как ни глуп* – *My dear boy, I'm awfully stupid about that.*
- [4] A fokozó és az általa módosított elem forrásnyelvi megfelelőjeként egy magasabb intenzitást kifejező elemet találunk, pl. *Что у меня в темноте презлупное лицо* – *That my face in the dark looked awfully silly.*

A fentebbi forrásnyelvi megfelelők gyakorisági megoszlását a 3. ábra mutatja be:





3. ábra. Az *awfully* mint fordítási megoldás forrásnyelvi megfelelőinek gyakorisági megoszlása

A kapott értékek alapján a következő legfontosabb megállapításokat tehetjük: A forrásnyelvi szövegrész az esetek többségében (61,95%) tartalmi és stilisztikai szempontból egyenértékű ekvivalenst tartalmaz akkor, ha az *awfully* megjelenik a célnyelvben. Ez az eredmény ugyancsak összhangban áll egyik kezdeti hipotézisünkkel, mely szerint a vizsgált elemek alapvetően akkor jelennek meg a célnyelvi oldalon, ha hasonló stílári értékkel rendelkező elemet találunk a forrásnyelvi oldalon is. Pragmatikai-stilisztikai szempontból nem volt jelentős eltérés a forrás- és a célnyelvi oldalon akkor sem, ha a fordító átalakítás során kompenzációs céllal alkalmazta a vizsgált elemet (9,78%). Mindössze 28,27%-ban talákoztunk olyan fordítási egységgel, ahol a forrásnyelvi szövegrész nem hordozott hasonló stílári értékű elemet vagy elemeket.

#### 4.2 A *чертовски* és az *ужасно* elemek vizsgálati eredményei

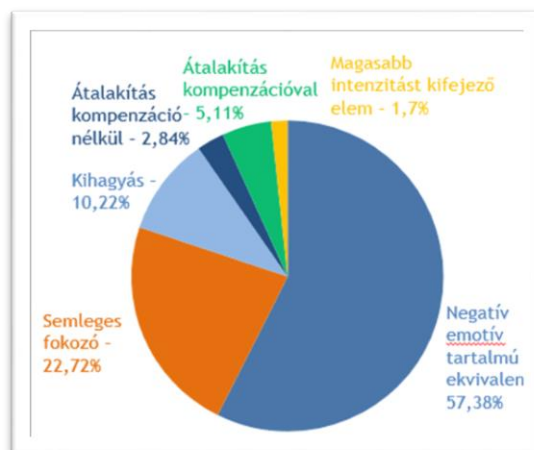
Az *ужасно* esetében összesen 480 fordítási egységet, míg a *чертовски* esetében 82 fordítási egységet kaptunk a korpusz lekérdezésével, azonban itt is figyelembe kellett vennünk, hogy nem minden adat felelt meg a kutatási céljainknak, illetve művenként maximálisan 10 egységet vontunk be a vizsgálat körébe. A szűrést követően, a *чертовски* esetében 81, az *ужасно* esetében 429 fordítási egység maradt. A vizsgált elemek fordítási irányonkénti megoszlását az alábbi táblázat mutatja be:

| Vizsgált elem    | Fordítási irány |       | Összesen: |
|------------------|-----------------|-------|-----------|
|                  | A → O           | O → A |           |
| <i>чертовски</i> | 78              | 3     | 81        |
| <i>ужасно</i>    | 253             | 176   | 429       |

**2. táblázat.** *Ачертовски* és az *ужасно* gyakorisági megoszlása fordítási irányonként és összesítve

A fentebbi eredmények közül számunkra a legmeglepőbb az volt, hogy amíg az *ужасно* mind a forrásnyelvi, mind pedig a célnyelvi szövegekben gyakori (253 és 176 előfordulás), addig a *чертовски* kifejezetten ritkán, mindösszesen három alkalommal fordul elő a forrásnyelvből, ugyanakkor a célnyelvi oldalon magas a gyakorisága. Bár mindkét vizsgált kifejezésnél elmondható, hogy azok a fordítási egységek forrásnyelvi oldalán szerepelnek ritkábban, az eltérés a *чертовски* esetében igazán szembeötlő. A tapasztalat véleményünk szerint a fordított és nem fordított szövegek eltérésének problémájára mutat rá, hiszen jelen esetben a fordítások során olyan lexéma kerül be gyakran a fordított orosz szövegekbe, amely az eredeti orosz szövegeknek egyáltalán nem jellemző eleme.

Először azokat a fordítási megoldásokat részletezzük, amelyekben a vizsgált elemek a forrásnyelvi oldalon találhatóak. Mivel a *чертовски* elem forrásnyelvi előfordulása ritka, azokat nem részletezzük. Az *ужасно* esetében a fordítási megoldások alapvetően egybeesnek azokkal, amelyeket az *awfully* esetében láttunk az angol–orosz fordítási irányban. Csupán egyetlen eltérést láttunk, azt, hogy amíg az *awfully* elemet néhány alkalommal pozitív szemantikai tartalmú fokozóval ültette célnyelvre a fordító, addig ezzel a megoldással az *ужасно* esetében egyszer sem élt. Az *ужасно* fordítási megoldásainak gyakorisági megoszlását a 4. ábrán látható diagram mutatja be:

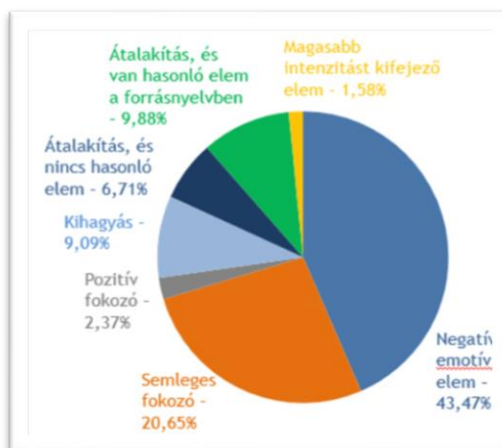


4. ábra. Az *ужасно* fordítási megoldásainak gyakorisági megoszlása

A diagramon szereplő adatok alapján a következő megállapításokat tehetjük: Viszonylag magas számban, az esetek 57,38 %-ban él a fordító tartalmi és stilisztikai szempontból is egyenértékű ekvivalenssel. A leggyakoribb ilyen elem az *awfully*. Kevés veszteség éri a szöveget, ha a fordító – bár átalakítást végez – törekszik a forrásnyelvi kifejezés stilisztikai és pragmatikai tartalmát más eszközökkel pótolni. A vizsgált elem esetében viszonylag kevés alkalommal élt a fordító az átalakítás eszközével kompenzációs törekvéssel együtt (5,11%). A fordítási egységek 37,51%-ában mondhatjuk azt, hogy komolyabb stiláris és pragmatikai veszteség érte a szöveget: 22,72%-ban semleges fokozót használt, 10,22%-ban kihagyta az elemet, 2,84%-ban kompenzáció nélküli átalakítással élt, és végül 1,7%-ban a teljes szókapcsolatot egy magasabb intenzitást kifejező elemre cserélte.

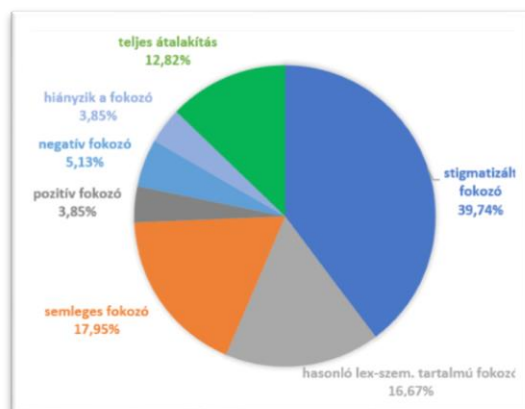
A következőkben azokat az eseteket vesszük górcső alá, amelyekben a vizsgált elemek fordítási megoldásként jelennek meg, tehát a kifejezés célnyelvi oldalán.

