

A negatív emotív elemek vizsgálata a nemek közötti nyelvhasználati különbségek szempontjábólⁱ

Szabó Martina Katalin^{1,2,3}, Lázár Bernadett¹, Nyíri Zsófi¹, Morvay Gergely¹

¹PrecognoX Informatikai Kft

²Szegedi Tudományegyetem, Bölcsészettudományi Kar, Szláv Intézet

³„Lendület” Research Center for Educational and Network Studies (RECENS)

szabo.martina@lit.u-szeged.hu

(mszabo, blazar, zsnnyiri, 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, rohadt 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.

ⁱ 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 fejtszenek ki (pl. *brutális volt a hétvégi koncert*) (vö. Szabó 2015). Az előbbi esetben a vizsgált elemek más, értékelő kifejezések jelentését fokozzák (ezt a jelenséget nevezzük **értékvesztésnek**), az utóbbi esetben pedig maguk töltik be értékelő 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) mellett 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) hozzáteszi, 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 *verdamm* ('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) mellett érvel, hogy a fenti megfigyelések, még ha csupán indirekt módon is, de azzal a kulturális sajátsággal vannak kapcsolatban, hogy míg a férfi beszélőktől a harag érzelm nyelvvi 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 érzelm kifejező fokozó elemek használati sajátságaival is.

3 A nyers korpusz bemutatása és elemzési eredményeiⁱⁱ

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.

ⁱⁱ 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 sajátosságokat 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 alkorpust automatikus módszerrel elemeztük a 6 és a 8 érzelemtípus nyelvi elemeit tartalmazó emóciólexikonunkkal.

Az emóciószótáraink létrehozásakor Ekman és Friesen (1969) érzelemkategorizálási 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 érzelemtípussal szükséges kibővítenünk: a FESZÜLTség 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 átlaga	Nők átlaga	Különbség	Standard hiba	t-érték (szabadság- fok)	Szórás	p-érték
bánat	0,00189	0,00207	-0,00018	0,00017	-1,07812 (36543)	0,02336	0,28099
düh	0,00098	0,00111	-0,00013	0,00014	-0,97661 (36716)	0,01856	0,32877
félelem	0,0006	0,00079	-0,00019	0,0001	-1,8909 (32144)	0,01412	0,05865
meglepődés	0,00047	0,00076	-0,00029	0,00012	-2,49625 (36667)	0,0159	0,01256
öröm	0,00801	0,00685	0,00116	0,00038	3,07792 (37360)	0,05206	0,00209
undor	0,00155	0,00182	-0,00027	0,00016	-1,71795 (34442)	0,02201	0,08582

2. táblázat. A 6 szólistával végzett elemzés eredményeiⁱⁱⁱ

Emóció	Férfiak átlaga	Nők átlaga	Különbség	Standard hiba	t-érték (szabadságfok)	Szórás	p-érték
bánat	0,00175	0,00152	0,00023	0,00015	1,48971 (37691)	0,02069	0,13631
düh	0,00033	0,00019	0,00014	0,00006	2,22208 (31482)	0,00871	0,02629
félelem	0,00038	0,00024	0,00014	0,00006	2,29052 (37622)	0,00844	0,022
meglepődés	0,00034	0,00022	0,00012	0,00008	1,44357 (32549)	0,01158	0,14887
öröm	0,00708	0,00579	0,00129	0,00036	3,64017 (37086)	0,04905	0,00027
undor	0,00141	0,00128	0,00013	0,00014	0,95721 (36853)	0,01914	0,33847
feszültség	0,001	0,00068	0,00032	0,00011	2,88852 (37162)	0,01536	0,00387
vonzalom	0,00076	0,00048	0,00028	0,0001	2,78117 (35838)	0,0139	0,00542

3. táblázat. A 8 szólistával végzett elemzés eredményei^{iv}

ⁱⁱⁱ A férfítweetek száma (N) 18236, a női tweetek száma 19559, az összevetés kétoldalú, kétmintás t-próbával készült.

^{iv} 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 elemeket 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.^v (A munkához az annotálni kívántak alapján az eszközt magunk állítottuk be (az eszközzől 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 trolí.

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!

^v 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 tudjuk, 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.