A fordítási egységek vizsgálata alapján megállapítható, hogy mind az *ужасно*, mind a *чертовски* elem esetében a fordítási megoldások teljesen egybeesnek azokkal a megoldásokkal, amelyeket az *awfully* esetében láttunk az angol–oros fordítási irányban. A forrásnyelvi megfelelők gyakorisági megoszlását az *ужасно* vonatkozásában az 5. ábra mutatja be:



5. ábra. Az *yachcho* mint fordítási megoldás forrásnyelvi megfelelőinek gyakorisági megoszlása

Az *yachcho* elemet illetően mindenekelőtt érdemes felhívni a figyelmet arra, hogy amíg az *yachcho* elemet egyetlen esetben sem fordították pozitív fokozóval (1. fentebb), addig a pozitív fokozó mint célnyelvi ekvivalens megjelenhet fordítási megoldásként (2,37%). A megoszlási arányok alapján a következő megállapításokat tehetjük: A forrásnyelvben az eseteknek megközelítőleg a felében (43,47%) található egy stilisztikai és pragmatikai szempontból egyenértékű elem, ha az *yachcho* megjelenik a célnyelvben. Ez szintén megfelel kezdeti hipotézisünknek (1. fentebb, 3. pont). Nincs jelentős eltérés azokban az esetekben sem, amikor a fordító átalakítás során kompenzációs céllal alkalmazta a vizsgált elemet (9,88%). Mindössze az esetek 28,27%-ában fordult elő, hogy a fordítási egység forrásnyelvi része nem tartalmazott hasonló stiláris értékű elemet. A *чертовски* elem mint fordítási megoldás forrásnyelvi megfelelőinek gyakorisági megoszlását a 6. ábrán látható diagram mutatja be:



6. ábra. A *chertovskiy* mint fordítási megoldás forrásnyelvi megfelelőinek gyakorisági megoszlása

A *чертовски* forrásnyelvi megfelelőit illetően mindenekelőtt említsük meg, hogy a forrásnyelvi ekvivalensek 39,74%-át stiláris értékét tekintve stigmatizált elem teszi ki. A diagramon szereplő adatok alapján a következő megállapításokat tehetjük: A fordítási egységek közel  $\frac{3}{4}$  részében (74,36%) a célnyelvi megoldás reflektál a forrásnyelvi elem vagy elemek tartalmára. Ezekben az esetekben egy stigmatizált (39,74%), egy hasonló lexikai-szemantikai tartalmú (16,67%) vagy pedig egy negatív fokozó elem (5,13%) található a forrásnyelvi oldalon. Mindössze 25,65%-ban fordul elő, hogy a forrásnyelvi oldal nem tartalmaz a stiláris-pragmatikai értéket tekintve hasonló elemet.

Összevetve a két elem, az *ужасно* és *чертовски* mint fordítási megoldás megoszlását a következőket állapíthatjuk meg: Amíg az *ужасно* egy tartalmi és stiláris szempontból egyenértékű kifejezés fordításaként 53,38%-ban került bele a célnyelvi szövegbe, addig a *чертовски* esetében ez a megoszlási arány jóval magasabb, mivel 73,07%-ban tartalmaz a forrásnyelvi szöveg hasonló értékű elemet. A fentebb említett megoszlási arányok véleményünk szerint a két vizsgált elem eltérő stiláris- pragmatikai értékével állnak összefüggésben.

## 5 Összegzés

A dolgozatban a negatív emotív szemantikai tartalmú fokozó elemeket vizsgáljuk, az angolról oroszra és oroszról angolra történő fordítás szempontjából. A kutatásunkhoz az Orosz nyelv nemzeti korpuszának (НКРЯ) angol–oros párhuzamos alkorpuszát használtuk. A vizsgálatot a következőképpen végeztük el: Először kiválasztottunk az orosz és az angol nyelvben 2-2 olyan elemet, amelyek lexikai szinten negatív emotív szemantikai tartalommal rendelkeznek, képesek értékelést megfogalmazó kifejezésekben fokozó funkciót betölteni, és gyakoriságuk a hasonló elemekhez viszonyítva magas. Az oroszban a *чертовски* ('szörnyen, pokolian, átkozottul'), és az *ужасно* ('borzasztóan, szörnyen'), az angolban a *damn(ed)* ('átkozott, átkozottul, rohadt, istenverte') és az *awfully* ('borzasztóan, szörnyen, rémesen') elemeket választottuk. Miután a korpuszból lekérdeztünk elemeket kvalitatív és kvantitatív szempontból rendszereztük, a vizsgálati eredményeket összevetettük korábban felállított hipotéziseinkkel.

### 5.1 A vizsgálat legfontosabb eredményei

Amennyiben az általunk vizsgált elem a forrásnyelvi oldalon jelent meg, a következő legfontosabb megállapításokat tettük: A *чертовски* elem csekély számban fordul elő forrásnyelvi szövegekben, így annak fordítási megoldásait ki is zártuk a vizsgálatból. Ennek kapcsán felhívtuk a figyelmet arra, hogy a tapasztalat a fordított és nem fordított szövegek eltérésének problémájára mutat rá, hiszen jelen esetben a fordítások során olyan lexéma kerül be gyakran a fordított orosz szövegekbe, amely az eredeti orosz szövegeknek egyáltalán nem jellemző eleme. Az *awfully* és az *ужасно* fordítási megoldásai hasonló megoszlást mutattak, ugyanakkor megállapítottunk néhány fontos különbséget: Egyrészt az *ужасно* elemet gyakrabban fordították negatív emotív tartalmú fokozóval (57,38%), mint az *awfully* fokozószót (36,6%). Utóbbi esetében azonban a kompenzációval történő átalakítás mutatkozott gyakoribbnak az előbbinél. A *damn(-ed)* elem vizsgálata igazolta a hipotézisünket, hiszen a szitokszóval történő

átültetés csupán a *damn(-ed)* elemnél volt jellemző, az *ужасно* és az *awfully* esetében nem. Emellett az átalakítást és a vizsgált elem kihagyását is nagyobb százalékban választották a fordítók e stiláris szempontból stigmatizált elem fordításakor. Érdeemes megjegyezni mindemellett azt is, hogy a semleges fokozók is jelentősen ritkábban jelennek meg a *damn(-ed)* fordítási megoldásaként.

Amennyiben az általunk vizsgált elem a célnyelvi oldalon, tehát fordítási megoldásként jelent meg, a következő legfontosabb megállapításokat tettük: A *damn(-ed)* elem a célnyelvi oldalon csekély gyakorisággal szerepelt. Ez a tapasztalat véleményünk szerint arra mutat, hogy a stigmatizált fokozó elemek használatát illetően a fordítók mintegy óvatosak, és a stigmatizált elemek alapvetően olyan esetekben jelennek meg fordítási megoldásként, ahol a forrásnyelvi szöveg is tartalmaz egy stiláris és tartalmi szempontból egyenértékű elemet.

Az *awfully* és az *ужасно* mint fordítási megoldás forrásnyelvi megfelelőinek megoszlása nagyobb egyezést mutat, mint ha ugyanezeket a *чертовски* adataival vetjük össze. Azt láttuk ugyanis, hogy a *чертовски* esetében a fordítási egységek közel  $\frac{3}{4}$  részében (73,07%) a célnyelvi megoldás reflektál a forrásnyelvi elem vagy elemek tartalmára, emellett 39,74%-ban stigmatizált elemet találunk a forrásnyelvi oldalon. Ugyanakkor az *ужасно* esetében ez az arány jóval kisebb. Ez az elem ugyanis egy tartalmi és stiláris szempontból egyenértékű kifejezés fordításaként csupán 53,38%-ban került bele a célnyelvi szövegbe. A fentebbi tapasztalatok véleményünk szerint a két vizsgált elem eltérő stiláris-pragmatikai értékével állnak összefüggésben. Ami az *awfully* elemet mint fordítási megoldást illeti, megállapítható, hogy az az *ужасно* fokozónál jóval gyakrabban jelentkezik tartalmi és stilisztikai szempontból egyenértékű elem fordítási megoldásaként (61,95%).

A vizsgálat eredményei igazolják kezdeti hipotéziseinket, miszerint, ha a vizsgált elemek a forrásszövegben jelennek meg, a fordítók – annak lexikai szintű affektív tartalma, stiláris értéke miatt – gyakran kerülnek a stiláris-pragmatikai szempontból egyenértékű ekvivalensek használatát a célnyelvi oldalon. Másrésztől a vizsgált elemek – a fentebb írt affektív tartalmi ok miatt – a célnyelvi oldalon gyakorta csak akkor jelennek meg, ha hasonló stiláris értékkel rendelkező elemet találunk a forrásnyelvi oldalon is. Mindemellett mindkét tendenciát a *damn* esetében feltételeztük a legjellemzőbbnek, annak lexikai szemantikai tartalma okán, és ezt a vizsgálati tapasztalatok ugyancsak igazolták a számunkra.

## 5.2 Az eredmények hasznosíthatósága

Reményeink szerint a jelen dolgozatban prezentált eredmények mind a lexikográfiában, mind a fordítástudományban adalékkul szolgálhatnak.

A lexikográfia számára az eredményeinket két szempontból is hasznosnak véljük: Egyrészt azért, mert a két- vagy többnyelvű szótárak szerkesztésekor elengedhetetlen tudnunk azt, hogy a fordítók mely kifejezéseknek mely ekvivalenseit alkalmazzák a célnyelvre ültetés során (vö. pl. Szabó 2012). Csupán ez nyit ugyanis lehetőséget arra, hogy a szótárírásakor a javasolt ekvivalensek közelítsenek a valós, a gyakorlatban előforduló kifejezésekhez. Másrészt az eredmények arra is rámutatnak, hogy az átváltási műveleteket a fordítók az egyes, szótári ekvivalensek esetében messze nem szimmetrikusan alkalmazzák (vö. például az *awfully* és a *чертовски* esetét).

A fordítástudomány számára az eredményeinket ugyancsak két szempontból is hasznosnak véljük: Egyrészt, betekintést engednek a fordítók által e nyelvi jelenség

esetében alkalmazott fordítási megoldásokba. Másrészt rávilágítanak arra, hogy amíg bizonyos elemeket a fordítók előszeretettel alkalmaznak a fordításokban, addig azok az eredeti szövegekben kifejezetten ritka előfordulást mutatnak. Ez a tendencia vezet részben a nem fordított és a fordított szövegek eltérő sajátságaihoz, amelynek a feltárásához, és ezáltal a fordítások minőségének javításához a jelen dolgozathoz hasonló vizsgálatok értékes adalékul szolgálhatnak.

### 5.3 A kutatás további tervezett lépései

A jelen tanulmányban megvizsgált orosz és angol nyelvű fokozó elemek egyaránt eltérő stílusréteghez tartoznak, ami kétségkívül hatással van a vizsgálati eredményekre.

A fordítástudományi irodalomban fordítási univerzáléként kezelt jelenség a normalizálás (vö. Vanderauwera 1985; Shlesinger 1991), nyelvi konzervativizmus (vö. Kenny 2000) vagy kiegyenlítés (levelling out, vö. Baker 1996) művelete, amelynek eredményeképpen a célnyelvi megoldás banálisabb, konvencionálisabb jelleget nyer a forrásnyelvi szöveg stílusához képest.

E sajátságot megfontolva a kutatás egy következő lépéseként azt tervezzük, hogy mindkét nyelv esetében összehasonlítható korpusz segítségével alaposabban megvizsgáljuk a kiválasztott nyelvi elemek gyakorisági megoszlását az adott nyelv fordított és nem fordított szövegeiben, hogy ezzel mélyebb betekintést nyerhessünk azok használati sajátságaiba.

## Irodalom

- Andor J. 2011. De durva ez a téma! – Megfigyelések a melléknévi polaritásváltásról. *Hungarológiai Évkönyv*, 12: 33–42.
- Andor J. 2003. Functional Studies in the Polarity and Gradation of Amplifier Adjectives and Adverbs in English. In: Andor, J., Horváth J., Nikolov M. (szerk.) *Studies in English Theoretical and Applied Linguistics*. Pécs: Lingua Franca Csoport. 43–59.
- Drávucz, F., Szabó M. K., Vincze V. 2017. Szentiment- és emóciósztárak eredményességének mérése emóció- és szentimentkorpuszokon. In: Vincze, V. (szerk.) *XIII. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY 2017*. Szeged: JATEPress. 228–239.
- Jing-Schmidt, Z. 2007. Negativity bias in language: A cognitive-affective model of emotive intensifiers. *Cognitive Linguistics*, 18(3): 417–443.
- Klaudy K. 1997. *A fordítás elmélete és gyakorlata. Angol–német–francia–orosz fordítástechnikai példatárral*. Budapest: Scholastica.
- Kugler N. 2014. A nyelvi polaritás kifejezésének egy mintázata, avagy milyen a félelmetesen jó? *Magyar Nyelvőr*, 138(2): 129–139.
- Szabó M. K. 2012. *A bárki és az akárki névmások fordítási kérdéseinek vizsgálata a magyarról oroszra történő fordítás tükrében*. Díjnyertes pályamunka a „A Magyar Tudományos Akadémia Szegedi Akadémiai Bizottsága és „A Tudomány Támogatásáért a Dél-Alföldön” Alapítvány pályázatán.
- Szabó M. K. 2015a. A polaritásváltás és -változás kezelési lehetőségei a szentimentelemzésben. In: Keresztes Gábor (szerk.) *Tavaszi Szél konferenciakötet*. III. kötet. Eger–Budapest: Linceum Kiadó –Doktoranduszok Országos Szövetsége. 629–643.
- Szabó M. K. 2015b. A polaritásváltás problémája a szentimentelemzés szempontjából. In: Várad T. (szerk.) *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből*. Budapest: MTA Nyelvtudományi Intézet. 51–61.
- Szabó, M. K., Nyíri Zs., Morvay G., Lázár B. 2017. A szentimentérték módosulásának vizsgálata szemantikai–pragmatikai szempontból annotált korpuszon. In: Vincze, V. (szerk.) *XIII. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY 2017*. Szeged: JATEPress. 251–262.

Wierzbicka, A. 2002. Australian cultural scripts – bloody revisited. *Journal of Pragmatics*, 34(9): 1167–1209.

### **Források**

Az orosz nyelv nemzeti korpusza: Национальный корпус русского языка, НКРЯ. Elérhető: <http://www.ruscorpora.ru>. Letöltve: 2017. október 6.



## Anaforafeloldás menet közben – névmások egy pszicholingvisztikailag motivált elemzőben

Vadász Noémi

PPKE BTK Nyelvtudományi Doktori Iskola  
vadasz.noemi@itk.ppke.hu

**Kivonat:** Jelen cikk a koreferenciafeloldás számítógépes nyelvészeti kezelésével foglalkozik egy pszicholingvisztikailag motivált elemzőrendszer keretein belül. A visszautalástípusok közül a profornás – tehát testes vagy zéró névmási – anafora kezelésére helyezek hangsúlyt. Megvizsgálom egy létező koreferenciafeloldó algoritmus (Pléh–Radics 1976) lehetséges alkalmazási lehetőségét az elemzőben. Az emberi szövegfeldolgozást modelláló, balról jobbra és szavanként történő elemzés során a grammatikai viszonyokat kifejező függőségi élek mellett a koreferenciát kifejező élek folyamatosan jönnek létre az elemek között. Az elemző kimenete egy többféle típusú függőségi éleket tartalmazó gráf.

### 1 Az ANAGRAMMA elemző

Az ANAGRAMMA elemző (Prószéky–Indig–Vadász 2016) egy pszicholingvisztikailag motivált (Indig–Vadász–Kalivoda 2016), performanciaalapú szövegelemző, amely az emberi szövegfeldolgozás mintájára balról jobbra és szavanként elemzi a bemeneti szöveget. Az elemzőrendszer működési alapelvei a **kereslet-kínálat** elvű keretrendszer (Prószéky–Indig 2015), a **tározó** és az **ablak**. Az elemző kimenete egy függőségi éleket tartalmazó gráf. Az elemző működési alapelveit a következő alfejezet ismerteti.

#### 1.1 Az ANAGRAMMA működési alapelvei

Az elemző architektúrája párhuzamos, tehát az éppen elemzendő szót a párhuzamosan jelen lévő szálak (morfológiai elemző, különböző grammatikai jelenségeket azonosító szálak, korpuszgyakorisági szálak, anaforafeloldó szálak, fókuszazonosító szál stb.) egyszerre elemzik. Jelen dolgozat a párhuzamosan jelen lévő feldolgozási szálak közül az anaforafeloldó szálra koncentrálna.