100	xnuffx_625896339566669824.txt,Tegnap durván leégett az a jóllakott óvodás fejem a szoliban...	deszemant	alappem
101	(nagymosoly) (2 hete nem voltam) [[durván]] durván		
102	MrSuperEgo_587687911019118593.txt,A váltsérülés a múlté.		
103	Ez ma kiderült edzés közben (halk juhé).		
104	Az is kiderült hogy három hónap alatt elképeszt?en sokat romlott a formám!	deszemant	alappem
105	[[elképeszt?en]],elképeszt?en		
106	wasandras_526364731100397568.txt,"épp az el7bb baszott le az Ocsém, hogy öregszem, mert nem értek a Snapchat-hez...	deszemant	alappem
107	(nagymosoly) [[baszott]] baszott		
108	BaracskaGreta_627620118747607040.txt,Itt szól a szomszédba brutál hangosan a mulatós zene 01:22-kor.. gratula.. nem lehet aludni! (semleges)?	deszemant	alappem
109	[[brutál]] brutál		
110	szokeptr_578180711032713216.txt,"fuu, kurva szar lett az OS X wifi kezelése 10.10 után [[kurva]] kurva	deszemant	alappem
111	szilagy_l_vivien_584998216049000448.txt,amikor 2 nap alatt alszol durván 10 órát (mosoly) felbecsülhetetlen [[durván]] durván		
112	kockasfulu_570244277810421760.txt,"nézd a jó oldalát: aki ennyiért marad, baromira elhivatott lehet.	deszemant	alappoz
113	[[baromira]] baromira		

1. ábra. Részlet a korpusz annotációjából a brat programban

Ahhoz, hogy az annotátorok közötti konszenzust mérni tudjuk, 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 értékváltást, és 308 esetben értékvesztést, tehát fokozó funkciót.

Megvizsgálva az értékváltás eseteit, a következő megállapításokat tehetjük: az összesen 75 értékváltás 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 értékváltás, 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 értékváltó 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 értékváltó 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 értékváltóként, 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 sajtóságot is lekérdeztük (vö. Szabó et al. 2017). Az adatokat a 3. táblázat mutatja be.

		A vizsgált elem aktuális szemantikai tartalma				ÖSSZESEN:
		deszem	negatív	pozitív	egyéb	
Az alaptag szemantikai tartalma	alappoz	128	1	1	1	131
	alapneg	86	1	–	8	95
	alapsem	42	37	3	–	82
	egyéb	–	–	–	–	–
	ÖSSZESEN:	256: 83,11%	39: 12,66%	4: 1,29%	9: 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 színészióriás

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.

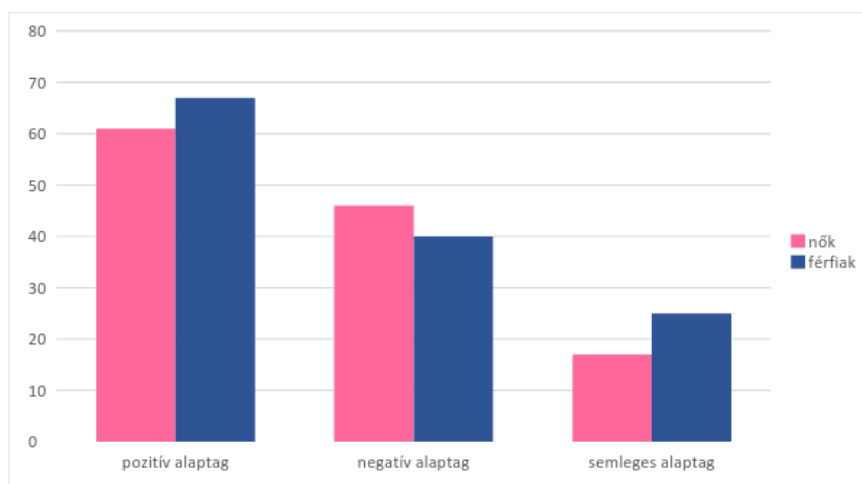
[8] Exportáltam pár képet erről a **brutálisan sűrű** és mozgalmas hétről, hogy legyen mit tolni instára, persze hogy minden az iMac-en maradt

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 mellett 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 sajátosságokkal 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], rohadt[-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. Érdemes 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 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 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 emóciósztárainkkal 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. Érdemi

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 DŰH, a FÉLELEM, a FESZÜLTSEG, 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 értékváltás 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 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”. 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 amellet, 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 értékváltás 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: Geccsó, 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.
- Tolecsvai 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.