A kereslet-kínálat elvű keretrendszerben az elemzés során az egyes tokenek kínálatként jelenhetnek meg – és így más elemek keresleteit kielégíthetik (pl. a melléknevek a főnevek melléknévkeresőit vagy az esetragos főnevek az ige vonzatkeresőit). A keresleteket keresőeljárások valósítják meg, amelyek megszorításokat is tartalmazhatnak (a keresés irányára, a keresett elem bizonyos jegyeire stb.).

A tározó a balról jobbra történő elemzés során az aktuálisan elemzett tokent megelőző tokeneket, azok keresőit, valamint a kész részszerkezeteket tartalmazza. A tározó tartalma lépcsőről lépésre változik az elemzés során, a mondat végére a tározóban nem maradhat kínálat vagy olyan kereső, amely kötelező elemet keres.

Pszicholingvisztikailag motivált elemzőként az ANAGRAMMA az emberi szövegfeldolgozást kívánja modellálni. Ehhez Frazier és Fodor *Sausage Machine* kétfázisú mondatfeldolgozó modelljét (Frazier–Fodor 1978) vettük alapul. A Sausage Machine első fázisában a szöveges bemenet szócsoportjaiból létrejönnek a frázisok, amelyek a második fázisban kapják meg a szerepüket a mondatban. A Sausage Machine első fázisa – az angol nyelvre kidolgozva – egy hat token méretű előretekintő ablakban dolgozik. A Sausage Machine elvét az ANAGRAMMA elemzőre áttűtetve a magyar nyelv agglutináló jellege miatt az előretekintő elemzési ablakot három token méretűnek határoztuk meg. Az előretekintő elemzési ablak azt jelenti, hogy a balról jobbra történő elemzés során egy token elemzésekor az elemzést a tározó tartalmán kívül az adott tokent követő két elem is segíti. Ennek jelentős szerepe van az esetegyértelműsítésben (Vadász–Indig 2016) és a vonzatkeret-egyértelműsítésben (Indig–Vadász 2016; Vadász–Kalivoda–Indig 2017).

## 1.2 Névszói frázisok az ANAGRAMMA elemzőben

Az ANAGRAMMA elemzőrendszerben a névszói frázis (NP) kínálatként viselkedő elem. Az NP-t az ige vonzatkeresője fogja keresni a mondatban, ami az NP fején lévő esetrag alapján történik. A balról jobbra történő kétfázisú elemzésnek általában az első fázisában történik az NP-k előkészítése a kínálati szerepre. Az első fázisban történik a szófai egyértelműsítés és a morfológiai elemzés (amelyek nem képezik jelen cikk tárgyát). Ezután kétféle problémával kell megküzdeni: (1) a testetlen esetragok kezelése és (2) a többértelmű esetragok egyértelműsítése.

Az NP-t mindig esetrag zárja, tehát a balról jobbra történő elemzés során az elemző arra törekszik, hogy az esetrag megtalálásával lezárja az aktuális NP-t, így előkészítve azt kínálatként a vonzatkeresők számára. Az alanyesetű névszók esetében azonban egy testetlen esetrag zárja az NP-t. Amikor az elemző egy testes esetrag nélküli elemmel találkozik, nem lehet biztos benne, hogy az alanyeset testetlen esetragjáról van-e szó, vagy pedig az elem nem zárja le a frázist (pl. a jelöletlen birtokos szerkezet vagy a névutós névszók esetében).

Az elemző a kétfázisú mondatelemzés első fázisában az előretekintő elemzési ablak segítségével deríti ki, hogy a testes esetrag nélküli elem NP-záró nominatívusz esetű elem-e. Ehhez az adott elemtől közvetlenül jobbra elhelyezkedő más elemekre támaszkodik. Például ha mondathatár vagy finit ige (vagy bármilyen, tipikusan NP-n kívüli elem) követi, akkor biztos, hogy nominatívusz esettel lezárt NP-vel talákoztunk. Ha azonban a kérdéses esetű elem elemzési ablakában egy birtokos ragozást magán viselő elem látszik – amely birtokos ragozás a kérdéses elemmel egyező birtokost sejt

<sup>1</sup> Vonzatkeresőnek azokat a keresőket nevezem, amelyek a vonzatkerettel rendelkező elemek összes bővítményét keresik, ideértve a kötelező és opcionális vonzatokat, valamint a szabad határozókat is.

–, akkor a kérdéses elem lehet egy jelöletlen birtokost tartalmazó birtokos szerkezet birtokosa.<sup>ii</sup>

Külön problémát jelentenek a névutót tartalmazó szerkezetek is. A névutók a testes esetragokhoz hasonlóan kezelendők, hiszen az NP a névutóval való összekapcsolódás után válik kínálattá. A névutót megelőző főnévi fej bizonyos névutók esetében nominatívuszi esetben áll, míg más névutók esetében valamilyen testes esetragot visel magán. A névutós névszók esetében is az előretekinthető elemzési ablak nyújt megoldást, hiszen ebben az esetben is a testetlen esetrag tisztázására van szükség. A névutók megfigyelt viselkedése alapján a kérdéses tokenhez nagyon közel, általában közvetlenül utána helyezkednek el, így nem okoz problémát az elemzési ablakban történő kezelésük. Az elemzés további lépéseiben – miután az elemzési ablakban tisztázódott a testes esetrag nélküli elem aktuális szerepe – a névutó a névszó különleges, külön tokenben álló esetragjaként fog funkcionálni: a névutó alapján lesz a névszói frázisból kínálat, hogy az ige vonzatigényét kielégítthesse.

A morfológiai elemző kimenete alapján a névszók címkéi a következőképpen épülnek fel:

| Névszó            |   | Címke  |
|-------------------|---|--|
| főnév             | → | CAS/Nom: tő+N(+PropN)+Sg/Pl(+PersSg/Pl1-3)   |
| melléknév         | → | CAS/Nom: NPMod+Adj+Sg/Pl(+PersSg/Pl1-3)      |
| számnév           | → | CAS/Nom: NPMod+Num+Sg/Pl(+PersSg/Pl1-3)      |
| melléknévi igenév | → |  |
| – folyamatos      | → | CAS/Nom: NPMod+Part+Pres+Sg/Pl               |
| – befejezett      | → | CAS/Nom: NPMod+Part+Past+Sg/Pl               |
| – beálló          | → | CAS/Nom: NPMod+Part+Fut+Sg/Pl(+PersSg/Pl1-3) |

1. táblázat. A névszók morfológiai elemzését tartalmazó címkék

A tokenek morfológiai címkéjének jegyei különböző eljárásokat indítanak az elemzés során. A tokenek főkategóriája – a kettőspont előtti jegy – alapján viselkedik kínálatként az adott elem. A további jegyei alapján elégítheti ki a kereső megszorításait (pl. a számrá és a személyre). Ezenkívül bizonyos jegyek alapján a tokenek saját keresői is elindulhatnak, például az olyan, saját vonzatkerettel rendelkező elemek esetében, mint a melléknévi igenevek.

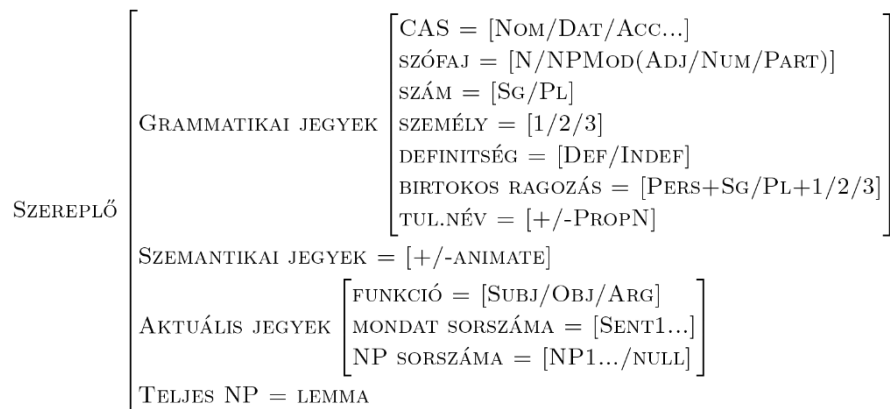
A címkéket minden esetben az esetrag jelölése indítja. Az esetrag jelölése a testes ragok esetében az esetrag nevének hárombetűs rövidítése (pl. Acc, Dat, Ela, Sub stb.), a testetlen esetragok esetében Nom. A testes esetragok a hierarchikus jegyszerkezetben a CAS főkategória alá esnek (így minden olyan folyamatot el tudnak indítani, amelyet az esetraggal rendelkező elemek szoptak, pl. a névelőkeresés). A Nom jegy nem tartozik a CAS főkategória alá, hiszen – amint fentebb már kifejtettem – előbb még tisztázni kell, hogy egy nominatívuszi esetű NP feje-e, vagy nem. Csak az esete tisztázása után tudja az esetraggal rendelkező elemek által indítandó folyamatokat elindítani, legyen szó akár tisztázott nominatívuszi esetről, akár névutós névszóról.

A címke kettőspont utáni részében a tő jelölése után a lexikális elem aktuális jegyei sorakoznak. A hagyományos szófajokat kiegészítve hierarchikus jegyszerkezetek

<sup>ii</sup> Az elemzési ablak működésének részleteit az eset-egyértelműsítés megoldásában, valamint az azokat megalapozó korpuszméréseket Vadász–Kalivoda–Indig 2017 részletezi.

fejézik ki az adott szófajokra jellemző tulajdonságokat. A főnevet módosító összes szófaj (Adj, Num, Part) rendelkezik a közös NPMod címkével, ami azt jelenti, hogy ők mindannyian a főnevet balról módosító elemek. A főnevek címkéje között megjelenhet a PropN jegy a tulajdonnevek esetében, amelynek a névelőkeresés és a definitiség beállítása során van jelentős szerepe. Az összes névszó rendelkezik számot kifejező jeggyel (Sg vagy Pl), valamint mindegyikhez járulhat birtokos ragozást kifejező jegy is (Pers), amely a keresett birtokos számára (Sg vagy Pl) és személyére (1, 2 vagy 3) tesz megszorítást.

A koreferenciafeloldáskor fontos információ az NP címkéjében található jegyek összessége, de ezeket a morfológiai jegyeket néhány további jeggyel is ki kell egészíteni. A kétfázisú mondaelemzés első fázisában lezárt – tisztázott esetű – NP-k egy, az összes szükséges információt tartalmazó jegyszerkezettel jelennek meg kínálatként a második fázisban működő anaforafeloldó szál keresői számára. A koreferenciafeloldó számára rendelkezésre álló kínálat az alábbi jegyszerkezetekkel rendelkező ún. Szereplő.



**1. ábra.** A Szereplők jegyszerkezete, amellyel a koreferenciakeresésben résztvevő elemek rendelkeznek

A Grammatikai jegyek mellett (ami a morfológiai elemzéből kinyert címke tartalma alapján töltődik ki) megjelennek Szemantikai jegyek is. Az elemző jelenlegi állapotában csak a [+/-élő] jegy tartozik a szükséges szemantikai jegyek közé. A [+/-élő] jegy a magyar nyelv esetében az *ember* kategóriára vonatkozik, nyelvi szempontból az állatok [-élő] jegyűek (Frawley 1992). Az elemzés során a vonatkeret lehívásakor szükség lehet arra az információra is, hogy az ige egyes vonzataira vonatkozik-e [+/-élő] megszorítás. Emellett a szótár a főnevek [+/-élő] szemantikai jegyeit is kell hogy tartalmazza, hiszen jelentősége lehet a koreferenciafeloldáskor. A vonatkeretszótár összeállításánál, valamint a főnevek [+/-élő] jegyének megállapításánál a MetaMorpho rendszer (Tihanyi–Merényi 2006) lexikonjára támaszkodunk.

A grammatikai és a szemantikai jegyek mellett az Aktuális jegyek is a koreferenciafeloldáshoz szükséges információkat szolgáltatnak, közöttük az NP mondatbeli funkcióját (alany: Subj, tárgy: Obj vagy egyéb vonat: Arg) és a

definittségét (Def vagy Indef) kifejező jegyeket. A koreferenciafeloldás során az NP több mondatban is kaphat funkciót, ennek megfelelően az aktuális funkciójához jegyezzük a mondat sorszámát is, valamint azt is, hogy a mondat hányadik NP-jeként töltötte be az adott funkciót (ennek értéke lehet Null is).

## 2 A visszautalások

Pléh (1998) alapján a visszautaló kifejezések egyik végpontját az eredetileg deiktikus, indexikus elemek képviselik (*az, ő, ez, oda* stb.), míg a másik végpontra a szemantikailag tartalmas, helyettesítő szerepű elemek helyezkednek el.

| Példa   | Visszautalás típusa      |
|---|--------------------------|
| A gyerek meglátta <i>a kutyát</i> .<br><i>A kutya</i> ugatni kezdett.             | ismétlés                 |
| Laci vett <i>egy pulit</i> .<br><i>A kutya</i> sokat ugat.                        | alárendelés, fölrendelés |
| Feri meglátta <i>a rendőrt</i> .<br><i>A zsarú</i> hirtelen megfordult.           | szinonima                |
| <i>A gyerek</i> megölelte a tanárt.<br>Leült és elmosolyodott.                    | zéró anafora             |
| <i>A gyerek</i> megpusztilta <i>a nénit</i> .<br>Ő meg elmosolyodott.             | névmási anafora          |
| <i>Ede</i> elhagyta Marit.<br><i>A disznó</i> még el sem köszönt.                 | epitheton                |
| <i>A rendőr</i> belépett <i>a szobába</i> .<br><i>Az ablak</i> nyitva volt.       | valószínű rész           |
| Tegnap lerobbant <i>a kocsim</i> .<br><i>A fém</i> romlott el.                    | szükségszerű rész        |
| <i>A rabló</i> <i>lelőtte</i> a pénztárost.<br><i>A pisztoly</i> a földön hevert. | esetkeret                |

**2. táblázat.** Visszautalási típusok Pléh (1998) példáival. A koreferenciaviszonyt dőlt betű jelzi

A koreferenciafeloldás körébe tartozó feladatok megoldása számítógépes eszközökkel két osztályba sorolható, léteznek (1) tudásalapú és (2) „tudásszegény” megoldások. Előbbire példa a magyar nyelv esetében Lejtovicz és Kardkovács (2006) megoldása, amely Brennan, Friedman és Pollard (1987) diskurzusalapú – a figyelem középpontjában álló entitások követését modellező – algoritmusára támaszkodik. Egy másik példa magyarra alkalmazott tudásalapú megoldásra Miháltz (2012) többféle tudásra támaszkodó szabályalapú rendszere; a MetaMorpho gépi fordítórendszer (Tihanyi–Merényi 2006) magyar mély nyelvi elemzőjének kimenetében található morfológiai, szintaktikai és szemantikai információkra; a kormányzás és kötés elméletének a magyar szintaxiselméletben megfogalmazott változatára, továbbá a magyar mondatmegértés pszicholingvisztikájában elért kutatási eredményekre támaszkodó szabályokra; a Magyar WordNetben található szemantikai tudásra; valamint karakteralapú heurisztikákra. A „tudásszegény” megoldások felügyelt gépi tanulást alkalmaznak, amelyhez nagyméretű, kézzel annotált tanítókorpusra van szükség (Ng 2005).

A proformás névmások működésének alapelve az, hogy használatukkal a szöveg lineáris sorrendjében közelebb kerülnek egymáshoz az azonos referenciájú elemek (Pléh–Radics 1976). Jelen dolgozatban egy olyan szabályalapú, algoritmikus megoldást javaslok, amely az input nyelvtani elemzővel történő feldolgozása révén rendelkezésre álló szintaktikai információkat használ. Pléh és Radics (1976) anaforafeloldó algoritmusát vettem alapul. Ez utóbbi a proformás anaforák használata mögött húzódó szabályrendszert implementálható algoritmusként mutatja be, amely elsősorban a mondatban szereplő névmások és NP-k felszíni nyelvtani tulajdonságaival operál. A 2. táblázatban felsorolt visszaütalás-típusok közül jelen dolgozatban a zéró és névmási anaforával foglalkozom részletesebben. Ezeket összefoglaló néven proformás anaforának nevezzük. A többi visszaütalás-típus kezeléséhez szükség van olyan információkra, amelyhez nem elegendő a morfológiai elemző kimenete.

### 3 A Pléh–Radics-algoritmus

Pléh és Radics munkája (1976) alapján kijelenthető, hogy a (zéró vagy testes) névmások használata mögött egy feltárható szabályrendszer húzódik, amely lehetővé teszi az elemek vonatkozásának visszaállítását (Pléh–Radics 1976). Megkísérelték ennek a szabályrendszernek a feltárását, amelyre a továbbiakban Pléh–Radics-algoritmusként fogok hivatkozni. Az algoritmus a mondat alanyának, tárgyának vagy egyéb argumentumának előző mondatbeli antecedensét keresi.

Az algoritmus alkalmazása azokban az esetekben érdekes, amelyekben az anaforikus vagy zéró névmást megelőző szövegrészben több főnévi csoport szerepel, hiszen a szabályrendszer azt határozza meg, hogy a lehetséges jelöltek közül melyik főnévi csoport a névmás antecedense.

| Az 1. mondat NP-je |             | A 2. mondatban |                  |
|--------------------|-------------|----------------|------------------|
| szerep             | egyéb jegy  | szerep         | megjelenés       |
| alany              |             |                | ∅                |
| tárgy              |             | alany          | az               |
| egyéb arg.         |             |                |                  |
| alany              | [-élő]      |                | ∅                |
| tárgy              | [+élő]      | tárgy          | ∅ vagy szem. nm. |
| egyéb arg.         | többes szám |                | szem. nm.        |
| alany              |             |                |                  |
| tárgy              |             | egyéb arg.     | szem. nm.        |
| egyéb arg.         |             |                |                  |

3. táblázat. A Pléh–Radics-algoritmus működésének összefoglaló táblázata

A 3. táblázat a Pléh–Radics-algoritmus működését mutatja be. A második mondat alanya zéró névmás akkor, ha koreferens az előző mondat alanyával. Ezt a jelenséget

alanytörlésnek nevezik, és több predikátum esetén is fennáll. Ha az előző mondat tárgyával vagy egyéb argumentumával koreferens, akkor az alanyváltást az *az* távolra mutató névmás jelzi. Az alany előző mondatbeli antecedensének megkeresése nem függ az előző mondatbeli főnévi csoport szemantikai jegyeitől.

A második mondat tárgyának antecedenskeresése ezzel szemben függ a főnévi csoport [+/-élő] szemantikai jegyétől. Ha a második mondat tárgyat zéró névmás fejezi ki, akkor az előző mondat [-élő] jegyű főnévi csoportjával lehet koreferens. Ha személyes névmás fejezi ki, akkor az előző mondat egy többes számú főnévi csoportjával lehet koreferens. Ha azonban a második mondat tárgya az előző mondat tárgyával koreferens, akkor az zéró és személyes névmásként is realizálódhat. Ez azt jelenti, hogy a második mondat tárgya előző mondatbeli antecedensének meghatározása a többértelmység miatt problémás, ugyanakkor a tárgyi szerepű személyes névmás antecedenskeresését segíti a számbeli és személybeli egyezés az előző mondat valamely nem alanyi argumentumával.

A második mondat egyéb argumentumának koreferenciáját az előző mondat bármilyen szerepű főnévi csoportjával személyes névmás fejezi ki, így ezekben az esetekben a számítógép a számbeli és személybeli egyezésekre támaszkodhat.

A fentieknek megfelelően a Pléh–Radics-algoritmus a koreferenciafeloldás folyamatát a hallgató szempontjából egy többlépcsős döntési algoritmussal írja le (Pléh–Radics 1976: 275.):

1. Keresd meg a mondat alanyi funkcióban levő főnévi csoportját; ha ez zéró, akkor az alany megfelel az előző mondat mélyszerkezetei alanyának; ha pedig *az*, akkor az előző mondat valamelyik nem alanyi főnévi csoportjának felel meg.
2. Ha az állítmánynak más vonzata is névmási vagy zéró formában van, akkor annak az előző mondat valamelyik főnévi csoportja felel meg, mégpedig úgy, hogy ha 1. fennáll, akkor a másik.
3. Ha a második mondat alanya nem azonos az előző mondat alanyával, akkor a másodikban levő személyes névmást inkább az első mondat alanyával értelmezzük koreferensnek.

Ezek mellé a perceptuális-értelmezési szabályok mellé járul a mélyszerkezeti alany megkeresésének problémája, amely azonban nem tárgya a jelenlegi cikknek.

#### 4 Koreferenciafeloldás az ANAGRAMMA elemzőben

A fejezet azt mutatja be, hogy hogyan lehet a Pléh–Radics-algoritmust az ANAGRAMMA balról jobbra és szavanként elemző keretrendszerében alkalmazni. A kétfázisú mondatelemzés első fázisában az egyértelműsített esetű NP-k az 1.2 fejezetben ismertetett módon Szereplővé válnak, tehát kínálatként készen állnak a második fázisban zajló anaforafeloldó szál keresői számára.

A koreferenciafeloldás a vonzatkereséssel összefonódva, azzal párhuzamosan történik. Az igei elem vonzatkeret-egyértelműsítése (Vadász–Kalivoda–Indig 2017) után kezdődik a vonzatok keresése, megfelelő megszorításokkal a keresett elemekre. A következőkben az alany és a többi vonzat keresésével párhuzamosan történő koreferenciafeloldást ismertetem. A példákban a finit ige vonzatait keresem, az infinitívusz és a többi vonzatkerettel rendelkező elem vonzatkeresését jelen cikkben nem tárgyalom.

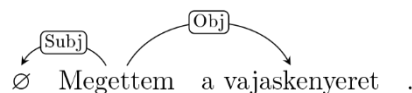
Az alfejezetek példái a vizsgált szempontok alapján létrejövő függőségi éleket szemléltetik. A példamondatok fölötti kerek élek a mondatok főigéje és vonzatai között létrejövő vonzatóéleket reprezentálják, a példamondatok alatti szögletes élek a mondatok szereplői között létrejövő koreferenciaéleket.

#### 4.1 Az alany keresése

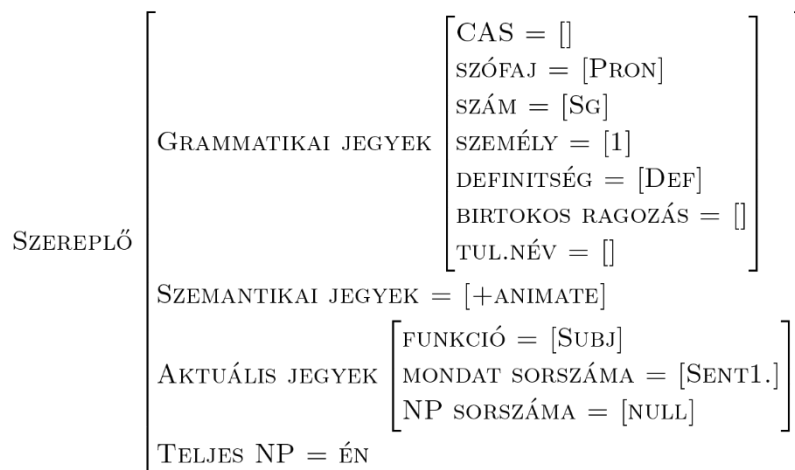
Az elemző alanykereséskor a tisztázott nominatívusz esetű elemeket keresi. Az alanykeresés legelső lépéseként a finit ige ragozásából kiszámolja a keresett alany számát és személyét, és létrehoz egy zérus csomópontot ezekkel az értékekkel. A zérus csomópont Szereplőként megkapja az 1.2 fejezetben ismertetett jegyszerkezetet, és rögtön megkapja az alany szerepét (és Subj jegyet kap).

A testetlen névmással kifejezett alanyt tartalmazó mondat függőségi elemzését az [1] példa szemlélteti.

[1]



Az alany zérus csomópontjának Szereplő jegyszerkezetét a 3. ábra mutatja be. A jegyszerkezeten látható, hogy a névmás szófajú (Szófaj = Pron) elem az első mondat (Mondat sorszáma = Sent1) néma elemeként (NP sorszáma = Null) alany (funkció = Subj) szerepet kap.



3. ábra. Az alany szerepű néma proformás anafora Szereplő jegyszerkezete

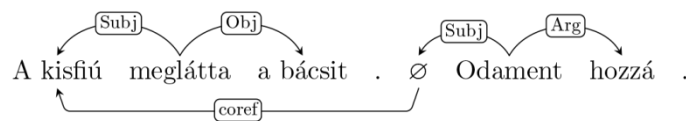
Az alanyok keresésekor – és a default zérus csomópont létrehozásakor – tehetünk további megszorításokat. Az alany hajlamos arra, hogy [+élő] jegyű, ágens tematikus



szerepű elem legyen (Frawley 1992). Ugyanakkor bizonyos igék csak egyes szám 3. személyű ragozásban állhatnak és alanyuk [-élő] jegyű (Komlósy 1992). Az 1. és 2. személy mindig élő alanyra utal, csak a 3. személyű ragozás esetében kell döntenünk. Sass (2007) kidolgozott egy eljárást, amely az élettelen alanyú igei keretek kiválasztására alkalmas (felszíni, morfológiai tulajdonságok alapján).

Az ismétlődő alany törlésének szabálya (Pléh–Radics 1976) alapján: ha volt előző mondat, és a finit ige ragozása az előző mondat finit igéjének ragozásához képest nem változott, akkor a zérus csomópont koreferens lesz az előző finit ige alanyával. Ezt az esetet a [2] példa illusztrálja.

[2]



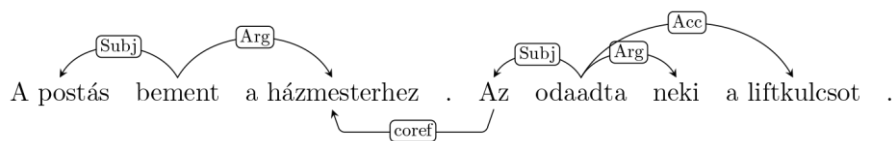
Ha az aktuális mondatban az elemzés során érkezik egy tisztázott nominatívusz esetű NP, amely szám és személy jegyeivel megfelel az ige által indított alanykereső megszorításainak, akkor az Szereplőként átveszi az alany szerepét a zérus csomóponttól (és Subj jegyet kap). A [3] példán látható, hogy a második mondat nem koreferens az előző mondat egyik Szereplőjével sem.

[3]



Egyébként ha a mondatban szerepelt egy egyértelműsített *az* távolra mutató névmás, akkor átveszi a zérus csomóponttól az alany szerepét (és Subj jegyet kap), valamint koreferenciaél jön létre közte és az előző mondat egy Arg jegyű elemével. Ezt az esetet mutatja be a [4] példa.

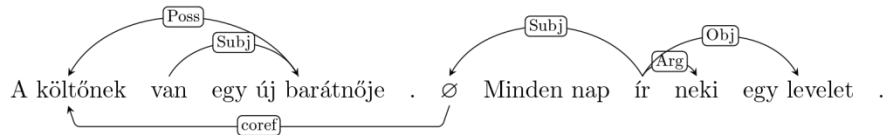
[4]



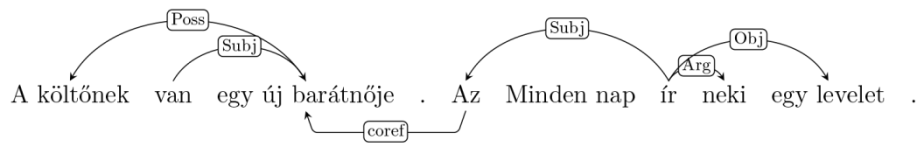
Mivel az alanykeresés az elemző jelenlegi állapotában csak a tisztázott nominatívusz esetű elemekre korlátozódik, ezért a datívusz esetű mélyszerkezeti alany problémájával nem képes megküzdeni. Pléh–Radics (1976) megállapította, hogy „a második mondat alanyi főnévi csoportja zéró formában realizálódik, ha az első mondat mélyszerkezeti alanyával koreferens, s mutató névmás (*az*) formában, ha az előző mondatbeli antecedense a mondatnak ott nem mélyszerkezeti alanya volt.” (Pléh–Radics 1976: 266). Az előbbi esetre lásd az [5] példát, az utóbbira a [6] példát, amelyek Pléh–Radics (1976) példái. Az [5] példában a második mondat zéró alanya az előző mondat datívusz

esetű elemével (a mélyszerkezeti alannyal), míg a [6] mondat távolra mutató névmással kifejezett alanya az előző mondat nominatívusz esetű bővítményével koreferens.

[5]



[6]

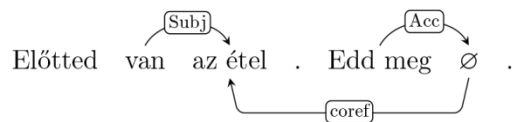


#### 4.2 A tárgy keresése

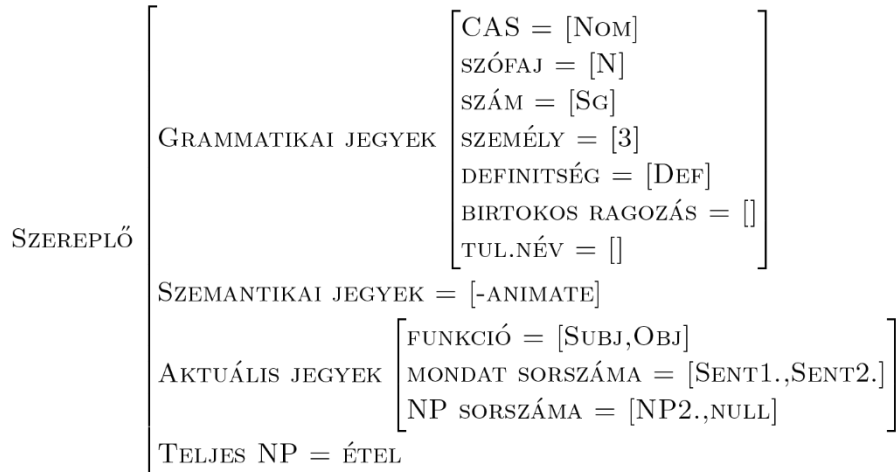
A kereső a tárgy esetében is az NP-k esetragja alapján keres. A tárgy keresésekor az alanyhoz hasonlóan az elemző a finit ige ragozásából kiszámítja a keresett tárgy számát, személyét és definitiségét, valamint létrehoz egy zérus default csomópontot a megfelelő jegyekkel. A zérus csomópont Szereplőként megkapja a jegyszerkezetet és rögtön megkapja a tárgy szerepét (Obj jeggyel).

Ha volt előző mondat, amelyben volt egy Subj jegyű, egyes számú elem, amelynek a jelenlegi mondatban nincs Subj jegye, akkor a tárgy azzal az elemmel koreferens lesz (így az a Szereplő megkapja a második mondat finit igéjének Obj jegyét). A [7] példa egy ilyen esetet mutat be.

[7]



Ebben az esetben a Szereplő jegyszerkezete a 4. ábrán látható módon fest.

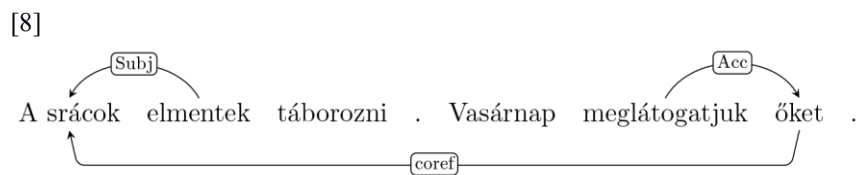


4. ábra. A tárgy szerepű elem jegyszerkezete, amely koreferens az előző mondat alanyával

A 4. ábrán látható példa azt illusztrálja, hogy hogyan változik egy Szereplő NP jegyszerkezete mondatról mondatra. Alanyesetű névszói frázisként az első mondatban az alany szerepét töltötte be, majd a második mondatban zéró névmásként a tárgy szerepét.

Egyéb esetben az előző mondat egyik nem Subj jegyű elemével lesz koreferens (amely szintén Obj jegyet kap).

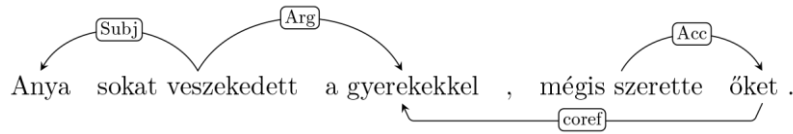
Ha az aktuális mondatban van Acc esetű személyes névmás, amely az előző mondat Subj jegyű elemével számban egyezik, és amelynek a jelenlegi mondatban nincs Subj jegye, akkor ezzel az elemmel lesz koreferens a tárgy (és Obj jegyet kap), amint a [8] példán látható.



Egyéb esetben az előző mondat egyik nem Subj jegyű, számban egyező elemével lesz koreferens.

Ha az aktuális mondatban szerepel egy megfelelő tárgyesetű, nem személyes névmási főnévi csoport, akkor az veszi át az ige tárgyi vonzatának helyét, és koreferenciát nem állapítunk meg. Ha a tárgy személyes névmás, akkor vele egy számban és személyben egyező nem alanyi vonzat lesz koreferens az előző mondatban. Ezt illusztrálja a [9] példa.

[9]

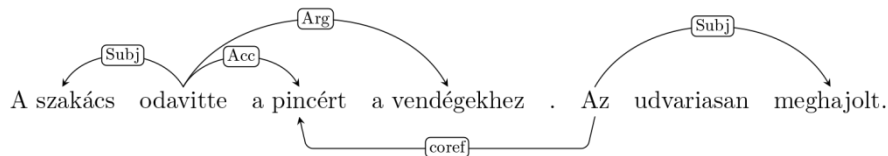


### 4.3 A többi vonzat keresése

Az elemző a nem alanyi vagy tárgyi vonzatokat is az esetragjuk alapján keresi. A megfelelő esetragos személyes névmásnak az előző mondatban egy számban és személyben egyező főnévi csoport lehet az antecedense. Ha egy proformás anafora számára több lehetséges jelölt is rendelkezésre áll antecedensként (a grammatikai és az aktuális jegyek alapján), akkor a szemantikai jegy segíthet az egyértelműsítésben.

A [10] példában látható, hogy az alanyváltás után a második mondat alanya csak az előző mondat tárgyával lehet koreferens, hiszen a másik argumentummal nem egyezik számban.

[10]



### 4.4 Többértelmű koreferenciaviszonyok

A nem alanyi vonzat szerepű proformás anafora antecedensének keresésekor előfordulhat, hogy a megelőző mondatban több lehetséges jelölt is szerepel. Ekkor pusztán a nyelvtanszerű szabályrendszerre támaszkodva nem tudunk dönteni.

A [11] példa második mondatában mindkét proformás anafora esetében több antecedensjelölt is felmerül a megelőző mondatban. Az *az* távolra mutató névmás mint az alanyváltást jelölő elem az előző mondat valamely nem alanyi vonzatával lesz koreferens (*a rendezővel* vagy *az operatőrrel*). A *hozzá* személyes névmás a második mondat egyéb argumentumaként az előző mondat bármely Szereplőjére vissza tud utalni, így koreferens lehet akár *a színésszel*, akár *a rendezővel*, akár *az operatőrrel*. *A rendező* és *az operatőr* a második mondatban természetesen csak egy szerepet kaphat, így ha a rendezővel koreferens az *az* távolra mutató névmási alany, akkor a *hozzá* névmás már nem lehet vele koreferens.

[11] A színész szolt a rendezőnek az operatőrrel. Az odajött hozzá.

## 5 Összefoglalás és távlatok

A cikkben egy koreferenciafeloldó algoritmus alkalmazását ismertettem az ANAGRAMMA elemzőrendszer keretein belül. Bemutattam, hogy hogyan működik az alany, a tárgy és az egyéb argumentumok antecedenskeresése a balról jobbra és szavanként történő elemzés során.

Az eljárást egyelőre csak a proformás (zéró- vagy névmási) anaforák antecedenskeresésére dolgoztam ki. A kutatás folytatásaként olyan megoldásokat vizsgálok meg, amelyek a többi visszautalástípust képesek kezelni, és alkalmazom őket az ANAGRAMMA keretein belül. A kutatást ezen kívül ki kell egészíteni a birtokosok antecedenskeresésével is. Pléh (1998) kísérletekkel bizonyította, hogy a proformás anaforák feloldásakor fontos szerepet játszik a szórend (különösen az *az* proforma esetében). Emellett a különböző tematikus szerepű elemek fókuszba emelése, valamint a morfológiai jelöltség is befolyásolja a proformás anaforák feloldását. A következőkben az ezen megfigyelések alapján levonható következtetéseket is szeretném hasznosítani az ANAGRAMMA számára.

## Irodalom

- Brennan, S. E. – Friedman M. W. – Pollard C. J. 1987. A centering approach to pronouns. In: C. L. Sidner (szerk.) *Proceedings of the 25th meeting of the Association for Computational Linguistics*. Stroudsburg: Association for Computational Linguistics. 155–162.
- Frawley, W. 1992. *Linguistic Semantics*. London: Lawrence Erlbaum.
- Frazier, L., Fodor, J. D. 1978. The Sausage Machine: A New Two-Stage Parsing Model. *Cognition* 6(4): 291–325.
- Indig, B., Vadász, N. 2016. *Window in Human Parsing – How Far can a Preverb Go?* Elhangzott: Tenth International Conference on Natural Language Processing (HrTAL2016), Dubrovnik, 2016. szeptember 29–30.
- Indig, B., Vadász N., Kalivoda A. 2016. Decreasing Entropy: How Wide to Open the Window? In: Martín-Vide, C., Mizuki, T., Vega-Rodríguez, M. A. (szerk.) *Theory and Practice of Natural Computing. 5th International Conference, TPNC 2016, Sendai, Japan, December 12–13 2016, Proceedings*. Cham: Springer. 137–148.
- Komlósy, A. 1992. Régensek és vonzatok. In: Kiefer, F. (szerk.) *Strukturális magyar nyelvtan I. Mondattan*. Budapest: Akadémiai Kiadó. 279–529.
- Lejtovicz, K., Kardkovács, Zs. 2006. Anaforafeloldás magyar nyelvű szövegekben. In: Alexin, Z., Csendes, D. (szerk.) *MSZNY2006. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem. 362–364.
- Miháltz M. 2012. Tudásalapú koreferencia- és birtokosviszony-feloldás magyar szövegekben. In: Prószycki, G., Váradi, T. (szerk.) *Általános Nyelvészeti Tanulmányok XXIV*. Budapest: Akadémiai Kiadó. 151–166.
- Ng, V. 2005. Machine learning for coreference resolution: From local classification to global ranking. In: Knight, K., Ng, H. T., Oflazer, K. (szerk.) *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. Ann Arbor, MI: Association for Computational Linguistics. 157–164.
- Pléh, Cs. 1998. *A mondatmegértés a magyar nyelvben*. Budapest: Osiris.
- Pléh, Cs., Radics, K. 1976. „Hiányos mondat”, pronominalizáció és a szöveg. *Általános Nyelvészeti Tanulmányok* 11(4): 261–277.
- Prószycki, G., Indig, B. 2015. Magyar szövegek pszicholingvisztikai indíttatású elemzése számítógéppel. *Alkalmazott Nyelvtudomány* 15(1–2): 29–44.
- Prószycki, G., Indig B., Vadász N. 2016. Performanciaalapú elemző magyar szövegek számítógépes megértéséhez. In: Kas, B. (szerk.) „*Szavad ne feledd!*”: *Tanulmányok Bánréti Zoltán tiszteletére*. Budapest: MTA Nyelvtudományi Intézet. 223–232.
- Sass, B. 2007. Élő vagy élettelen? In: Tanács, A., Csendes, D. (szerk.) *MSZNY2007. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem. 195–203.

- Tihanyi, L., Merényi, Cs. 2006. A MetaMorpho fordítóprogram projekt 2006-ban. In: Alexin, Z., Csendes, D. (szerk.) *MSZNY2006. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem. 169–179.
- Vadász, N., Indig, B. 2016. *A birtokos esete az ablakkal*. Elhangzott: LingDokKonf 20. Nyelvészdoktoranduszok Országos Konferenciája, Szeged, 2016. november 30. – december 1.
- Vadász, N., Kalivoda Á., Indig B. 2017. Ablak által világosan – Vonzatkeret-egyértelműsítés az igekötők és az infinitívuszi vonzatok segítségével. In: Vincze, V. (szerk.) *XIII. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY 2017*. Szeged: JATEPress. 3–12.